



UNIOESTE – Universidade Estadual do Oeste do Paraná

CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS

Colegiado de Ciência da Computação

Curso de Bacharelado em Ciência da Computação

**Utilização de um redes neurais *feedforward* com
backpropagation na modelagem de Distribuição de
Espécies**

Gean Carlo Peixoto

CASCABEL

2014

GEAN CARLO PEIXOTO

**UTILIZAÇÃO DE UMA REDE NEURAL *FEEDFORWARD* COM
BACKPROPAGATION NA MODELAGEM DE DISTRIBUIÇÃO DE
ESPÉCIES**

Monografia apresentada como requisito parcial para obtenção do grau de Bacharel em Ciência da Computação, do Centro de Ciências Exatas e Tecnológicas da Universidade Estadual do Oeste do Paraná - Campus de Cascavel

Orientador: Adair Santa Catarina

CASCADEL

2014

GEAN CARLO PEIXOTO

**UTILIZAÇÃO DE UMA REDE NEURAL *FEEDFORWARD* COM
BACKPROPAGATION NA MODELAGEM DE DISTRIBUIÇÃO DE
ESPÉCIES**

Monografia apresentada como requisito parcial para obtenção do Título de *Bacharel em Ciência da Computação*, pela Universidade Estadual do Oeste do Paraná, Campus de Cascavel, aprovada pela Comissão formada pelos professores:

Prof. Dr. Adair Santa Catarina
Colegiado de Ciência da Computação,
UNIOESTE

Prof. M. Eng. Carlos José Maria Olguín
Colegiado de Ciência da Computação,
UNIOESTE

Prof^a. Dr^a. Claudia Brandelero Rizzi
Colegiado de Ciência da Computação,
UNIOESTE

Cascavel, 08 de Dezembro de 2014.

“You can't process me with a normal brain” –
Charlie Sheen

AGRADECIMENTOS

Aos meus pais, Emilson e Maria, pela educação, amor e confiança a mim dedicado em todos os anos de minha vida. Em especial, ao meu pai, pelas incontáveis horas de trabalho que possibilitaram minha formação.

Aos meus irmãos, Marcos e Polyanna, por todos os momentos de diversão que passamos juntos.

À minha namorada, Jéssica, pela paciência, amor e carinho a mim destinado.

Aos meus professores e colegas de curso por sempre me auxiliarem.

Ao meu orientador, Adair, pela grande contribuição e dedicação que teve na realização deste trabalho.

Lista de Figuras

Figura 1.1 - Classificação dos modelos de distribuição.....	2
Figura 2.1 - O método de distâncias	9
Figura 2.2 - Espaço ROC com quatro modelos arbitrários.....	14
Figura 2.3 - Curva ROC.....	15
Figura 3.1 - Neurônio Biológico.....	18
Figura 3.2 - Neurônio Artificial.....	19
Figura 3.3 - Tipos de problemas	21
Figura 3.4 – <i>Multilayer Perceptron</i>	22
Figura 4.1 - Esquema geral de uma rede neural aplicada à modelagem de distribuição de espécies	26
Figura 4.2 - <i>Thalurania Furcata Boliviana</i> , o beija-flor tesoura verde.....	27
Figura 4.3 - <i>Strix Varia</i> , a coruja listrada	28

Lista de Tabelas

Tabela 2.1 - Matriz de confusão	11
Tabela 2.2 - Métricas derivadas da matriz de confusão.....	12
Tabela 5.1 - Resultados do Experimento 1 no Conjunto A	31
Tabela 5.2 - Resultados do Experimento 1 no Conjunto B	31
Tabela 5.3 - Resultados obtidos por Santa Catarina (2009)	32
Tabela 5.4 - Resultados do Experimento 2 no Conjunto A	32
Tabela 5.5 - Resultados do Experimento 2 no Conjunto B	32
Tabela 5.6 - Resultados do Experimento 3 no Conjunto A	33
Tabela 5.7 - Resultados do Experimento 3 no Conjunto B	33
Tabela 5.8 - Resultados do Experimento 4 no Conjunto A	33
Tabela 5.9 - Resultados do Experimento 5 no Conjunto A	34
Tabela 5.10 - Resultados do Experimento 5 no Conjunto B	34
Tabela A.1 - Dados originais da <i>Thalurania Furcata Boliviana</i>	41
Tabela A.2 – Dados resultantes de <i>Thalurania Furcata Boliviana</i>	46
Tabela A.3 – Dados originais da <i>Strix Varia</i>	54
Tabela A.4 - Dados resultantes de <i>Strix Varia</i>	58

Lista de Abreviaturas e Siglas

RNA	Rede Neural Artificial
MLP	<i>Multilayer Perceptron</i>
CCM	Coefficiente de Correlação de Matthews
ROC	<i>Receiver Operating Characteristic</i>

Lista de Símbolos

P_v	Positivo verdadeiro
P_F	Falso positivo
A_v	Ausência verdadeira
A_F	Falsa ausência

Sumário

1	Introdução.....	1
1.1	Motivações.....	3
1.2	Objetivos.....	4
1.3	Organização do Texto.....	5
2	Modelos de distribuição de espécie.....	6
2.1	O processo de modelagem.....	6
2.2	Algoritmos de modelagem.....	8
2.3	Avaliação dos modelos.....	10
3	Redes neurais artificiais.....	17
3.1	O neurônio.....	17
3.1.1	Neurônio biológico.....	18
3.1.2	Neurônio artificial.....	19
3.2	A rede neural.....	20
3.3	Mecanismos de aprendizagem.....	22
3.3.1	Aprendizado supervisionado.....	23
3.3.2	Aprendizado não-supervisionado.....	24
3.3.3	Aprendizado por reforço.....	24
4	Metodologia.....	25
4.1	Seleção da rede neural artificial.....	25
4.2	Base de dados utilizada.....	27
4.3	Métodos de análise.....	29
5	Análise dos resultados.....	31
6	Considerações finais.....	35
6.1	Trabalhos futuros.....	35
	Referências bibliográficas.....	36

Anexo A - A base de dados	39
---------------------------------	----

Resumo

Um dos principais desafios que a humanidade enfrenta atualmente é identificar, de maneira rápida e barata, as áreas mais importantes para a conservação da biodiversidade, de modo que auxilie na implantação de estratégias adequadas para o desenvolvimento sustentável. O uso de técnicas de modelagem de distribuição é indicado em situações onde é preciso tomar decisões mas não há disponível um conjunto significativo de informações. O processo de modelagem de distribuição de espécies pode ser visto como um problema de predição, uma vez que o objetivo principal é determinar a probabilidade de ocorrência de uma dada espécie biológica em cada ponto de um mapa georreferenciado. Vários algoritmos de modelagem podem ser utilizados nesta tarefa. Neste trabalho foi utilizado um algoritmo pertencente à categoria de aprendizado de máquina, as Redes Neurais Artificiais. Nesse contexto, o objetivo principal deste trabalho foi avaliar a eficácia das RNAs como método alternativo para gerar modelos de distribuição de espécies. Como estudo de caso, o método proposto foi avaliado utilizando dados obtidos de duas espécies, a *Strix Varia* e a *Thalurania Furcata Boliviana*. Além da validação do método, os testes tiveram como objetivo avaliar seu desempenho em diferentes aspectos como, por exemplo, com poucos exemplos de treinamento ou então nenhum ponto de ausência. Os resultados mostraram que o algoritmo baseado em RNAs apresentaram resultados semelhantes ao desempenho dos demais algoritmos e, portanto, é adequado para a tarefa de modelagem. Todavia, não foi possível estabelecer uma estratégia para auxiliar no processo de seleção do tamanho da camada escondida. Fato este que dificulta a implantação real do método, pois transfere ao usuário final a responsabilidade de decidir quantos neurônios serão utilizados.

Palavras-chave: Aprendizado de Máquina, Redes Neurais Artificiais, Modelos de Distribuição de Espécies.

1 Introdução

O desenvolvimento econômico, principalmente nas últimas décadas, tem acelerado o processo de destruição de ecossistemas primários, tornando crescente a demanda por respostas rápidas que visem solucionar problemas como a perda de biodiversidade, controle de espécies invasoras e de pragas agrícolas, dentre outros fatores (RODRIGUES, RODRIGUES e CORRÊA, 2010). Dito isso, a gestão dos recursos naturais se mostra um dos principais desafios, senão o principal, que a humanidade enfrenta atualmente e enfrentará nas próximas décadas (RODRIGUES, 2012).

Nesse contexto, um dos principais problemas enfrentados é identificar, de maneira rápida e barata, as áreas mais importantes para a conservação da biodiversidade, de modo que a implantação de estratégias adequadas para o desenvolvimento sustentável seja eficaz. Para tanto é necessário o desenvolvimento de processos que auxiliem o trabalho de produzir, trabalhar e utilizar a informação disponível (TÔRRES e VERCILLO, 2012).

Segundo Canhos (2003), no passado a base de conhecimento sobre a diversidade biológica era incipiente e desagregada, as informações existentes estavam dispersas e, via de regra, inacessíveis. Com o passar dos anos, avanços significativos ocorreram na definição de padrões e protocolos para integração de dados distribuídos, fundamentais para a construção da infraestrutura global de informação sobre biodiversidade. Contribuíram para isto o aumento na oferta de imagens de satélites em diferentes resoluções espaciais, espectrais e temporais, o uso de GPS para coleta direta de dados e o acesso a bases de dados demográficos mais detalhados (SANTA CATARINA, 2009). Como resultado, na última década os avanços em tecnologia possibilitaram a criação de um ambiente cooperativo, envolvendo o compartilhamento de conhecimentos, sem precedentes (SIQUEIRA, 2005).

O uso de técnicas de modelagem de distribuição é particularmente indicado em situações nas quais é preciso tomar decisões mas ainda não está disponível um conjunto significativo de informações, como é o caso de várias áreas dos principais biomas brasileiros, nas quais a coleta de dados ainda é bastante precária para o tamanho e riqueza biológica de cada região (SIQUEIRA, 2005). De forma geral, o resultado desse tipo de modelagem é um mapa de distribuição, de uma determinada região de estudos, que representa a probabilidade de ocorrência de uma determinada espécie, em cada um de seus pontos (RODRIGUES, 2012).

Há três pilares no estudo de modelos aplicados à ecologia, sendo eles: generalidade, realidade e precisão, ilustrados na Figura 1.1. Desses pilares são derivados três grupos de modelos, onde em cada grupo dois desses aspectos são focados em detrimento do terceiro, são eles: modelos empíricos, modelos mecanicistas e modelos analíticos (GUISAN e ZIMMERMANN, 2000).

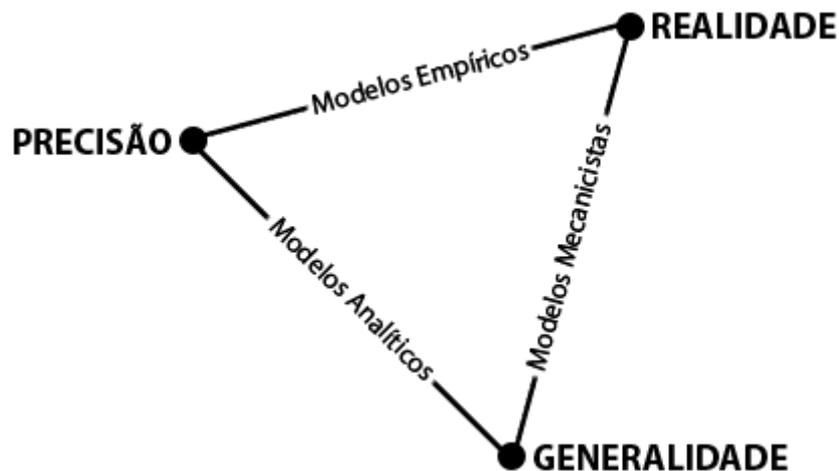


Figura 1.1 - Classificação dos modelos de distribuição

Fonte: Adaptação de Guisan e Zimmermann (2000)

Os modelos do primeiro grupo são chamados analíticos e focam a generalidade e a precisão, como as equações de crescimento populacional logístico, por exemplo. Os do segundo grupo são chamados mecanicistas, casuais ou modelos de processos; são desenvolvidos visando ser realistas e generalistas e suas previsões são baseadas nas relações de causa e efeito. Os modelos do terceiro grupo sacrificam a generalidade pela precisão e realidade; são chamados modelos empíricos, estatísticos ou fenomenológicos (SANTA CATARINA, 2009).

Dentre os tipos de modelos citados, é abordado nesse trabalho um modelo empírico para criação de Modelos da Distribuição de Espécies, do inglês *Species Distribution Models* (SDM). Sua formulação matemática não objetiva descrever as relações de causa e efeito entre os parâmetros do modelo e a resposta preditiva, nem descrever os mecanismos e funções ecológicas básicas, sendo seu principal objetivo aglutinar fatos empíricos (SANTA CATARINA, 2009). Os SDMs buscam encontrar relações não aleatórias entre os dados de ocorrência da espécie com variáveis ecológicas/ambientais relevantes para a espécie, como: temperatura, precipitação, topografia, tipo de solo, entre outros (SIQUEIRA, 2005).

A partir da década de 1990 foram desenvolvidas e utilizadas várias técnicas de modelagem de SDM com os mais variados objetivos. Algumas dessas técnicas envolvem modelagem baseada em análise ambiental, nas quais os algoritmos procuram por condições ambientais semelhantes àquelas onde as espécies foram encontradas, resultando em áreas potenciais onde as condições ambientais seriam propícias para o desenvolvimento dessas espécies. Esses métodos de modelagem podem ser divididos em dois tipos, os chamados modelos estatísticos e os modelos baseados em regras. No primeiro caso, temos os seguintes métodos: regressões múltiplas, modelos lineares, modelos aditivos, árvores de classificação. Como exemplo de modelos baseados em regras existem: algoritmos genéticos e redes neurais. Todos esses métodos buscam definir as limitações ambientais das espécies nas dimensões para as quais o modelo é desenvolvido (SIQUEIRA, 2005).

1.1 Motivações

Uma vez que os recursos humanos e financeiros para o meio ambiente geralmente são escassos frente à demanda (RODRIGUES, RODRIGUES e CORRÊA, 2010), as ferramentas de modelagem podem contribuir para a conservação ambiental, fornecendo elementos importantes em processos de tomada de decisão, direcionando investimentos e esforços para áreas de maior risco (SIQUEIRA, 2005). Seu uso pode auxiliar em diversos aspectos como, por exemplo, direcionar a pesquisa de campo, auxiliar planos de conservação da biodiversidade, identificar áreas de maior riqueza biológica, prever o efeito de mudanças climáticas globais, delimitar áreas ricas em espécies ameaçadas ou endêmicas e auxiliar na busca por novas populações de espécies conhecidas (RODRIGUES, 2012; SANTA CATARINA, 2009; SIQUEIRA 2005; CANHOS, 2003).

O uso das ferramentas de modelagem pode, ainda, indicar áreas de distribuição potencial para as espécies em risco de extinção. Neste caso, o modelo resultante determina os possíveis locais de ocorrência dessas espécies, isto é, auxilia na indicação de áreas onde existam condições ambientais favoráveis para a sobrevivência da espécie (SIQUEIRA, 2005). Tais regiões podem ser recomendadas como áreas prioritárias para a conservação ambiental, caso ela já ocupe o local, ou então definidas como áreas potenciais para a reintrodução desta espécie (TÔRRES e VERCILLO, 2012).

Como dados em escalas precisas de distribuição dos organismos são escassos e muito caros para serem adquiridos em quantidades suficientes e de maneira sistemática, é

necessário estabelecer estratégias para a conservação da biodiversidade, ainda que com dados incompletos, antes que as alterações geradas pelo desenvolvimento exterminem espécies (TÔRRES e VERCILLO, 2012). A fim de suprimir esta demanda, surgiram os modelos de distribuição potencial de espécies. Estes modelos predizem a distribuição espacial potencial, baseados nos requisitos ecológicos da espécie, extrapolando para áreas desconhecidas a partir de pontos conhecidos (SIQUEIRA, 2005). Entretanto, a eficácia destes modelos continua altamente dependente da qualidade e quantidade dos dados biológicos e ambientais disponíveis (TÔRRES e VERCILLO, 2012).

O processo de modelagem de distribuição geográfica de espécies biológicas pode ser visto como um problema de predição, uma vez que o objetivo principal é determinar a probabilidade da existência de uma determinada espécie biológica em cada ponto de um mapa georreferenciado (RODRIGUES, RODRIGUES e CORRÊA, 2010). Para geração dos modelos apresentados neste trabalho será utilizado um algoritmo pertencente à categoria de aprendizado de máquina, chamado de Redes Neurais Artificiais (RNA). As redes neurais podem, matematicamente, serem entendidas como modelos de inferência multidimensionais e não-lineares (VILLANUEVA, 2011).

O grande apelo das RNA está na sua capacidade de “aprender”, isto é, em sua habilidade de generalizar, ou extrair, automaticamente as regras que compõem um conjunto de dados complexos (GALO, FERREIRA e UTSUMI, 2013). Somando-se a esta capacidade os bons resultados obtidos pela técnica no reconhecimento de padrões (RODRIGUES, 2012), resolveu-se experimentar as RNA no processo de modelagem da distribuição potencial de espécies.

1.2 Objetivos

O objetivo geral deste trabalho consiste em verificar a eficácia de uma rede neural artificial *multilayer perceptron* como método alternativo para gerar modelos de distribuição de espécies. Para o treinamento da rede é utilizado o método de aprendizado supervisionado conhecido como *Backpropagation*.

Como objetivos secundários deste trabalho consistem em avaliar o desempenho do modelo proposto em diferentes aspectos: poucos, ou nenhum, exemplos negativos de treinamento; acurácia do modelo com diferentes quantidades de exemplos de treinamento; verificar se existe uma quantidade ideal de neurônios nas camadas intermediárias.

1.3 Organização do Texto

Além deste capítulo introdutório, o texto deste trabalho está organizado em outros 5 capítulos. O Capítulo 2 possui como objetivo apresenta os conceitos necessários para compreensão da modelagem de distribuição de frequência. O Capítulo 3 apresenta redes neurais artificiais, explicando seu funcionamento. No Capítulo 4 discute-se a metodologia, bases de dados utilizadas e detalhes sobre a implementação e experimentos realizados. A discussão dos resultados é feita no Capítulo 5. As considerações finais e os trabalhos futuros serão abordados no Capítulo 6.

2 Modelos de distribuição de espécie

O objetivo deste capítulo é apresentar os conceitos e técnicas necessários para a compreensão da modelagem de distribuição de espécies. Na primeira sessão está descrito o processo para a construção de um modelo de distribuição de espécies. Na segunda sessão estão listados os principais algoritmos para a modelagem de um SDM, juntamente com uma breve descrição dos mesmos. A última sessão apresenta os métodos utilizados na avaliação dos SDM.

2.1 O processo de modelagem

A distribuição de uma espécie é afetada por quatro fatores: fatores abióticos, ou seja, o conjunto de condições ambientais que limitam a capacidade de persistir em uma determinada região; conjunto de interações com outras espécies que interferem na capacidade de manter populações, chamado de fatores bióticos; fatores de acessibilidade, isto é, quais regiões estão acessíveis a partir de uma região original; os fatores evolucionários que, como o próprio nome diz, estão relacionados com a capacidade de adaptação às novas condições (SOBERÓN e PETERSON, 2005).

Dois conceitos são fundamentais para o entendimento do processo de modelagem de distribuição de espécies, o nicho ecológico e o nicho fundamental (RODRIGUES, 2012). O nicho ecológico é a união de todos os fatores citados anteriormente e pode ser descrito como sendo uma região n -dimensional, onde cada dimensão corresponde a uma condição ambiental ou recurso necessário para que uma determinada espécie possa sobreviver (HUTCHINSON, 1957). O nicho fundamental é um subdomínio do primeiro, onde apenas as condições ambientais são consideradas (SOBERÓN e PETERSON, 2005), sem considerar a influência de competição interespecífica ou predação por outras espécies (IWASHITA, 2008). Atualmente apenas o conceito de nicho fundamental tem sido comumente utilizado para produzir os mapas de distribuição de espécies (RODRIGUES, 2012). A principal razão para isto encontra-se na dificuldade de se obter variáveis que representam condições bióticas, cuja interpretação é complexa (SOBERÓN e PETERSON, 2005).

O processo de modelagem consiste na conversão de dados de ocorrência de espécies em mapas de distribuição geográfica indicando a provável presença ou ausência de espécies, através da aplicação de um algoritmo que busca encontrar relações não

aleatórias entre os dados de ocorrência das espécies com dados ecológicos/ambientais pertinentes a espécie (MATOS, 2010). Os estudos que envolvem modelagem da distribuição geográfica de espécies biológicas possuem três componentes básicos. O primeiro é um conjunto de dados de ocorrência e outro conjunto contendo variáveis ambientais que, unidos, formam um conjunto de pontos de nicho. Já o segundo é a técnica que será usada no processo de modelagem, cujo objetivo é relacionar a espécie com as variáveis ambientais. O último consiste na avaliação da utilidade do modelo através de validação ou por modelos de robustez (SANTA CATARINA, 2009).

Os dados de ocorrência são pontos georreferenciados (coordenadas geográficas), que representam a existência ou ausência da espécie naquele local. Os pontos de presença indicam a existência comprovada da espécie naquela coordenada, são obtidos por meio de coleta de campo (SIQUEIRA, 2005). Já os pontos de ausência representam a inexistência da espécie no local indicado. Entretanto, pontos de ausência raramente estão disponíveis e são difíceis de determinar, pois em um determinado local pode ser registrada a ausência da espécie por diferentes motivos: a espécie não pode ser detectada, embora presente; a sua ocorrência é sazonal (SANTA CATARINA, 2009); devido a fatos históricos a espécie está ausente, apesar das condições ambientais serem favoráveis; as condições são realmente inadequadas para a espécie (SIQUEIRA, 2005). Apesar de escassos, os pontos de ausência são informações necessárias na maioria dos algoritmos de modelagem para construir a representação de nicho fundamental. Visando contornar essa situação, normalmente adota-se alguma estratégia para obtenção de pontos de ausência, que passam a ser chamados de pontos de pseudoausência (RODRIGUES, 2012). A geração dos pontos de pseudoausência deve ser criteriosa a fim de reduzir possíveis ruídos, pois a escolha destes pontos reflete diretamente no resultado do modelo (FINAMORE, 2010).

Os fatores abióticos discretizados em camadas *raster* georreferenciadas constituem as variáveis ambientais (RODRIGUES, 2012) que, em geral, são constituídas de informações climáticas, topológicas e de uso e ocupação do solo (MATOS, 2010). As variáveis ambientais podem exercer efeitos diretos ou indiretos sobre a espécie, e devem ser escolhidas de modo a representar os principais fatores que atuam sobre a espécie (AUSTIN, 2002). Por exemplo, a temperatura pode não influenciar diretamente a fisiologia de uma espécie de planta, mas tem correlação com a luminosidade, esta sim determinante na viabilidade deste indivíduo. Atualmente é quase nulo o emprego de variáveis ambientais com

efeito direto sobre as espécies, e este é ainda um ponto crítico no processo de modelagem (IWASHITA, 2008).

A junção dos dados de ocorrência com os dados ambientais produz pontos de nicho, isto é, pontos no espaço ambiental. Eles representam os valores que as variáveis ambientais assumem em cada coordenada georreferenciada (RODRIGUES, 2012). No processo de modelagem de distribuição de espécies o objetivo é encontrar relações não aleatórias entre os dados de ocorrência da espécie e os dados ambientais relevantes para a espécie (SIQUEIRA, 2005). Para tanto, os modelos são executados com o objetivo de produzir uma aproximação do modelo de nicho, em outras palavras, buscam encontrar relações não aleatórias entre os pontos de nicho identificados. Esse modelo de nicho é, então, projetado sobre uma dada região geográfica, produzindo um mapa georreferenciado que contém, em cada coordenada, a probabilidade de ocorrência da espécie (RODRIGUES, 2012). Alguns dos principais algoritmos de modelagem serão apresentados na próxima seção.

O processo de modelagem de distribuição geográfica de espécies pode ser resumida em três etapas: primeiramente, um conjunto de pontos de ocorrência é combinado com um conjunto de variáveis ambientais, dando origem a um conjunto de pontos de nicho; em seguida um algoritmo de modelagem utiliza-se destes pontos para construir um modelo de nicho; finalmente, o modelo de nicho é projetado em uma área geográfica (RODRIGUES, 2012).

2.2 Algoritmos de modelagem

Como discutido na sessão anterior, a função do algoritmo de modelagem é executar o passo de inferência do processo de modelagem, isto é, ele busca extrapolar para as demais localidades as relações não aleatórias obtidas a partir dos pontos de nicho. Há uma grande variedade de formas de modelagem aplicadas ao problema de prever a distribuição de espécies, a seguir serão apresentadas algumas delas.

O método de distâncias é, provavelmente, o método mais simples da lógica de nicho ecológico. Este método assume o centroide dos pontos de ocorrência como sendo ótimo. A distância entre o ótimo estimado e os valores observados para cada célula da grade ambiental é inversamente proporcional à adequabilidade do ambiente naquele local (MARCO JR. e SIQUEIRA, 2009). Esta abordagem gera, claramente, um envelope circular ao redor do ponto ótimo no espaço ecológico, como mostra a Figura 2.1.



Figura 2.1 - O método de distâncias

Fonte: Mateo et al., 2010

Um dos mais antigos métodos para geração de modelos de distribuição de espécies, o envelope bioclimático, prediz locais com condições climáticas favoráveis baseado no cálculo retilíneo mínimo no espaço multidimensional (IWASHITA, 2008). Em outras palavras, para cada variável ambiental, o algoritmo, determina, com base em medidas estatísticas, um limite superior e inferior para a ocorrência da espécie (MARCO JR. e SIQUEIRA, 2009). O Bioclim é um algoritmo de modelagem que implementa o conceito de envelope bioclimático. Considerando o conjunto de pontos de ocorrência, o algoritmo calcula a média e o desvio padrão para cada variável ambiental. Segundo Rodrigues (2010) “o intervalo $[m - c * s, m + c * s]$ representa o envelope de cada variável ambiental”, onde m é a média, s é o desvio padrão e c é um limiar definido pelo usuário.

O Maxent, também conhecido como Entropia Máxima, inicia a lista dos modelos mais complexos. A ideia principal do método é encontrar, a partir de um conjunto de possíveis distribuições de probabilidade, aquela distribuição que tem a máxima entropia; este processo pode ser considerado um problema de otimização com restrições (RODRIGUES, 2012). Isso significa que o objetivo se baseia em elaborar uma aproximação onde sejam respeitadas todas as restrições conhecidas acerca da distribuição da espécie (IWASHITA, 2008).

O GARP (*Genetic Algorithm for Rule-set Production*) é um algoritmo baseado em aprendizado de máquina, e como o próprio nome sugere, utiliza-se de algoritmo genético para selecionar o conjunto de regras que melhor se adapta ao conjunto de pontos de nicho (MATOS, 2010). O GARP opera sobre o conjunto de regras, realizando uma “seleção

natural”, excluindo regras menos eficientes e criando um novo conjunto de regras a partir das sobreviventes (IWASHITA, 2008). Essas regras são definidas usando, como base, os dados ambientais somados aos dados de ocorrência da espécie (RODRIGUES, 2012).

As redes neurais artificiais são técnicas computacionais inspiradas em sistemas biológicos de aprendizado e são consideradas uma das técnicas de aprendizagem mais eficientes (RODRIGUES, 2012). Elas podem ser entendidas como dispositivos de processamento de informação caracterizados pela interconexão de unidades elementares de processamento, os neurônios artificiais, simples e similares entre si (VILLANUEVA, 2011). Mais informações sobre as redes neurais serão discutidas no terceiro capítulo deste trabalho.

Pertencente à família de classificadores generalizados, o SVM (*Support Vector Machine*) foi introduzido recentemente como uma técnica de reconhecimento de padrões. De acordo com a teoria de SVMs, enquanto técnicas tradicionais são baseadas na minimização do risco empírico, ou seja, buscam otimizar o desempenho sobre o conjunto de treinamento, as SVMs minimizam o risco estrutural, isto é, a probabilidade de classificar de forma incorreta os padrões ainda não vistos pela distribuição de probabilidade dos dados (MARCO JR. e SIQUEIRA, 2009).

2.3 Avaliação dos modelos

Como discutido nas sessões anteriores, a modelagem de distribuição de espécies pode ser vista como um problema de classificação binária, pois apenas duas classes são consideradas, presença ou ausência. Apesar de alguns algoritmos de modelagem serem baseados em técnicas de regressão, ou seja, produzem uma saída contínua ou probabilística, é comum a utilização de um limiar (*threshold*) para determinar a classe do resultado (RODRIGUES, 2012).

Uma maneira natural de apresentar estatísticas para avaliação de um modelo de classificação é por meio de uma tabulação cruzada entre a classe prevista pelo modelo e a classe real dos exemplos (PRATI, BATISTA e MONARD, 2008). Essa tabulação é conhecida como matriz de confusão, também chamada de tabela de contingência. Na modelagem de distribuição de espécies, a matriz de confusão é utilizada para auxiliar no processo de quantificar a qualidade dos modelos (SANTA CATARINA, 2009). Na Tabela 2.1 é apresentada uma matriz de confusão:

Tabela 2.1 - Matriz de confusão

Classe	Preditada	
	P	A
Observada P	P_V Presenças Verdadeiras	A_F Ausências Falsas
Observada A	P_F Presenças Falsas	A_V Ausências Verdadeiras

Fonte: Adaptado de Rodrigues (2012).

Os valores de presença verdadeira e ausência verdadeira localizadas na diagonal principal da matriz, representam os acertos na classificação. Os dados encontrados na diagonal secundária da matriz, ausência falsa e presença falsa, constituem as falhas de predição do método. O termo ausência falsa corresponde à quantidade de exemplos da classe de presença classificados incorretamente como ausência, este tipo de falha é chamado de erro de omissão. Os chamados de erros de comissão, ou superestimativa, são erros de presença falsa, ou seja, a quantidade de exemplos de ausência classificados como presença pelo SDM.

Os erros de comissão não são considerados graves, pois podem ocorrer devido a diversos fatores como, por exemplo, o fato das condições ambientais serem favoráveis, mas fatores topológicos ou biológicos impedem que a espécie ocupe a área. Já os erros de omissão são considerados graves, isto é, para um SDM é inadmissível que um local onde se sabe que a espécie é encontrada está sendo predito como ausente (SANTA CATARINA, 2009). Algumas outras métricas, derivadas da matriz de confusão, são apresentadas na Tabela 2.2.

Tabela 2.2 - Métricas derivadas da matriz de confusão

Métrica	Cálculo
Taxa de erro	$\frac{P_F + A_F}{P_V + P_F + A_V + A_F}$
Acurácia	$\frac{P_V + A_V}{P_V + P_F + A_V + A_F}$
Sensibilidade	$\frac{P_V}{P_V + A_F}$
Especificidade	$\frac{A_V}{P_F + A_F}$
Taxa de falso positivo (comissão)	$\frac{P_F}{P_F + A_V}$
Taxa de falso negativo (omissão)	$\frac{A_F}{P_V + A_F}$
Coefficiente de correlação de Mattheuws	$\frac{P_V * A_V + P_F * A_F}{\sqrt{(P_V + P_F) * (P_V + A_F) * (A_V + P_F) * (A_V + A_F)}}$

Fonte: Adaptado de Santa Catarina (2009).

Uma das maneiras mais comuns de avaliar modelos são a derivação de medidas que, de alguma maneira, buscam medir a qualidade do modelo. Essas medidas, geralmente, podem ser obtidas a partir da matriz de confusão, reduzindo suas quatro células principais a um único índice numérico de qualidade. Duas medidas de desempenho, amplamente utilizadas em problemas de classificação, são adotadas para a avaliação de modelos de distribuição de espécies: a taxa de erros e a acurácia. A primeira, também chamada de taxa de classificação incorreta, corresponde ao percentual de pontos de localização classificados de maneira incorreta; ela pode se especializar na taxa de erro da classe, que corresponde ao percentual de classificações incorretas relacionados a uma determinada classe, falso positivo, por exemplo (RODRIGUES, 2012).

A acurácia mede o acerto global do modelo, isto é, porcentagem de vezes que os pontos foram classificados de maneira correta (SANTA CATARINA, 2009). A sensibilidade é uma medida que descreve a probabilidade de uma amostra ser corretamente classificada como presença (SANTA CATARINA, 2009). Especificidade é o inverso da sensibilidade, isto é, a probabilidade de uma amostra ser classificada como ausência, de maneira correta (RODRIGUES, 2012).

O coeficiente de correlação de Mattheuws (CCM) é uma medida de qualidade em classificações binárias. Ele considera todas as informações advindas da matriz de confusão e é uma medida balanceada utilizada mesmo que as classes possuam diferentes tamanhos. Sua imagem está contida no intervalo $[-1,1]$ onde 1 corresponde à predição perfeita e -1 significa uma predição inversa (SANTA CATARINA, 2009).

A análise ROC (*Receiver Operating Characteristic*) é um método gráfico para avaliação, organização e seleção de sistemas de diagnóstico e/ou predição (PRATI, BATISTA e MONARD, 2008). O gráfico ROC é plotado em um espaço denominado espaço ROC e permite uma avaliação visual de aspectos relacionados com o desempenho das técnicas de classificação. No contexto da modelagem de distribuição de espécies, o espaço ROC bidimensional é definido através da taxa de verdadeiros positivos (P_V), plotada no eixo das ordenadas (eixo Y), e da taxa de falsos positivos (P_F), plotada no eixo das abscissas (eixo X). Em outras palavras, podemos dizer que o eixo X representa a especificidade e o eixo Y a sensibilidade (RODRIGUES, 2012).

Um modelo de classificação pode ser representado no espaço ROC por um simples ponto dado (1 - especificidade, sensibilidade), calculado a partir da matriz de confusão. Quanto mais próximo do ponto (0,1) estiver um modelo, melhor ele será, uma vez que este ponto representa uma classificação perfeita. Por outro lado, modelos que frequentemente classificam incorretamente, serão plotados próximo ao ponto (1, 0) (RODRIGUES, 2012). Enquanto modelos próximos ao canto inferior esquerdo (0, 0) podem ser considerados conservativos – fazem uma classificação positiva somente se têm grande segurança e, como consequência, possuem baixa taxa de falsos positivos – os modelos próximos ao ponto (1, 1) podem ser classificados como liberais – classificam a maior parte dos exemplos como positivos, mas, geralmente, com altas taxas de falsos positivos (PRATI, BATISTA e MONARD, 2008).

Modelos cujas predições são aleatórias estarão representados por pontos próximos a linha diagonal ($x = y$) que vai do ponto (0, 0) ao ponto (1, 1). Modelos posicionados acima desta diagonal apresentam desempenho melhor que modelos preditivos. Ao contrário, modelos posicionados abaixo dela possuem desempenho inferior comparados aos modelos aleatórios (RODRIGUES, 2012).

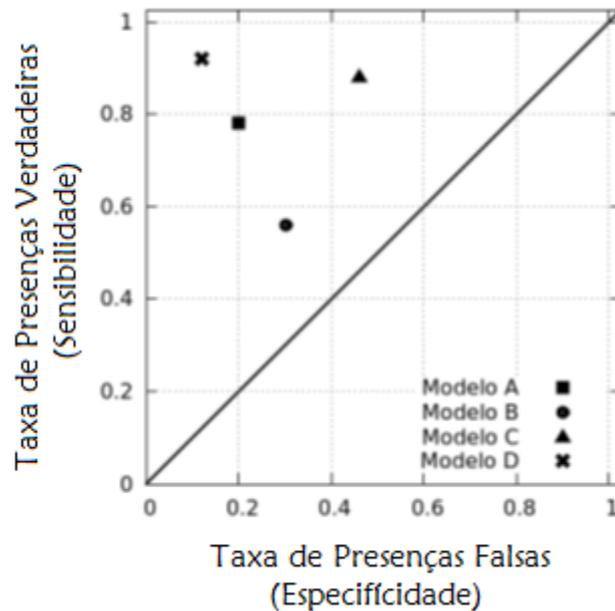


Figura 2.2 - Espaço ROC com quatro modelos arbitrários

Fonte: Adaptado de Rodrigues (2012)

A Figura 2.2 ilustra um espaço ROC com quatro modelos arbitrários. O Modelo D é o que apresenta melhor desempenho, devido à proximidade com o ponto (0, 1); já o modelo B possui resultados muito próximos à aleatoriedade.

Alguns modelos de distribuição de frequência produzem uma saída contínua ao invés de uma saída binária, então por que não analisá-los de maneira diferenciada? Para diminuir os erros de análise intrínsecos ao processo de limiarização, a curva ROC, ao invés de utilizar um único limiar, passa a “simular” a escolha de vários limiares. Neste caso, varia-se o limiar em todo o seu espectro, desde o valor mais restritivo até o valor mais liberal. Dessa maneira, a análise é feita independentemente da escolha de um limiar e o desempenho do sistema é, então, representado por uma curva no espaço ROC – a curva ROC (PRATI, BATISTA e MONARD, 2008).

Para construir a curva ROC é necessário, primeiramente, ordenar todas as saídas obtidas pelo modelo e, então, para cada exemplo do conjunto de teste, partindo do ponto (0, 0), dá-se um passo no espaço ROC de acordo com as seguintes regras: caso o exemplo pertença à classe positiva (presença), um passo de tamanho $\frac{1}{P}$ é dado na direção do eixo Y, onde P é o número total de exemplos da classe presença; caso contrário, o exemplo pertence à classe negativa (ausência), um passo de tamanho $\frac{1}{A}$ é dado na direção do eixo X, onde A é o número total de exemplo da classe ausência. Cada passo t deve ser dado a

partir de $t-1$, produzindo, assim, uma curva (PRATI, BATISTA e MONARD, 2008). Na Figura 2.3 é mostrado um exemplo hipotético de curva ROC.

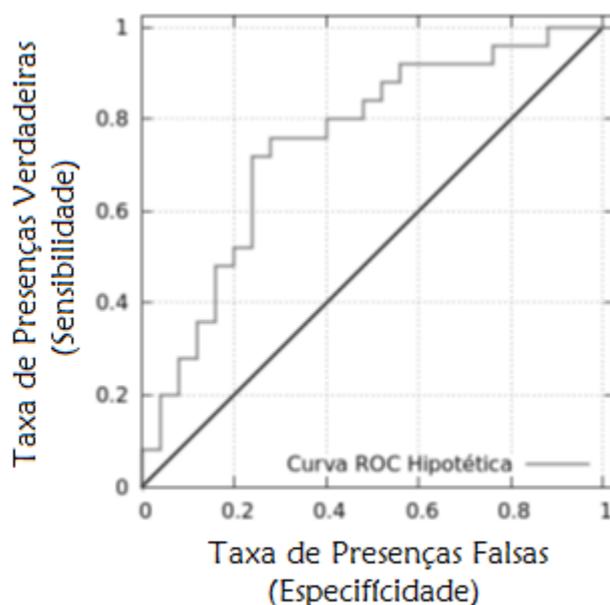


Figura 2.3 - Curva ROC

Fonte: Rodrigues (2012)

Quanto mais distante uma curva ROC estiver da linha diagonal ($x = y$), melhor será o desempenho do modelo. Desse modo é possível compara duas ou mais curvas ROC em um determinado domínio. Neste caso, o modelo com melhor desempenho será aquele cuja curva mais se aproximar do ponto ótimo, o ponto (0, 1) (RODRIGUES, 2012).

As curvas ROC possibilitam avaliar e comparar visualmente o desempenho de dois ou mais modelos de classificação, em um certo domínio. No entanto, é possível simplificar a curva ROC a um valor escalar, tornando mais direta a avaliação e a comparação entre modelos. Uma vez obtida a curva de um modelo no espaço ROC, é possível calcular a área sob esta curva, denominada área sob a curva ROC ou, simplesmente, AUC (*Area Under the ROC Curve*) (RODRIGUES, 2012). Deste modo, a AUC representa um valor que pode ser visto como uma medida simples de um desempenho global do modelo.

Uma vez que a área abaixo da curva ROC é uma fração da área de um quadrado de lado um, o valor do AUC sempre irá variar entre 0 e 1 (PRATI, BATISTA e MONARD, 2008). Sendo que, um AUC próximo a 0,5 equivale a uma predição aleatória (SANTA CATARINA, 2009). Vale lembrar que, ao comparar dois modelos de classificação

através dos valores de AUC, o modelo com melhor desempenho, ou seja, com o maior valor, pode apresentar desempenho inferior em uma região específica do espaço ROC.

A análise ROC é, atualmente, a principal técnica aplicada na avaliação de modelos de distribuição de espécies, nesse contexto, a AUC é uma medida que recebe destaque. No entanto, é importante que o desempenho de um modelo não seja reduzido a uma medida apenas, o que pode contribuir para a obtenção de conclusões equivocadas.

3 Redes neurais artificiais

Inspiradas no sistema nervoso humano, as redes neurais artificiais, possuem a capacidade de aquisição e manutenção do conhecimento e podem ser definidas como um conjunto de unidades de processamento, caracterizadas por neurônios artificiais, que são interligados por um grande número de interconexões (SILVA, SPATTI e FLAUZINO, 2010). Uma rede neural artificial não é simplesmente um sistema complexo, mas sim um sistema adaptativo complexo, isto significa que ela pode alterar suas estruturas internas baseada nas informações que trafegam por ela (SHIFFMAN, 2012).

O principal atrativo para a utilização de redes neurais está, sem dúvida, relacionado com sua capacidade de aprender, através de exemplos, e generalizar a informação aprendida. A generalização – capacidade da rede aprender através de um conjunto reduzido de exemplos e, em seguida, dar respostas coerentes para dados não-conhecidos – é uma demonstração de que a capacidade das RNAs vai muito além de, simplesmente, mapear relações de entrada e saída (BRAGA, CARVALHO e LUDEMIR, 2000).

Este capítulo possui como objetivo fornecer um embasamento teórico a respeito de redes neurais artificiais, principalmente da Perceptron de Multicamadas. Nele serão abordados desde conhecimentos básicos até conhecimentos específicos utilizados neste trabalho. A primeira sessão fala sobre o neurônio, tanto o biológico quanto o artificial. A Sessão 3.2 fala brevemente sobre a topologia em redes neurais artificias. Os métodos de treinamentos das RNAs são discutidos na sessão 3.3.

3.1 O neurônio

Um cérebro pode ser descrito como sendo uma rede neural biológica, isto é, uma rede de neurônios interligados transmitindo elaborados padrões elétricos (SHIFFMAN, 2012). O neurônio recebe sinais de entrada, processa estes sinais e, então, envia para os próximos neurônios seu sinal de saída. Este processo ocorre, tanto nos neurônios biológicos, quanto no artificial (CORRÊA e PORTUGAL, 1998). Note que o funcionamento individual do neurônio é simples, logo a riqueza computacional do cérebro humano está associada ao grande número de neurônios interconectados por uma rede complexa de sinapses (PORTUGAL e FERNANDES, 1996).

3.1.1 Neurônio biológico

A célula elementar do sistema nervoso cerebral é o neurônio e seu papel se resume a conduzir impulsos – estímulos elétricos advindos de relações físico-químicas – sob determinadas condições de operação. Tal elemento biológico pode ser dividido em três partes principais: dendritos, corpo celular e axônio (SILVA, SPATTI e FLAUZINO, 2010). O corpo do neurônio mede apenas alguns milésimos de milímetros, e os dendritos apresentam poucos milímetros de comprimento. O axônio, contudo, pode ser mais longo e, em geral, tem espessura uniforme (BRAGA, CARVALHO e LUDEMIR, 2000). A Figura 3.1 ilustra, de maneira simplificada, os componentes de um neurônio biológico.

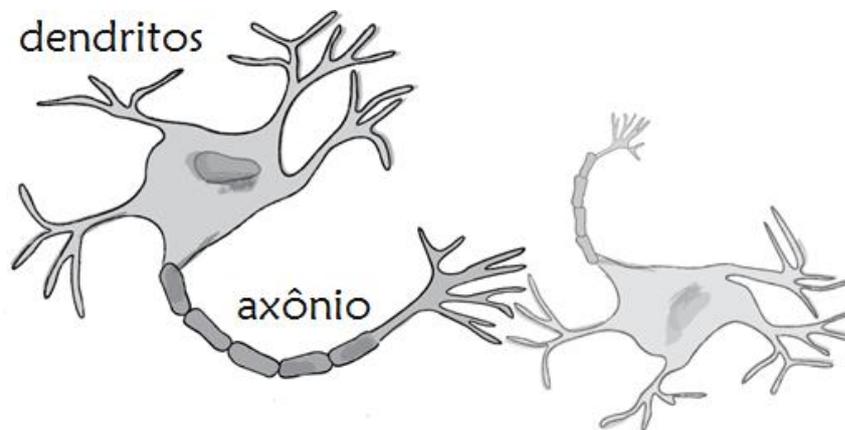


Figura 3.1 - Neurônio Biológico

Fonte: Adaptação de Shiffman (2012)

Os dendritos têm, por função, receber as informações, ou impulsos nervosos, oriundos de outros neurônios e conduzi-los até o corpo celular. O corpo do neurônio, também chamado de soma, é responsável por coletar e combinar informações vindas de outros neurônios e, então, gerar novos estímulos. Finalmente, o axônio é responsável por transmitir os estímulos para outras células (LUDWING JR e MONTGOMERY, 2007). A união de um axônio com um dendrito de outro neurônio é chamada de sinapse; elas funcionam como válvulas e são capazes de controlar a transmissão de impulsos – isto é, o fluxo da informação – entre os nodos na rede neural. Os efeitos das sinapses são variáveis, e é esta variação que dá ao neurônio capacidade de adaptação (BRAGA, CARVALHO e LUDEMIR, 2000).

Estima-se que a quantidade de neurônios existentes no cérebro humano esteja na casa dos bilhões. Contudo, a velocidade de processamento de cada neurônio é relativamente baixa, quando comparado aos computadores tradicionais. Esta deficiência é superada pela

imensa quantidade de neurônios existentes operando de forma paralela (PORTUGAL e FERNANDES, 1996). Tais características permitem ao cérebro humano executar rapidamente algumas funções – como, por exemplo, reconhecimento de imagens – que computadores convencionais não conseguem executar com o mesmo desempenho.

3.1.2 Neurônio artificial

O neurônio artificial é uma simplificação do neurônio biológico. Sua descrição matemática resultou em um modelo com vários terminais de entrada e apenas um terminal de saída. Enquanto que, em sistemas biológicos, não existe mecanismo algum que sincroniza as ações dos nodos, no sistema artificial considera-se que os nodos em cada camada da rede disparam sincronamente, ou seja, todos eles são avaliados no mesmo instante. Sabe-se também que o valor da próxima saída, em neurônios biológicos, depende enormemente das ativações de estados anteriores (BRAGA, CARVALHO e LUDEMIR, 2000).

Para emular o comportamento das sinapses, os terminais de entrada do neurônio (x_1, x_2, \dots, x_n) têm pesos acoplados (w_1, w_2, \dots, w_3), cujos valores podem ser positivos ou negativos. O corpo do neurônio, por sua vez, é emulado por um mecanismo simples, chamado soma ponderada, isto é, todos os valores de entrada são multiplicados pelos seus respectivos pesos sinápticos e, então, somados. A saída do neurônio (y) é obtida através da aplicação de uma função de ativação (f) no resultado da soma (Σ) (BRAGA, CARVALHO e LUDEMIR, 2000). Uma descrição do modelo está ilustrada na Figura 3.2.

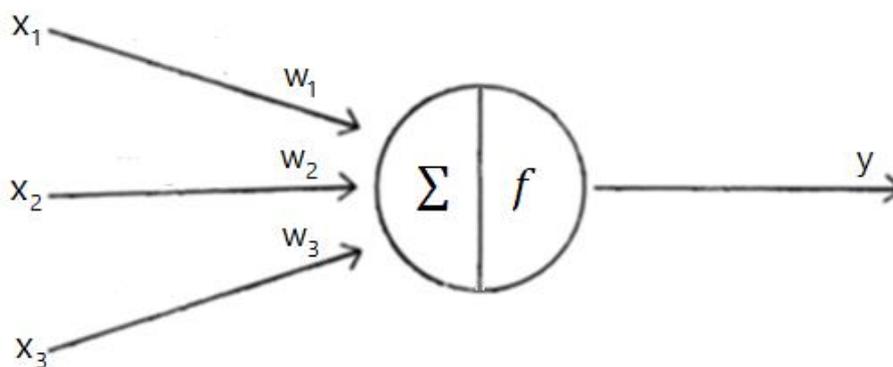


Figura 3.2 - Neurônio Artificial
Fonte: Adaptação de Shiffman (2012)

A função de ativação, também chamada de função de transferência, é uma função matemática que, aplicada à combinação linear entre as variáveis de entrada e pesos sinápticos, retorna seu valor de saída. Ela é utilizada para restringir a amplitude de saída de um neurônio (SILVA, 2005). Existem diversas funções matemáticas que são utilizadas como funções de ativação, dentre elas, as mais comumente utilizadas são a função de ativação logística e a função de ativação hiperbólica (SANTOS, SEIXAS, *et al.*, 2005).

3.2 A rede neural

Uma Rede Neural Artificial é composta por certo número de neurônios artificiais organizados em diferentes camadas. A primeira camada, chamada usualmente de camada de entrada, não executa qualquer computação, serve apenas para proporcionar a entrada de dados. O processamento ocorre somente nas camadas intermediárias – usualmente chamadas de camadas ocultas – e na camada de saída da rede (CORRÊA e PORTUGAL, 1998).

Em termos de topologia, para implementar uma rede neural multicamadas devem ser determinadas, a priori, as seguintes variáveis: o número de nós da camada de entrada, o número de camadas escondidas e, obviamente, o número de neurônio em cada uma delas e o número de neurônios na camada de saída (RODRIGUES, 2012). Os arranjos estruturais dos neurônios formam grafos orientados, onde o fluxo de sinais depende do número de camadas e o tipo de interligação entre elas (SILVA, 2005).

Não há uma regra que defina o número exato de neurônios necessários nas camadas escondidas. Em geral, redes com poucos neurônios são preferidas, visto que elas tendem a possuir um melhor poder de generalização, reduzindo o problema de super-ajuste (SANTOS, SEIXAS, *et al.*, 2005). Quanto maior o número de neurônios, maior a complexidade da rede e maior a sua abrangência em termos de soluções possíveis, isto é, redes com poucos neurônios escondidos podem não possuir habilidade suficiente para modelar e aprender os dados em problemas complexos, resultando em uma não convergência da rede (ANDRADE, 2011). Mais informações a respeito de erros de aprendizado são apresentadas na sessão 3.3. O número de camadas de uma rede neural multicamadas influencia, diretamente, a quantidade de problemas que a rede pode resolver. Redes com uma única camada resolvem apenas problemas linearmente separáveis. Ao adicionar, ao menos, uma camada escondida, a rede passa a resolver

qualquer tipo de problema contínuo. A utilização de duas camadas escondidas permite a aproximação de qualquer função (BRAGA, CARVALHO e LUDEMIR, 2000).

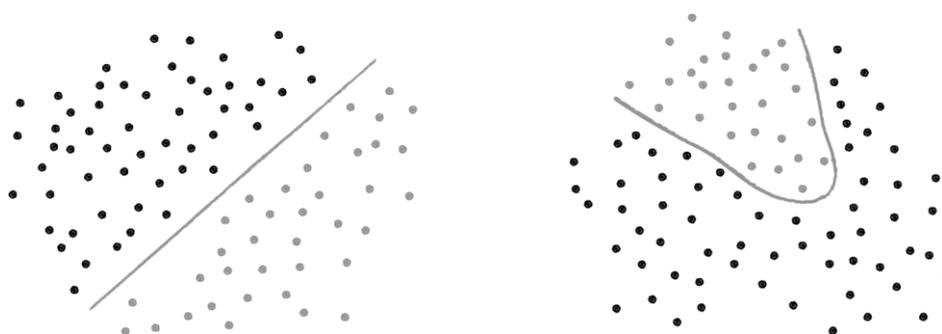


Figura 3.3 - Tipos de problemas

Fonte: Shiffman (2012)

A esquerda da Figura 3.3 existe um clássico problema linearmente separável, a saber: “Desenhe uma linha que separe dois conjuntos de dados”. Caso consiga separar os dados com uma única linha, este é um problema linearmente separável. À direita, no entanto, há um problema não linearmente separável – impossível de separar os dados desenhando uma única linha. O exemplo mais simples de um problema não linearmente separável é a porta lógica XOR, o “ou exclusivo” (SHIFFMAN, 2012).

As redes podem ser classificadas quanto ao tipo de conexão entre os nodos. Quando as conexões entre os neurônios formam um grafo acíclico – neurônios de uma camada estão conectados a, apenas, neurônios da camada imediatamente posterior, não havendo realimentação – a rede é chamada de *feedforward* (SANTOS, SEIXAS, *et al.*, 2005). Contudo, é possível conceber redes onde suas conexões formam ciclos, realimentando sua própria camada, ou camadas anteriores; redes deste tipo são conhecidas como *feedback*, usualmente chamadas de redes recorrentes. Segundo Silva (2005, p. 16), “quando todas as ligações da rede formam ciclos então a rede é dita auto associativa, sendo usada para recuperação de um padrão de entrada”.

Uma das redes neurais mais disseminadas na literatura, a rede *Multilayer Perceptron* (MLP) é composta por uma camada de entrada, composta por unidades sensoriais; uma ou mais camadas de neurônios ocultas; e uma camada de neurônios de saída. Como essas redes possuem, ao menos, uma camada intermediária de processamento, são capazes de aproximar funções complexas e não lineares, além de classificar padrões não linearmente separáveis (SILVA, 2005). Além disso, caracteristicamente, essas redes são completamente conectadas, isto é, todos os nós de uma camada são conectados a todos os

nós da camada subjacente (SANTOS, SEIXAS, *et al.*, 2005). Um exemplo de rede MLP pode ser observado na **Erro! Fonte de referência não encontrada.**, note que ela possui apenas uma camada escondida (*hidden layer*).

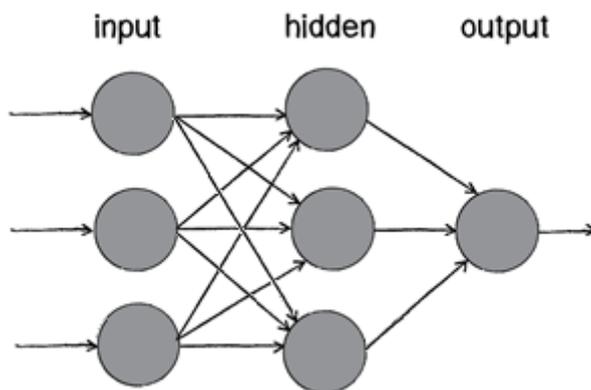


Figura 3.4 – *Multilayer Perceptron*

Fonte: Shiffman (2012)

3.3 Mecanismos de aprendizagem

As Redes Neurais Artificiais, ao contrário de outras técnicas de Inteligência Artificial, armazenam o conhecimento de forma não compreensível ao ser humano. De modo que, segundo Rezende (2005, p. 92), elas enquadram-se nos sistemas tipo caixa-preta: “desenvolvem sua própria representação do conceito, isto é, sua representação pode não ser facilmente interpretada por humanos e não fornecem esclarecimento, nem explicação do processo de reconhecimento”.

Consequentemente, a principal questão para operacionalizar uma RNA é obter uma maneira de ajustar as conexões, ou pesos sinápticos, de modo que a rede obtenha uma boa capacidade de resposta para o maior número de situações possíveis (CORRÊA e PORTUGAL, 1998). Análogo ao caso biológico – onde os estímulos recebidos promovem modificações nas intensidades das sinapses – os mecanismos de aprendizado são responsáveis por ajustar os pesos sinápticos (VILLANUEVA, 2006). O processo de aprendizado é iterativo e por meio dele a RNA deve melhorar seu desempenho, gradativamente, à medida que interage com o meio externo (REZENDE, 2005).

Durante o processo de aprendizagem, o fundamental é definir como seus pesos serão alterados. A fim de cumprir este objeto surgiram vários tipos de algoritmo de treinamentos específicos para determinados tipos de redes neurais, que diferem entre si principalmente pelo modo como os pesos são modificados (RODRIGUES, 2012). Nas

subseções seguintes serão abordados os três principais métodos de aprendizado: o aprendizado supervisionado, o aprendizado não supervisionado e o aprendizado por reforço.

Alguns problemas podem ocorrer durante o processo de aprendizado, tais como o *underfitting* (sub-ajuste) e o *overfitting* (super-ajuste). No primeiro caso, o *underfitting*, a rede não conseguiu convergir, ou uma amostra muito pouco representativa é assimilada pela rede em detrimento de amostras mais representativas. Já o *overfitting* ocorre quando há consideração excessiva do ruído presente nas amostras ou, simplesmente, uma amostra anômala acaba desviando o classificador de forma que ele passe a considerar uma extensão maior que a ideal.

3.3.1 Aprendizado supervisionado

O método de aprendizado mais comum é chamado de aprendizado supervisionado. Tem este nome porque a entrada e saída desejadas são fornecidas por um supervisor (professor) externo. O professor indica, explicitamente, um comportamento bom ou ruim para a rede, visando direcionar o processo de treinamento (BRAGA, CARVALHO e LUDEMIR, 2000). Normalmente, o treinamento supervisionado de uma rede neural consiste em um problema de minimização não linear sem restrições, em que os pesos sinápticos da rede são interativamente modificados para minimizar o erro médio quadrático entre a resposta desejada a partir dos dados de entrada, e a saída pelos neurônios de saída (SANTOS, SEIXAS, *et al.*, 2005).

A desvantagem do aprendizado supervisionado é que, na ausência de um professor, a rede não conseguirá aprender novas estratégias para situações não cobertas pelos exemplos de treinamento. Os exemplos mais conhecidos de algoritmos de aprendizado supervisionado são a regra delta e sua generalização para redes de múltiplas camadas, o algoritmo *backpropagation* (BRAGA, CARVALHO e LUDEMIR, 2000).

O processo de treinamento pelo algoritmo de retropropagação de erros, o *backpropagation*, inicia com um conjunto arbitrário de pesos – comumente utiliza-se um conjunto aleatório de pesos – e envolve duas fases bem definidas. Na primeira etapa um conjunto de entrada é apresentado à rede, que computará uma saída. Em seguida, a saída obtida é comparada com a saída esperada e, a partir disso, é obtido o erro cometido pela rede. A segunda fase do processo é o passo reverso, onde o erro é repassado – da camada de saída, para a de entrada – e as conexões serão atualizadas com base neste erro (ANDRADE, 2011).

3.3.2 Aprendizado não-supervisionado

No aprendizado não supervisionado, como o próprio nome sugere, não há um professor ou supervisor para acompanhar o processo de aprendizado. Apesar da semelhança entre o aprendizado supervisionado e o aprendizado dos seres humanos, muitos dos sistemas biológicos ocorrem através de aprendizado não-supervisionado, como, por exemplo, os estágios iniciais dos sistemas de visão e audição (BRAGA, CARVALHO e LUDEMIR, 2000).

Para estes algoritmos, somente os padrões de entrada estão disponíveis para a rede, ao contrário dos pares de entrada e saída dos algoritmos de aprendizado supervisionado. A partir do momento em que a rede estabelece uma harmonia com as regularidades estatísticas da entrada de dados, desenvolve-se nela uma habilidade de formar representações internas para codificar características de entrada e criar novas classes ou grupos automaticamente. Este tipo de aprendizado só se torna possível quando existe redundância nos dados de entrada. Sem redundância é impossível de encontrar quaisquer padrões ou características (BRAGA, CARVALHO e LUDEMIR, 2000).

3.3.3 Aprendizado por reforço

O aprendizado por reforço pode ser visto como um caso particular de aprendizado supervisionado. A principal diferença entre o aprendizado supervisionado clássico e o aprendizado por reforço é a medida de desempenho usado em cada um dos sistemas. Enquanto no aprendizado supervisionado comum a medida de desempenho é baseada na diferença entre o resultado esperado e a saída obtida pela rede, no aprendizado por reforço a única informação de realimentação fornecida é a classificação da saída, isto é, se a saída obtida é boa ou ruim.

Imagine um pequeno rato correndo dentro de um labirinto. Caso ele vire a esquerda, ele obtém um pedaço de queijo; do contrário, se virar para direita, ele recebe um pequeno choque. Presumidamente, o rato irá aprender a sempre virar à esquerda. Sua rede neural toma uma decisão e observa seu resultado. Caso o resultado seja ruim, a rede ajusta seus pesos de maneira que uma decisão diferente seja tomada da próxima vez (SHIFFMAN, 2012).

4 Metodologia

Este capítulo apresenta a metodologia utilizada para a execução dos experimentos. Ele é dividido em 3 sessões: a primeira sessão explica qual arquitetura foi utilizada para implementação do trabalho¹ e o porquê de sua escolha; a segunda sessão fala sobre o conjunto de dados utilizado; a terceira sessão fala sobre como foram executados os testes e quais são seus objetivos.

4.1 Seleção da rede neural artificial

Diferentes topologias de redes neurais são encontradas na literatura. Para cumprir o objetivo desse trabalho foi adotada a rede neural *Multilayer Perceptron* utilizando, como algoritmo de treinamento, o *Backpropagation*. O fato de ser uma rede clássica, amplamente utilizada e adequada tanto a tarefa de predição quanto classificação contribuíram significativamente para a escolha. Além disso, a maneira como os dados estão dispostos – pontos de nicho, descritos na Sessão 2.1 – facilitaram o processo de obtenção do par (entrada-saída) utilizado pelo algoritmo de treinamento supervisionado.

A topologia é um parâmetro fundamental a ser estabelecido na definição de uma MLP. Como dito anteriormente, o próprio problema auxilia na escolha da quantidade de neurônios nas camadas de entrada e saída. Na camada de entrada, a rede irá receber um vetor de dados, onde cada elemento é correspondente ao valor assumido por uma coordenada geográfica em uma determinada variável ambiental. Deste modo, podemos assumir que o número de neurônios na camada de entrada é igual ao número de *layers* geográficas que compõe o espaço de características e é definido automaticamente. Como saída, espera-se da rede um valor equivalente à probabilidade da presença da espécie no local informado, ou seja, um único valor de saída e, conseqüentemente, um único neurônio na camada de saída.

¹ Todo o código escrito durante o desenvolvimento do trabalho pode ser obtido através do link: <https://code.google.com/p/tcc-nightmare/>

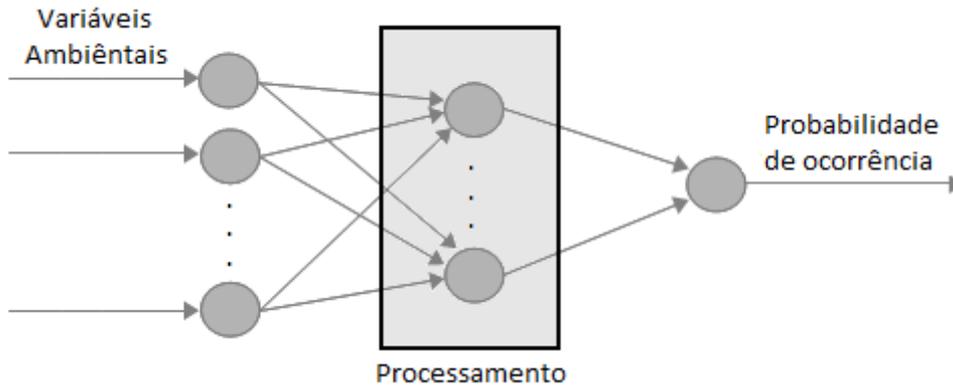


Figura 4.1 - Esquema geral de uma rede neural aplicada à modelagem de distribuição de espécies

Fonte: Adaptação de Rodrigues (2012)

Determinar o número de neurônios presente nas camadas escondidas – problema clássico das redes neurais artificiais – é, normalmente, resolvido de forma empírica. Como o número de neurônios nas camadas escondidas depende do número de camadas de entrada – varia, também, de acordo com a espécie estudada – caberá ao usuário do sistema informar a quantidade de neurônios nas camadas escondidas. A Figura 4.1 mostra um esquema geral da arquitetura de uma rede MLP aplicada à modelagem de distribuição de espécies.

Em um neurônio pertencente a uma rede MLP o sinal de ativação é obtido através de uma função não linear que recebe a soma ponderada dos valores de entrada (Σ), a função de ativação. Foi adotada a função logística, pertencente à classe sigmoideal, para este propósito. A Equação (4.1) representa a função de ativação tomada.

$$y = \frac{1}{1 + e^{-\Sigma}} \quad (4.1)$$

A estratégia de treinamento adotada foi o treinamento supervisionado, através do algoritmo de retro propagação do erro. Resumidamente, o *backpropagation* busca ajustar as conexões de forma a minimizar os erros de previsão dentro do período amostral onde, em cada interação, os pesos são revisados de acordo com a inclinação da superfície de erros, tentando atingir o mínimo global. O Ajuste de pesos no *backpropagation* é feito pela Equação (4.2), onde: w_{ji} é o peso do j -ésimo neurônio para a i -ésima entrada no instante t ; $w_{ji}(t + 1)$ é o novo valor para o peso; η é a taxa de aprendizagem; $\delta(t)$ é o erro cometido pelo neurônio e $x_i(t)$ é a i -ésima entrada.

$$w_{ji}(t + 1) = w_{ji} + \eta \delta(t) x_i(t) \quad (4.2)$$

No entanto, optou-se por utilizar uma variação da regra original de ajustes de pesos, incluindo o termo *momentum*. Este termo é utilizado para evitar a instabilidade da rede quando é usada uma taxa de aprendizagem alta (RODRIGUES, 2012). O termo *momentum* é representado pela equação:

$$\psi = \alpha (w_{ji}(t) + w_{ji}(t - 1)) \quad (4.3)$$

A formula completa de ajusto de pesos utilizando o termo *momentum* passa então a ser representada pela Equação 4.4. Neste trabalho optou-se por utilizar valores fixos para a taxa de aprendizagem (α) e o termo *momentum* (ψ), sendo eles 0.3 e 0.1 respectivamente.

$$w_{ji}(t + 1) = w_{ji} + \eta \delta(t) x_i(t) + \alpha (w_{ji}(t) + w_{ji}(t - 1)) \quad (4.4)$$

4.2 Base de dados utilizada

Durante a realização deste trabalho foi aproveitada a mesma base de dados utilizada por Santa Catarina (2009) e consiste em dados obtidos de duas espécies: *Strix Varia* e *Thalurna Furcata Boliviana*. Estes conjuntos foram obtidos, primeiramente, junto ao instalador do *software* DesktopGarp (Biodiversity Institute, 2007).



Figura 4.2 - *Thalurnia Furcata Boliviana*, o beija-flor tesoura verde

Fonte: Bartley (2009)

Um exemplar do *Thalurnia Furcata Boliviana*, conhecido popularmente como beija-flor tesoura verde, é mostrado na Figura 4.2. A base de dados do beija-flor, aqui chamado

de Conjunto A, contém 65 pontos de presença e 8 variáveis ambientais: precipitação acumulada no trimestre mais úmido, precipitação acumulada no trimestre mais quente, precipitação anual, temperatura média anual, temperatura média no trimestre mais frio, temperatura média no trimestre mais seco, temperatura média no trimestre mais quente e temperatura média no trimestre mais úmido. Assim como no primeiro conjunto, Santa Catarina (2009), utilizou o algoritmo BIOCLIM para auxiliar na seleção de pontos de pseudo-ausência, neste caso foram 50 pontos.



Figura 4.3 - *Strix Varia*, a coruja listrada

Fonte: Lewis (2013)

A Figura 4.3 mostra um exemplar da espécie *Strix Varia*, conhecida como coruja listrada e nativa da América do Norte (LEWIS, 2013). Seu conjunto original de dados, aqui chamado de Conjunto B, contém 1218 pontos de presença e 7 *layers* geográficas, sendo: temperatura média diária, variação da temperatura, precipitação anual, número de dias úmidos, elevação do terreno, declividade do terreno e exposição. Santa Catarina (2009) selecionou, aleatoriamente, 100 pontos de presença do conjunto fornecido e, com o auxílio do algoritmo BIOCLIM, selecionou 100 pontos de pseudo-ausência.

A junção dos pontos de ocorrência com as variáveis ambientais dá origem a um conjunto de vetores ou tuplas de valores de atributos, chamados comumente de exemplos. Cada variável ambiental é caracterizada de acordo com a escala dos atributos. A fim de normalizar os exemplos, foi utilizado, primeiramente, a distribuição normal padrão descrita pela Equação (4.5) onde: x_{ij} é conjunto de valores assumidos pela i -ésima

variável ambiental, no j -ésimo ponto; \bar{x}_i é a média amostral da i -ésima variável; e s_i é o desvio padrão amostral da i -ésima variável ambiental.

$$x'_{ij} = \frac{x_{ij} - \bar{x}_i}{s_i} \quad (4.5)$$

Entretanto, para tornar o conjunto de dados adequado para a entrada da rede neural foi necessário, ainda, utilizar a transformação linear descrita pela Equação de modo que a imagem pertença ao conjunto $[0, 1]$. O Anexo A apresenta tanto os dados iniciais como os obtidos após as transformações.

$$x''_{ij} = \frac{(\text{abs}(\min(x'_i)) + x'_{ij})}{\text{abs}(\max(x'_i)) + \text{abs}(\min(x'_i))} \quad (4.6)$$

4.3 Métodos de análise

Grande parte dos algoritmos aplicados à modelagem de distribuição de espécies são baseados em técnicas preditivas, cujos modelos produzidos são avaliados, geralmente, através de análises estatísticas. Para auxiliar a avaliação e a consequente validação dos modelos, diferentes medidas estatísticas podem ser consideradas. Em seu trabalho, Iwashita (2008) ressalta a necessidade de mais de uma métrica de avaliação pois, assim como os modelos, elas possuem sensibilidade e restrições de uso. Portanto, neste trabalho foram utilizadas as métricas variadas da matriz de confusão como medidas para validação dos modelos.

O objetivo do trabalho é ambicioso e busca avaliar a utilização de redes neurais artificiais na modelagem de distribuição de espécies, deste modo, algumas perguntas devem ser respondidas: Quão dependente de pontos de ausência o algoritmo é? Como o método se comporta com poucos pontos de treinamento? Há um arranjo de neurônios que se sobressai aos demais? Para responder essas perguntas foi imprescindível a execução de diversos experimentos, com diferentes variáveis:

- Número de camadas escondidas;
- Quantidade de neurônios em cada camada;
- Quantidade de pontos de exemplos de treinamento;
- Número de pontos de ausência.

Existem inúmeros mapeamentos que respondem igualmente bem a um conjunto finito de exemplos (VILLANUEVA, 2006). Ao realizar medições com diferentes configurações de redes, busca-se reduzir o espaço amostral a um pequeno intervalo de possibilidades.

Medir a dependência de exemplos de treinamento é um aspecto essencial aos SDMs pois, em muitos casos, é extremamente difícil encontrar exemplares de uma espécie, resultando em poucos pontos coletados e, conseqüentemente, poucos exemplos para o treinamento e validação de modelos. A fim de cumprir esta tarefa, foram executadas avaliações com diferentes quantidades de exemplos de treinamento.

Como discutido brevemente no capítulo 2, a imprecisão e difícil acesso a pontos de ausência tornam os algoritmos, pouco dependes de pontos de ausência, mais robustos. Ao efetuar avaliações com quantidades distintas de pontos de ausência, permite inferir quanto o método proposto é dependente desses exemplos.

Duas estratégias foram adotadas para minimizar as conseqüências de exemplos discrepantes. A primeira estratégia consiste na repetição dos experimentos; todos os experimentos foram repetidos cinco vezes e sua avaliação consiste na média das medidas obtidas. Já a segunda estratégia diz respeito à divisão do conjunto de treinamento em treino e teste, aleatorizados a cada execução.

Em um conjunto inicial de experimentos, não documentado, buscou-se definir quais valores seriam usados para a taxa de aprendizado, erro máximo aceitável e quantidade de épocas de treinamento. Entretanto, não foi possível encontrar um conjunto de valores onde a rede concluía seu treinamento com uma taxa de erro inferior a aceitável, mesmo com uma quantidade extremamente alta de épocas de treinamento; visivelmente a rede estava convergindo, pois conforme o número de épocas crescia, a acurácia da rede melhorava. Acredita-se que isto ocorra devido à utilização de pontos de pseudo-ausência que podem ser gerados em locais onde há condições favoráveis a existência da espécie.

5 Análise dos resultados

Nesta sessão é apresentado o resultado de cinco experimentos que foram feitos com o objetivo de avaliar o comportamento do método proposto em diferentes circunstâncias. É importante frisar que os resultados aqui apresentados foram obtidos com 500.000 épocas de treinamento, resultados melhores são obtidos ao aumentar este número.

Todos os experimentos foram executados com uma única camada intermediária, onde diversas quantidades de neurônios foram avaliadas. Nas tabelas seguintes, a primeira coluna descreve a quantidade de neurônios de cada camada: o primeiro valor corresponde a camada de entrada; o segundo, o número de neurônios nas camadas escondidas; o último, o número de neurônios na camada de saída.

Tabela 5.1 - Resultados do Experimento 1 no Conjunto A

	matthews	acurácia	comissão	omissão	sensibilidade	especificidade
8 5 1	0,5321	0,7517	0,2720	0,2303	0,7697	0,7280
8 10 1	0,7664	0,8828	0,1600	0,0848	0,9152	0,8400
8 15 1	0,7240	0,8621	0,2080	0,0848	0,9152	0,7920
8 20 1	0,7470	0,8724	0,1920	0,0788	0,9212	0,8080
8 25 1	0,6827	0,8379	0,3040	0,0545	0,9455	0,6960
8 30 1	0,7742	0,8828	0,1840	0,0667	0,9333	0,8160

Tabela 5.2 - Resultados do Experimento 1 no Conjunto B

	matthews	acurácia	comissão	omissão	sensibilidade	especificidade
7 5 1	0,6317	0,8139	0,2039	0,1680	0,8320	0,7961
7 10 1	0,7222	0,8594	0,1451	0,1360	0,8640	0,8549
7 15 1	0,7221	0,8594	0,1451	0,1360	0,8640	0,8549
7 20 1	0,7601	0,8752	0,1608	0,0880	0,9120	0,8392
7 25 1	0,7439	0,8693	0,1804	0,0800	0,9200	0,8196
7 30 1	0,8039	0,8990	0,1451	0,0560	0,9440	0,8549

O primeiro experimento teve como objetivo avaliar o método em uma condição normal de execução, com certa abundância de exemplos para o treinamento. Ambos os conjuntos foram treinados com 50% dos exemplos: Conjunto A com 32 pontos de presença e 25 pontos de ausência; Conjunto B com 50 pontos de ausência e 50 pontos de presença. Considerando o melhor ajuste de neurônios, em ambos os casos, o modelo chegou próximo a 90% de acurácia e obteve CCM maior que 0.75 o que demonstra que a previsão é muito melhor que uma previsão aleatória. A tabela 5.3 apresenta os resultados obtidos

por Santa Catarina (2009), ao compará-los com o método proposto nota-se uma clara semelhança. Deste modo, podemos afirmar que as RNAs são, de fato, um método válido para a geração de SDMs.

Tabela 5.3 - Resultados obtidos por Santa Catarina (2009)

	matthews	acurácia
Conjunto A	0,640	0,82
Conjunto B	0,821	0,91

Fonte: Adaptado de Santa Catarina (2009)

Tabela 5.4 - Resultados do Experimento 2 no Conjunto A

	matthews	acurácia	comissão	omissão	sensibilidade	especificidade
8 5 1	0,6317	0,8139	0,2039	0,1680	0,8320	0,7961
8 10 1	0,7222	0,8594	0,1451	0,1360	0,8640	0,8549
8 15 1	0,7221	0,8594	0,1451	0,1360	0,8640	0,8549
8 20 1	0,7601	0,8752	0,1608	0,0880	0,9120	0,8392
8 25 1	0,7439	0,8693	0,1804	0,0800	0,9200	0,8196
8 30 1	0,8039	0,8990	0,1451	0,0560	0,9440	0,8549

Tabela 5.5 - Resultados do Experimento 2 no Conjunto B

	matthews	acurácia	comissão	omissão	sensibilidade	especificidade
7 5 1	0,6515	0,8236	0,2148	0,1375	0,8625	0,7852
7 10 1	0,6070	0,8012	0,2444	0,1525	0,8475	0,7556
7 15 1	0,6767	0,8335	0,1975	0,1350	0,8650	0,8025
7 20 1	0,7150	0,8559	0,1778	0,1100	0,8900	0,8222
7 25 1	0,7321	0,8634	0,1580	0,1150	0,8850	0,8420
7 30 1	0,7295	0,8596	0,1852	0,0950	0,9050	0,8148

Com o objetivo de medir a eficácia das RNAs com um conjunto menor de pontos de treinamento, 20% dos exemplos disponíveis: 13 pontos de presença e 10 pontos de ausência para o Conjunto A; 20 pontos de presença e 20 pontos de ausência para o Conjunto B. O segundo experimento, mostrou que apesar de uma pequena queda na acurácia, o método continuou com um bom índice de acerto, mesmo no pior caso.

Tabela 5.6 - Resultados do Experimento 3 no Conjunto A

	matthews	acurácia	comissão	omissão	sensibilidade	especificidade
8 5 1	0,4652	0,7288	0,3644	0,2000	0,8000	0,6356
8 10 1	0,5045	0,7519	0,4133	0,1220	0,8780	0,5867
8 15 1	0,4534	0,7269	0,4933	0,1051	0,8949	0,5067
8 20 1	0,6090	0,8058	0,3644	0,0644	0,9356	0,6356
8 25 1	0,5318	0,7635	0,3156	0,1763	0,8237	0,6844
8 30 1	0,5299	0,7654	0,2844	0,1966	0,8034	0,7156

Tabela 5.7 - Resultados do Experimento 3 no Conjunto B

	matthews	acurácia	comissão	omissão	sensibilidade	especificidade
7 5 1	0,5468	0,7669	0,2549	0,2111	0,7889	0,7451
7 10 1	0,5896	0,7856	0,1736	0,2556	0,7444	0,8264
7 15 1	0,5775	0,7834	0,1495	0,2844	0,7156	0,8505
7 20 1	0,6571	0,8265	0,1758	0,1711	0,8289	0,8242
7 25 1	0,5546	0,7724	0,1846	0,2711	0,7289	0,8154
7 30 1	0,6108	0,7989	0,1495	0,2533	0,7467	0,8505

No terceiro experimento foi constatado uma redução significativa da acurácia, quando comparado com o primeiro experimento. O número de exemplos de treinamento foi reduzido para 10% do conjunto total disponível: 6 pontos de presença e 5 de ausência no conjunto A; 10 pontos de ambos para o conjunto B. Os resultados obtidos mostram que, assim como o esperado, o método é sensível ao tamanho do conjunto de treinamento. Não foi encontrado na literatura nenhum resultado com um conjunto tão reduzido de treinamento impossibilitando, deste modo, uma comparação com os demais métodos.

Tabela 5.8 - Resultados do Experimento 4 no Conjunto A

layers	matthews	acurácia	comissão	omissão	sensibilidade	especificidade
8 5 1	-	0,3976	1,0000	0,0000	1,0000	0,0000
8 10 1	-	0,3976	1,0000	0,0000	1,0000	0,0000
8 15 1	-	0,3976	1,0000	0,0000	1,0000	0,0000
8 20 1	-	0,3976	1,0000	0,0000	1,0000	0,0000
8 25 1	-	0,3976	1,0000	0,0000	1,0000	0,0000
8 30 1	-	0,3976	1,0000	0,0000	1,0000	0,0000

O experimento 4 foi executado apenas no conjunto A, com 32 pontos de presença e nenhum ponto de ausência. Nota-se que sem pontos de ausência o modelo não se comporta como esperado e passa a classificar todos como pontos de presença, isto fica

claro ao notar que a coluna de sensibilidade assumiu seu valor máximo (1) em todos os arranjos da RNA, somado a inexistência de especificidade.

Tabela 5.9 - Resultados do Experimento 5 no Conjunto A

layers	matthews	acurácia	comissão	omissão	sensibilidade	especificidade
8 5 1	0,4625	0,6767	0,5550	0,0424	0,9576	0,4450
8 10 1	0,6467	0,7973	0,3450	0,0303	0,9697	0,6550
8 15 1	0,6761	0,8219	0,2850	0,0485	0,9515	0,7150
8 20 1	0,5561	0,7260	0,4950	0,0061	0,9939	0,5050
8 25 1	0,5988	0,7589	0,3800	0,0727	0,9273	0,6200
8 30 1	0,4960	0,7014	0,4950	0,0606	0,9394	0,5050

Tabela 5.10 - Resultados do Experimento 5 no Conjunto B

layers	matthews	acurácia	comissão	omissão	sensibilidade	especificidade
7 5 1	0,6102	0,7695	0,3531	0,0320	0,9680	0,6469
7 10 1	0,6992	0,8382	0,2296	0,0520	0,9480	0,7704
7 15 1	0,6931	0,8366	0,2247	0,0640	0,9360	0,7753
7 20 1	0,6741	0,8183	0,2642	0,0480	0,9520	0,7358
7 25 1	0,6599	0,8061	0,2815	0,0520	0,9480	0,7185
7 30 1	0,7514	0,8672	0,1778	0,0600	0,9400	0,8222

Como o experimento 4 demonstrou que o método proposto necessita de pontos de ausência para operar, o experimento 5 teve como objetivo avaliar o desempenho da RNA com uma diferença significativa na quantidade de pontos de presença e ausência: 50% das presenças e 20% das ausências. Esta diferença ocasionou uma maior sensibilidade da rede aumentando, assim, os erros de comissão. Surpreendentemente, os resultados deste experimento se mostraram inferiores ao obtidos no experimento 2 que possuía menos exemplos de treinamento.

Ao analisar os experimentos como um todo, nota-se que, assim como o esperado, não há uma topologia que responda melhor ao problema. A quantidade de neurônios nas camadas intermediárias varia tanto de acordo com a espécie, quanto com a variação no número de exemplos de treinamento.

6 Considerações finais

O objetivo principal deste trabalho foi avaliar a utilização de redes neurais artificiais na modelagem de distribuição potencial de espécies. Apesar dos resultados obtidos serem semelhantes aos de Santa Catarina (2009), espera-se que o trabalho desenvolvido possa ser utilizado como base para trabalhos futuros, principalmente como comparativo para novos algoritmos de modelagem. A fundamentação teórica necessária ao desenvolvimento do trabalho foi obtida com base no estudo do processo de modelagem, somado a técnicas de aprendizado de máquina. Com base nessa fundamentação, a acurácia foi estabelecida como principal fator relacionado ao desempenho do algoritmo de modelagem.

Os resultados obtidos mostraram que redes neurais são, de fato, eficazes na tarefa de modelagem, chegando a resultados satisfatórios mesmo com um conjunto pequeno de exemplos de treinamento. Entretanto, o método possui algumas restrições de uso pois, o conjunto de treinamento deve, necessariamente, possuir dados de ausência e, de preferência, em um número semelhante aos de presença. Vícios de coleta ou baixo nível de ocorrência, problemas oriundos do processo de coletada em campo, influenciam significativamente no desempenho do algoritmo. Entretanto, este não é um problema exclusivo das RNAs, são problemas comuns na maior parte dos algoritmos de modelagem existentes.

Determinar qual a melhor quantidade de neurônios nas camadas escondidas – problema comum em projetos que utilizam redes neurais artificiais – é a principal complicação do método. Fato este que dificulta a implantação real do método, pois caberá ao usuário final do modelador decidir quantos neurônios serão utilizados. Automatizar este processo com, por exemplo, algoritmos genéticos, pode vir a sanar este problema, mas acarretará em um maior custo computacional.

6.1 Trabalhos futuros

- Utilizar algoritmos genéticos para obter o melhor ajuste de rede.
- Considerar erros de posicionamento na avaliação de desempenho dos modelos.
- Avaliar o custo computacional dos algoritmos de modelagem.

Referências bibliográficas

- ANDRADE, L. N. D. **Redes neurais artificiais aplicadas na identificação automática de áreas cafeeiras em imagens de satélite**. Universidade Federal de Minas Gerais. Belo Horizonte. 2011.
- AUSTIN, M. Species distribution models and ecological theory: A critical assessment and some possible new approaches. **Ecological Modelling**, v. 157, n. 2, p. 101-118, Novembro 2002.
- BARTLEY, G. Birds of Bolivia. **Nature Photography**, 2009. Disponível em: <<http://www.glennbartley.com/naturephotography/birds/EMERALD-BELLIED%20WOODNYMPH.html>>. Acesso em: 9 Outubro 2014.
- BIODIVERSITY Institute. **Desktop Garp**, 2007. Disponível em: <<http://www.nhm.ku.edu/desktopgarp/>>. Acesso em: 8 Outubro 2014.
- BRAGA, A. D. P.; CARVALHO, A. P. D. L.; LUDEMIR, T. B. **Redes Neurais Artificiais: Teoria e Aplicações**. Rio de Janeiro: Livros Técnicos e Científicos Editora S.A, 2000.
- CANHOS, V. P. Informática para Biodiversidade: Padrões, Protocolos e Ferramentas. **Ciência e Cultura**, v. 55, n. 2, p. 45-47, 2003.
- CORRÊA, W. R.; PORTUGAL, M. S. Previsão de séries de tempo na presença de mudança estrutural: redes neurais artificiais e modelos estruturais. **Economia Aplicada**, v. 2, p. 487-514, 1998. ISSN 3.
- FINAMORE, P. P. **Avaliação de Três Métodos para Geração de Pontos de Pseudo-Ausência Sobre a Qualidade dos Modelos de Distribuição de Espécies Ajustados pelo Sistema SAHGA SDM**. Universidade Estadual do Oeste do Paraná. Cascavel. 2010.
- GALO, M. D. L. B. T.; FERREIRA, M. S.; UTSUMI, A. G. Utilização da Geoestatística e Redes Neurais Artificiais na inferência espacial de fitoplâncton em estágios iniciais de floração. **XVI Simpósio Brasileiro de Sensoriamento Remoto - SBSR**, Foz do Iguaçu, Abril 2013.
- GUISAN, A.; ZIMMERMANN, N. E. Predictive habitat distribution models in ecology. **Ecological Modelling**, v. 135, p. 147-186, 2000.
- HUTCHINSON, G. E. Concluding Remarks. **Cold Spring Harbour Symposium on Quantitative Biology**, v. 22, p. 415-427, 1957.

IWASHITA, F. **Sensibilidade de modelos de distribuição de espécies a erros de posicionamento de dados de coleta.** Instituto Nacional de Pesquisas Espaciais. São José dos Campos. 2008.

LEWIS, D. Barred Owl - Strix varia. **The Owl Pages**, 2013. Disponível em: <<http://www.owlpages.com/owls.php?genus=Strix&species=varia>>. Acesso em: 09 Outubro 2014.

LUDWING JR, O.; MONTGOMERY, E. **Redes Neurais - Fundamentos e Aplicações com Programas em C.** Rio de Janeiro: Ciência Moderna Ltda, 2007.

MARCO JR., P. D.; SIQUEIRA, M. F. Como determinar a distribuição potencial de espécies sob uma abordagem conservacionista? **Megadiversidade**, v. 5, Dezembro 2009. ISSN 1-2.

MATEO, R. G. et al. Profile or group discriminative techniques? Generating reliable species distribution models using pseudo-absences and target-group absences from natural history collections. **Diversity and Distributions**, 2010. 84-94.

MATOS, R. D. S. **Avaliação das ferramentas de modelagem preditiva de nicho fundamental para espécies de aves do Parque Estadual da Serra do Mar e Núcleo São Sebastião - SP.** Escola Superior de Agricultura "Luiz de Queiroz". Piracicaba. 2010.

PORTUGAL, M. S.; FERNANDES, L. G. L. Redes neurais artificiais e previsão de séries econômicas: uma introdução. **Nova Economica**, Belo Horizonte, v. 6, p. 51-74, Julho 1996. ISSN 1.

PRATI, R. C.; BATISTA, G. E. A. P. A.; MONARD, M. C. Curvas ROC para avaliação de classificadores. **IEEE Latin America Transactions**, v. 6, p. 215-222, 2008. ISSN 2.

REZENDE, S. O. **Sistemas Inteligentes: Fundamentos e Aplicações.** 1ª. ed. Barueri - SP: Manole, 2005.

RODRIGUES, F. A. **Um Método de Referência para Análise de Desempenho Preditivo de Algoritmos de Modelagem de Distribuição de Espécies.** Universidade de São Paulo. São Paulo. 2012.

RODRIGUES, F. A.; RODRIGUES, S. D. C.; CORRÊA, P. L. P. Modelagem da Biodiversidade Utilizando Redes Neurais Artificiais. **XXX Congresso da Sociedade Brasileira de Computação - Computação Verde: Desafios Científicos e Tecnológicos**, São Paulo, 2010. Disponível em: <<http://ppgee.poli.usp.br/Producao/12945/Modelagem-da-Biodiversidade-Utilizando-Redes-Neurais-Artificiais.html>>. Acesso em: 24 Julho 2014.

- SANTA CATARINA, A. **SAHGA – Um Algoritmo Genético Híbrido com representação explícita de relacionamentos espaciais para análise de dados geoespaciais**. Instituto Nacional de Pesquisas Espaciais. São José dos Campos. 2009.
- SANTOS, A. M. D. et al. Usando Redes Neurais Artificiais e Regressão Logística na Predição da Hepatite A. **Bras Epidemiol**, v. 8, p. 117-126, 2005. ISSN 2.
- SHIFFMAN, D. **The Nature of Code**. 1ª. ed. [S.l.]: [s.n.], 2012.
- SILVA, I. N.; SPATTI, D. H.; FLAUZINO, R. A. **Redes Neurais Artificiais para engenharia e ciências aplicadas**. 1ª. ed. São Paulo: Arliber, 2010.
- SILVA, P. C. D. L. E. **Paralelização do Algoritmo de Backpropagation para Clusters Beowulf**. Instituto Educacional Santo Agostinho. Monte Carlos. 2005.
- SIQUEIRA, M. F. D. **Uso de modelagem de nicho fundamental na avaliação**. Universidade de São Paulo. São Carlos. 2005.
- SOBERÓN, J.; PETERSON, A. T. Interpretation of models of fundamental ecological niches and species' distributional areas. **Biodiversity Informatics** , v. 2, p. 1-10, 2005.
- TÔRRES, N. M.; VERCILLO, U. E. Como Ferramentas de Modelagem de Distribuição de Espécies Podem Subsidiar Ações de Governo? **Brazilian Journal of Nature Conservation**, v. 10, n. 2, p. 228-230, Dezembro 2012.
- VILLANUEVA, W. J. P. **Comitê de Máquinas em Predição de Séries Temporais**. Universidade Estadual de Campinas. Campinas. 2006.
- VILLANUEVA, W. J. P. **Síntese automática de redes neurais artificiais com conexões à frente arbitrárias**. Universidade Estadual de Campinas. Campinas. 2011.

Anexo A - A base de dados

A base de dados utilizada neste trabalho é a mesma adotada por Santa Catarina (2009) e consiste em pontos de nicho de duas espécies: *Thalurna Furcata Boliviana* e *Strix Varia*. As tabelas seguintes representam os valores iniciais de cada um dos pontos de nichos e, em seguida, será apresentado os valores utilizados para o treinamento e avaliação do método.

	x1	x2	x3	x4	x5	x6	x7	x8	y
1	533	675	2.772	2.643	2.581	2.581	2.698	2.640	-100
2	180	401	1.679	2.606	2.383	2.421	2.733	2.702	-100
3	160	403	1.574	2.612	2.394	2.408	2.742	2.699	-100
4	116	391	1.164	2.665	2.355	2.377	2.863	2.832	-100
5	44	229	535	2.570	2.232	2.232	2.850	2.850	-100
6	188	414	1.659	2.625	2.328	2.606	2.853	2.720	-100
7	1	386	584	1.432	956	979	1.758	1.758	-100
8	177	49	448	845	246	1.449	1.449	309	-100
9	857	190	2.145	2.605	2.546	2.713	2.713	2.546	-100
10	23	81	178	2.176	1.484	1.986	2.761	2.670	-100
11	14	171	519	553	400	400	642	622	-100
12	99	376	1.499	2.691	2.588	2.592	2.784	2.690	-100
13	100	347	1.410	2.649	2.490	2.521	2.763	2.667	-100
14	92	326	810	2.545	2.067	2.268	2.937	2.937	-100
15	5	160	240	706	380	448	976	976	-100
16	846	339	2.367	2.571	2.546	2.591	2.611	2.547	-100
17	19	206	335	1.897	1.056	1.547	2.649	2.541	-100
18	171	498	1.791	2.719	2.618	2.618	2.800	2.745	-100
19	66	462	949	597	342	342	745	745	-100
20	185	66	1.116	2.589	2.490	2.597	2.732	2.507	-100
21	308	489	1.789	2.541	2.307	2.433	2.699	2.662	100
22	118	520	1.271	2.462	2.271	2.287	2.580	2.570	100
23	88	457	1.100	2.037	1.858	1.880	2.153	2.140	100
24	115	599	1.440	2.344	2.141	2.150	2.473	2.463	100
25	129	635	1.539	2.254	2.049	2.056	2.380	2.373	100
26	119	625	1.505	2.125	1.921	1.927	2.250	2.243	100
27	85	421	1.209	2.207	2.030	2.032	2.316	2.303	100
28	80	393	1.169	1.806	1.634	1.637	1.909	1.895	100
29	402	790	2.797	2.479	2.170	2.251	2.695	2.694	100
30	200	515	1.799	2.410	2.118	2.209	2.632	2.601	100
31	48	217	719	2.146	1.898	1.931	2.341	2.284	100
32	10	195	597	1.296	1.011	1.013	1.512	1.464	100
33	185	451	1.156	2.480	2.120	2.315	2.714	2.702	100
34	29	334	682	2.290	1.958	1.967	2.531	2.526	100

35	33	353	718	2.008	1.622	1.639	2.290	2.272	100
36	428	921	2.826	2.531	2.376	2.449	2.621	2.617	100
37	63	268	858	2.023	1.879	1.879	2.136	2.063	100
38	246	742	2.215	2.533	2.398	2.407	2.617	2.590	100
39	429	872	2.695	2.355	2.204	2.268	2.450	2.441	100
40	312	806	2.423	2.505	2.371	2.375	2.590	2.565	100
41	271	763	2.287	2.518	2.384	2.388	2.602	2.573	100
42	718	1.346	4.159	2.457	2.314	2.386	2.542	2.536	100
43	502	977	3.020	2.108	1.945	2.015	2.205	2.199	100
44	408	906	2.737	2.538	2.382	2.465	2.630	2.617	100
45	1.132	1.698	5.273	1.800	1.672	1.740	1.875	1.871	100
46	132	442	1.902	2.593	2.526	2.539	2.668	2.590	-100
47	477	705	2.642	2.569	2.512	2.512	2.610	2.584	-100
48	1.143	170	2.959	2.561	2.513	2.640	2.640	2.524	-100
49	1.001	410	2.393	2.667	2.606	2.699	2.745	2.606	-100
50	107	382	1.154	2.649	2.339	2.348	2.862	2.812	-100
51	129	442	1.101	2.598	2.267	2.292	2.826	2.818	-100
52	17	404	1.580	2.792	2.669	2.669	2.863	2.859	-100
53	18	430	684	142	362	193	536	536	-100
54	140	412	1.459	2.747	2.483	2.571	2.939	2.898	-100
55	28	398	1.616	2.591	2.458	2.458	2.674	2.650	-100
56	210	441	1.741	2.619	2.323	2.597	2.844	2.715	-100
57	186	488	2.195	2.541	2.490	2.508	2.603	2.519	-100
58	21	116	349	1.667	1.325	1.436	1.937	1.832	-100
59	17	0	25	1.197	868	1.440	1.472	946	-100
60	272	165	957	2.725	2.693	2.728	2.766	2.710	-100
61	11	227	378	837	455	459	1.104	1.104	-100
62	916	249	2.050	2.589	2.492	2.635	2.695	2.505	-100
63	833	631	2.976	2.663	2.597	2.707	2.707	2.650	-100
64	81	301	751	2.541	2.155	2.351	2.815	2.812	-100
65	123	391	1.010	2.555	2.184	2.200	2.834	2.834	-100
66	8	223	363	930	558	569	1.198	1.198	-100
67	20	311	741	636	415	415	779	761	-100
68	120	198	698	1.302	675	675	1.971	1.848	-100
69	44	307	717	624	489	489	696	696	-100
70	354	523	1.887	2.706	2.654	2.664	2.746	2.702	-100
71	227	490	1.625	2.623	2.549	2.566	2.668	2.622	-100
72	45	93	293	1.149	366	366	1.916	1.582	-100
73	146	427	1.825	2.593	2.492	2.495	2.679	2.615	-100
74	114	373	1.591	2.650	2.539	2.557	2.754	2.652	-100
75	78	545	1.532	2.627	2.445	2.445	2.717	2.674	-100
76	92	451	1.733	2.445	2.265	2.265	2.553	2.522	100
77	299	491	1.776	2.569	2.336	2.460	2.727	2.691	100
78	307	488	1.785	2.450	2.221	2.346	2.606	2.571	100
79	144	613	1.513	2.287	2.087	2.100	2.412	2.398	100
80	86	467	1.109	2.134	1.949	1.971	2.252	2.239	100

81	125	621	1.502	2.320	2.115	2.125	2.448	2.437	100
82	128	632	1.533	2.280	2.075	2.082	2.405	2.398	100
83	125	555	1.424	2.146	1.948	1.950	2.266	2.257	100
84	121	632	1.520	2.125	1.920	1.926	2.252	2.246	100
85	94	533	1.273	1.885	1.697	1.699	2.006	1.987	100
86	219	658	2.079	2.571	2.291	2.392	2.754	2.742	100
87	82	423	1.179	2.053	1.872	1.876	2.170	2.154	100
88	65	336	1.010	1.400	1.219	1.227	1.518	1.493	100
89	70	347	1.061	1.803	1.608	1.616	1.926	1.910	100
90	396	780	2.761	2.462	2.154	2.235	2.679	2.676	100
91	188	508	1.780	2.446	2.156	2.246	2.668	2.636	100
92	12	146	480	2.213	1.996	1.996	2.369	2.313	100
93	202	635	1.465	2.376	2.051	2.052	2.598	2.598	100
94	215	768	1.725	2.335	2.017	2.136	2.554	2.554	100
95	212	631	1.482	2.319	1.994	2.074	2.543	2.543	100
96	93	333	803	2.110	1.822	1.924	2.316	2.314	100
97	38	342	675	2.051	1.743	1.743	2.283	2.282	100
98	24	337	679	2.068	1.733	1.750	2.307	2.299	100
99	140	609	1.857	2.414	2.272	2.272	2.494	2.481	100
100	318	752	2.300	2.579	2.420	2.489	2.673	2.670	100
101	249	742	2.217	2.527	2.392	2.399	2.610	2.584	100
102	607	1.159	3.568	2.505	2.361	2.434	2.589	2.586	100
103	372	864	2.610	2.502	2.363	2.404	2.584	2.562	100
104	425	861	2.658	2.361	2.200	2.270	2.460	2.451	100
105	380	789	2.433	2.297	2.129	2.199	2.399	2.391	100
106	282	759	2.284	2.529	2.389	2.393	2.615	2.592	100
107	427	860	2.654	1.900	1.728	1.798	2.002	1.998	100
108	856	1.535	4.756	2.306	2.161	2.234	2.390	2.387	100
109	815	1.498	4.568	2.268	2.118	2.194	2.353	2.350	100
110	1.123	1.687	5.239	1.707	1.577	1.646	1.783	1.780	100
111	417	952	2.877	2.441	2.275	2.357	2.535	2.535	100
112	428	992	2.882	2.329	2.154	2.235	2.431	2.431	100
113	370	888	2.600	2.214	2.035	2.117	2.320	2.320	100
114	457	1.091	2.924	1.552	1.355	1.441	1.664	1.664	100
115	343	916	2.425	1.772	1.574	1.663	1.886	1.886	100

Tabela A.1 - Dados originais da *Thalurania Furcata Boliviana*

	x1	x2	x3	x4	x5	x6	x7	x8	y
1	0,465	0,397	0,523	0,943	0,963	0,949	0,899	0,887	0
	5	6	3	9	4	8	7	0	
2	0,156	0,236	0,315	0,929	0,898	0,894	0,914	0,910	0
	8	1	2	8	4	9	4	4	
3	0,139	0,237	0,295	0,932	0,902	0,890	0,917	0,909	0
	2	0	1	0	0	3	9	4	

4	0,100 4	0,229 8	0,217 1	0,952 0	0,889 3	0,879 7	0,968 4	0,960 0	0
5	0,037 4	0,134 7	0,097 3	0,916 3	0,849 0	0,830 0	0,962 7	0,966 7	0
6	0,163 4	0,243 3	0,311 4	0,936 8	0,880 4	0,958 2	0,964 0	0,917 3	0
7	0,000 0	0,227 1	0,106 6	0,486 6	0,431 6	0,401 2	0,508 6	0,551 5	0
8	0,153 8	0,028 6	0,080 7	0,265 2	0,199 1	0,562 2	0,380 1	0,000 0	0
9	0,749 6	0,111 9	0,403 9	0,929 4	0,952 0	0,995 0	0,906 0	0,851 3	0
10	0,019 3	0,047 4	0,029 2	0,767 6	0,604 4	0,745 9	0,925 8	0,898 5	0
11	0,011 0	0,100 4	0,094 1	0,155 0	0,249 4	0,202 9	0,044 3	0,119 2	0
12	0,085 4	0,221 0	0,280 9	0,961 9	0,965 8	0,953 3	0,935 5	0,905 9	0
13	0,086 9	0,204 3	0,263 9	0,946 0	0,933 6	0,929 1	0,926 8	0,897 3	0
14	0,079 5	0,191 9	0,149 6	0,906 9	0,795 1	0,842 6	0,999 1	1,000 0	0
15	0,003 5	0,094 1	0,041 1	0,212 7	0,242 9	0,219 6	0,183 3	0,254 0	0
16	0,740 2	0,199 4	0,446 3	0,916 5	0,951 8	0,952 9	0,863 3	0,851 4	0
17	0,015 5	0,121 2	0,059 2	0,662 4	0,464 2	0,595 7	0,879 2	0,849 3	0
18	0,148 8	0,293 2	0,336 6	0,972 6	0,975 5	0,962 4	0,941 9	0,926 9	0
19	0,057 0	0,271 9	0,176 1	0,171 5	0,230 7	0,183 3	0,086 9	0,165 9	0
20	0,161 4	0,038 4	0,207 9	0,923 4	0,933 7	0,955 1	0,913 8	0,836 3	0
21	0,269 2	0,287 9	0,336 1	0,905 1	0,873 7	0,899 1	0,900 1	0,895 5	1
22	0,102 2	0,306 1	0,237 4	0,875 5	0,861 7	0,849 1	0,850 6	0,860 4	1
23	0,076 5	0,269 0	0,204 8	0,715 1	0,726 8	0,709 7	0,673 0	0,696 6	1
24	0,099 8	0,352 7	0,269 7	0,830 9	0,819 3	0,802 1	0,805 9	0,819 7	1
25	0,111 8	0,373 9	0,288 5	0,796 9	0,789 1	0,770 1	0,767 2	0,785 2	1
26	0,103 1	0,367 9	0,282 0	0,748 1	0,747 2	0,725 9	0,713 2	0,735 9	1

27	0,073 9	0,248 0	0,225 6	0,779 2	0,782 9	0,761 8	0,740 7	0,758 9	1
28	0,069 5	0,231 0	0,218 0	0,627 9	0,653 5	0,626 5	0,571 2	0,603 6	1
29	0,351 3	0,465 4	0,528 3	0,881 7	0,828 7	0,836 6	0,898 5	0,907 4	1
30	0,174 3	0,302 9	0,338 1	0,855 7	0,811 8	0,822 2	0,872 3	0,872 3	1
31	0,041 1	0,127 4	0,132 3	0,756 1	0,739 8	0,727 2	0,751 1	0,751 6	1
32	0,007 4	0,114 4	0,109 0	0,435 3	0,449 6	0,412 8	0,406 2	0,439 7	1
33	0,161 4	0,265 2	0,215 5	0,882 1	0,812 5	0,858 5	0,906 4	0,910 5	1
34	0,024 7	0,196 2	0,125 2	0,810 6	0,759 4	0,739 5	0,830 2	0,843 8	1
35	0,027 9	0,207 7	0,132 0	0,704 1	0,649 6	0,627 1	0,730 0	0,747 0	1
36	0,374 2	0,542 4	0,533 8	0,901 4	0,896 1	0,904 4	0,867 5	0,878 3	1
37	0,054 4	0,157 6	0,158 7	0,709 6	0,733 7	0,709 5	0,665 9	0,667 5	1
38	0,214 8	0,436 7	0,417 3	0,902 4	0,903 6	0,889 9	0,865 7	0,867 9	1
39	0,374 8	0,513 6	0,508 7	0,834 9	0,840 0	0,842 5	0,796 5	0,811 2	1
40	0,272 5	0,474 5	0,456 8	0,891 8	0,894 5	0,879 1	0,854 6	0,858 3	1
41	0,236 7	0,449 6	0,431 0	0,896 4	0,898 8	0,883 5	0,859 8	0,861 4	1
42	0,628 2	0,792 6	0,787 8	0,873 6	0,875 8	0,882 9	0,834 6	0,847 4	1
43	0,438 4	0,575 5	0,570 7	0,741 8	0,755 1	0,755 9	0,694 4	0,719 2	1
44	0,356 6	0,533 4	0,516 8	0,904 1	0,898 1	0,909 8	0,871 4	0,878 3	1
45	0,990 6	1,000 0	1,000 0	0,625 5	0,665 7	0,661 6	0,557 1	0,594 6	1
46	0,114 3	0,260 2	0,357 7	0,925 0	0,945 4	0,935 1	0,887 0	0,867 8	0
47	0,416 7	0,415 2	0,498 7	0,915 9	0,940 8	0,926 1	0,863 1	0,865 5	0
48	1,000 0	0,100 0	0,559 0	0,912 9	0,941 2	0,969 8	0,875 5	0,842 7	0
49	0,875 6	0,241 4	0,451 2	0,952 6	0,971 5	0,990 2	0,919 3	0,874 0	0

50	0,092 4	0,224 9	0,215 1	0,946 1	0,884 0	0,870 0	0,967 7	0,952 5	0
51	0,112 2	0,260 3	0,205 1	0,926 7	0,860 5	0,850 7	0,952 8	0,954 7	0
52	0,013 5	0,237 6	0,296 3	1,000 0	0,992 3	0,979 9	0,968 3	0,970 3	0
53	0,014 7	0,253 2	0,125 5	0,000 0	0,000 0	0,000 0	0,000 0	0,086 4	0
54	0,121 4	0,242 6	0,273 3	0,983 0	0,931 2	0,946 3	1,000 0	0,985 2	0
55	0,023 6	0,234 1	0,303 1	0,924 1	0,922 9	0,907 6	0,889 8	0,890 6	0
56	0,183 0	0,259 3	0,326 9	0,934 7	0,878 9	0,955 1	0,960 4	0,915 6	0
57	0,162 0	0,287 4	0,413 5	0,905 2	0,933 7	0,924 6	0,860 0	0,841 1	0
58	0,017 6	0,068 0	0,061 7	0,575 2	0,552 3	0,557 8	0,583 0	0,579 7	0
59	0,014 3	0,000 0	0,000 0	0,398 1	0,402 8	0,559 0	0,389 3	0,242 4	0
60	0,237 0	0,096 7	0,177 6	0,974 6	1,000 0	1,000 0	0,928 0	0,913 4	0
61	0,008 4	0,133 2	0,067 3	0,262 2	0,267 7	0,223 1	0,236 5	0,302 7	0
62	0,801 8	0,146 6	0,385 9	0,923 3	0,934 2	0,968 1	0,898 3	0,835 7	0
63	0,729 0	0,371 7	0,562 3	0,951 5	0,968 7	0,992 9	0,903 5	0,890 9	0
64	0,069 9	0,177 3	0,138 3	0,905 2	0,823 8	0,870 7	0,948 4	0,952 4	0
65	0,106 7	0,230 3	0,187 8	0,910 6	0,833 4	0,819 4	0,956 2	0,960 6	0
66	0,006 1	0,131 4	0,064 5	0,297 3	0,301 1	0,260 9	0,275 5	0,338 3	0
67	0,016 5	0,183 1	0,136 5	0,186 5	0,254 3	0,208 0	0,101 2	0,172 0	0
68	0,104 0	0,116 5	0,128 2	0,437 7	0,339 6	0,297 2	0,597 1	0,585 7	0
69	0,037 9	0,180 6	0,131 8	0,181 6	0,278 7	0,233 5	0,066 7	0,147 4	0
70	0,309 4	0,308 0	0,354 8	0,967 5	0,987 1	0,978 1	0,919 8	0,910 5	0
71	0,198 2	0,288 2	0,304 8	0,936 2	0,952 7	0,944 7	0,886 9	0,880 2	0
72	0,038 3	0,054 7	0,051 2	0,380 1	0,238 5	0,191 5	0,574 4	0,484 5	0

73	0,126 7	0,251 5	0,342 9	0,925 0	0,934 2	0,920 2	0,891 6	0,877 4	0
74	0,098 6	0,219 3	0,298 4	0,946 5	0,949 4	0,941 3	0,922 8	0,891 5	0
75	0,067 2	0,320 9	0,287 2	0,937 9	0,918 9	0,903 2	0,907 6	0,900 1	0
76	0,079 9	0,265 6	0,325 5	0,869 2	0,860 0	0,841 6	0,839 2	0,842 2	1
77	0,260 9	0,288 9	0,333 7	0,916 0	0,883 1	0,908 4	0,911 6	0,906 3	1
78	0,267 8	0,287 2	0,335 4	0,871 0	0,845 4	0,869 1	0,861 3	0,860 6	1
79	0,124 9	0,360 9	0,283 6	0,809 4	0,801 7	0,785 0	0,780 4	0,795 0	1
80	0,074 1	0,275 1	0,206 5	0,751 5	0,756 6	0,740 9	0,713 9	0,734 3	1
81	0,108 7	0,365 6	0,281 4	0,821 8	0,810 9	0,793 6	0,795 7	0,809 8	1
82	0,111 4	0,372 1	0,287 3	0,806 6	0,797 6	0,778 9	0,777 9	0,794 9	1
83	0,108 7	0,326 8	0,266 5	0,756 2	0,756 2	0,733 5	0,719 8	0,741 3	1
84	0,104 6	0,372 0	0,284 9	0,748 4	0,747 1	0,725 4	0,714 0	0,737 0	1
85	0,081 3	0,313 7	0,237 9	0,657 5	0,674 0	0,647 6	0,611 6	0,638 4	1
86	0,191 1	0,387 4	0,391 5	0,916 5	0,868 6	0,884 8	0,922 7	0,925 8	1
87	0,071 1	0,249 1	0,220 0	0,721 2	0,731 2	0,708 5	0,679 7	0,702 2	1
88	0,055 9	0,197 4	0,187 7	0,474 8	0,517 5	0,486 1	0,408 8	0,450 4	1
89	0,060 0	0,204 2	0,197 3	0,626 7	0,645 0	0,619 2	0,578 6	0,609 0	1
90	0,345 9	0,459 6	0,521 4	0,875 5	0,823 7	0,831 3	0,891 8	0,900 7	1
91	0,163 8	0,298 9	0,334 4	0,869 4	0,824 1	0,835 1	0,886 9	0,885 4	1
92	0,009 5	0,085 6	0,086 7	0,781 4	0,771 8	0,749 3	0,762 8	0,762 5	1
93	0,175 7	0,373 9	0,274 3	0,843 0	0,789 8	0,768 7	0,858 2	0,871 0	1
94	0,187 5	0,452 4	0,323 8	0,827 4	0,778 7	0,797 5	0,839 6	0,854 1	1
95	0,184 6	0,371 3	0,277 7	0,821 3	0,771 2	0,776 0	0,835 0	0,850 0	1

96	0,080 5	0,196 0	0,148 2	0,742 6	0,714 8	0,724 8	0,740 5	0,763 0	1
97	0,032 0	0,201 0	0,124 0	0,720 4	0,689 0	0,662 7	0,727 0	0,750 8	1
98	0,020 2	0,198 3	0,124 7	0,726 6	0,685 7	0,665 1	0,737 0	0,757 2	1
99	0,121 6	0,358 5	0,349 1	0,857 1	0,862 3	0,844 0	0,814 8	0,826 6	1
100	0,277 4	0,442 9	0,433 6	0,919 6	0,910 5	0,918 3	0,889 2	0,898 2	1
101	0,217 0	0,436 9	0,417 7	0,899 9	0,901 5	0,887 4	0,863 0	0,865 8	1
102	0,530 8	0,682 3	0,675 1	0,891 6	0,891 5	0,899 2	0,854 2	0,866 4	1
103	0,325 0	0,508 7	0,492 5	0,890 5	0,892 1	0,889 1	0,852 2	0,857 3	1
104	0,371 2	0,506 8	0,501 7	0,837 4	0,838 6	0,843 1	0,800 4	0,815 1	1
105	0,331 9	0,464 4	0,458 8	0,813 3	0,815 4	0,818 7	0,775 3	0,792 3	1
106	0,245 7	0,446 8	0,430 4	0,900 8	0,900 4	0,885 2	0,865 1	0,868 9	1
107	0,372 9	0,506 4	0,501 0	0,663 2	0,684 1	0,681 5	0,610 0	0,642 6	1
108	0,748 7	0,903 9	0,901 3	0,816 5	0,825 9	0,831 0	0,771 4	0,790 5	1
109	0,712 8	0,882 1	0,865 7	0,802 3	0,811 9	0,817 2	0,756 2	0,776 8	1
110	0,983 2	0,993 9	0,993 5	0,590 5	0,634 6	0,629 5	0,518 9	0,559 6	1
111	0,364 2	0,560 9	0,543 4	0,867 7	0,863 0	0,873 1	0,831 8	0,847 0	1
112	0,374 0	0,584 2	0,544 3	0,825 1	0,823 7	0,831 2	0,788 5	0,807 4	1
113	0,322 8	0,523 0	0,490 6	0,781 7	0,784 5	0,790 7	0,742 5	0,765 3	1
114	0,399 0	0,642 7	0,552 3	0,531 9	0,562 1	0,559 5	0,469 4	0,515 6	1
115	0,299 3	0,539 4	0,457 3	0,615 0	0,633 9	0,635 4	0,561 9	0,600 3	1

Tabela A.2 – Dados resultantes de *Thalurania Furcata Boliviana*

	x1	x2	x3	x4	x5	x6	x7	y
1	159	111	9.000	414	4	27	-15	0
2	143	120	15.19 9	371	38	25	12	0

3	176	75	7.412	292	92	37	35	0
4	47	162	21.55 4	1.278	12	10	150	0
5	160	105	8.409	360	21	28	-18	0
6	71	204	8.772	2.987	448	14	6	0
7	45	164	24.90 2	1.524	600	13	158	0
8	44	128	17.85 3	177	28	14	224	0
9	32	134	7.528	284	161	10	224	0
10	28	184	22.77 4	1.294	148	6	193	0
11	169	102	24.76 2	621	13	24	-32	0
12	44	134	9.308	485	120	11	205	0
13	45	128	9.25	175	31	14	221	0
14	132	122	25.15 7	305	2	19	-12	0
15	140	114	28.36 7	397	27	25	1	0
16	55	152	12.09 6	1.067	29	12	143	0
17	26	172	1.062	865	23	5	207	0
18	151	113	13.50 0	397	5	29	-5	0
19	76	143	12.05 8	564	73	13	74	0
20	47	136	19.18 1	275	129	23	225	0
21	159	103	33.12 6	571	51	25	-34	0
22	80	117	34.15 7	240	2	13	29	0
23	144	112	12.17 6	278	35	21	-15	0
24	68	165	11.72 8	794	77	14	103	0
25	46	151	33.82 0	820	62	11	179	0
26	172	105	22.50 0	610	1	24	-45	0
27	132	106	33.34 3	305	3	17	-15	0
28	114	111	9.140	380	29	16	4	0
29	67	154	23.57 8	975	43	12	106	0

30	38	162	35.82 1	1.152	46	10	183	0
31	76	122	5.285	475	30	11	-1	0
32	116	126	33.17 0	466	11	20	1	0
33	38	139	34.64 8	438	58	12	204	0
34	50	146	8.619	706	75	13	180	0
35	152	109	34.15 7	393	2	22	-21	0
36	83	128	28.40 4	325	18	13	41	0
37	66	148	5.631	784	10	13	122	0
38	83	125	5.446	348	18	15	20	0
39	72	149	6.343	662	13	17	102	0
40	78	128	23.75 3	463	19	15	85	0
41	52	147	2.180	723	15	14	174	0
42	60	166	4.295	1.098	28	11	126	0
43	178	72	29.93 9	350	72	40	33	0
44	75	134	35.26 7	610	101	11	49	0
45	162	105	17.08 4	421	45	21	-38	0
46	40	119	18.21 2	39	19	13	233	0
47	151	111	31.20 5	228	98	29	12	0
48	77	153	15.98 0	2.236	147	8	43	0
49	73	151	3.297	733	63	16	82	0
50	74	153	7.855	914	112	12	81	0
51	93	122	24.34 3	171	48	37	151	1
52	108	117	24.89 9	277	222	31	86	1
53	84	127	14.56 2	136	33	34	163	1
54	93	135	17.12 5	105	19	33	166	1
55	78	127	7.999	257	12	28	146	1
56	95	128	12.61 9	74	73	43	189	1
57	116	135	17.13 8	487	48	27	108	1

58	96	132	25.64 3	371	21	19	51	1
59	127	112	7.010	544	88	26	61	1
60	88	124	18.00 0	31	4	37	175	1
61	79	132	9.000	152	9	30	158	1
62	94	122	18.27 3	214	15	25	104	1
63	118	103	4.378	463	284	28	86	1
64	94	132	26.42 9	144	58	39	183	1
65	122	112	11.10 4	537	40	28	110	1
66	102	117	24.82 0	308	8	22	71	1
67	95	122	1.843	118	9	31	136	1
68	67	128	10.84 3	365	7	21	160	1
69	130	110	21.86 1	159	118	24	65	1
70	116	111	13.50 0	396	162	30	84	1
71	92	136	2.022	94	44	31	183	1
72	115	114	32.55 7	251	138	30	55	1
73	100	123	7.596	208	41	34	135	1
74	103	120	10.90 0	352	46	22	53	1
75	123	114	32.38 8	306	537	28	93	1
76	118	120	19.18 2	225	31	21	52	1
77	138	126	3.922	284	241	29	34	1
78	119	110	3.422	231	65	20	64	1
79	74	136	9.000	204	4	26	147	1
80	80	123	3.400	31	37	33	195	1
81	93	139	15.22 4	135	31	33	170	1
82	88	128	15.59 5	236	139	31	150	1
83	67	134	23.13 4	335	28	24	170	1
84	99	130	27.59 1	386	21	19	52	1
85	104	123	7.701	353	19	22	52	1
86	106	121	13.70 5	417	57	22	44	1

87	91	118	1.324	200	50	25	111	1
88	95	115	7.474	22	8	39	195	1
89	94	115	-1	12	-2	32	153	1
90	118	111	28.69 7	274	88	26	101	1
91	123	108	17.37 9	186	212	25	84	1
92	103	124	31.56 5	906	879	40	136	1
93	124	112	13.24 0	335	22	23	48	1
94	130	110	3.025	187	492	24	65	1
95	108	120	15.76 2	425	13	19	35	1
96	129	122	25.15 7	518	70	31	86	1
97	101	124	9.929	596	191	39	142	1
98	62	124	10.65 0	120	40	20	212	1
99	125	115	8.472	407	257	22	27	1
100	103	119	15.94 4	117	24	29	125	1
101	115	107	8.619	266	11	12	-61	0
102	123	100	25.15 7	457	41	13	-44	0
103	106	112	16.86 9	189	7	13	-23	0
104	69	127	2.790	562	14	9	33	0
105	89	124	1.515	731	197	12	-1	0
106	65	139	11.34 3	732	47	8	61	0
107	64	160	16.35 8	1.950	180	10	89	0
108	38	129	15.05 9	66	45	16	240	0
109	52	146	18.28 6	708	29	13	167	0
110	170	106	31.50 0	611	10	25	-33	0
111	163	77	5.006	549	241	33	15	0
112	148	109	22.90 9	171	14	19	-25	0
113	148	109	25.15 7	365	11	22	-16	0
114	146	109	-1	374	23	23	-15	0
115	135	107	519	182	8	18	-20	0
116	121	121	7.869	98	18	18	-9	0

117	125	113	5.904	381	4	19	3	0
118	123	107	22.29 5	449	28	18	1	0
119	86	111	17.36 6	296	39	13	21	0
120	31	142	3.261	1.038	264	6	197	0
121	30	164	5.404	803	341	6	197	0
122	31	164	25.50 2	732	53	8	204	0
123	38	160	29.80 7	865	24	9	181	0
124	47	152	20.53 5	924	15	11	171	0
125	53	154	11.44 4	1.166	26	11	146	0
126	56	154	24.34 3	1.127	22	11	140	0
127	62	147	26.89 9	730	82	15	143	0
128	61	152	17.51 8	761	60	14	138	0
129	61	168	14.69 8	940	34	11	125	0
130	67	152	18.58 3	793	35	14	112	0
131	75	142	3.254	457	89	15	83	0
132	77	140	19.84 3	392	2	13	61	0
133	109	106	27.84 3	426	78	18	16	0
134	116	106	35.72 7	403	30	17	10	0
135	108	103	16.98 2	357	57	17	21	0
136	85	126	27.89 7	373	14	15	41	0
137	82	127	9.000	306	6	14	57	0
138	80	132	3.749	478	225	13	54	0
139	68	146	21.48 2	563	40	15	121	0
140	59	150	14.18 8	853	59	14	142	0
141	52	128	14.38 6	688	112	15	195	0
142	145	108	6.071	398	67	28	-2	0
143	146	100	14.06 6	520	216	30	-14	0

144	122	123	5.019	230	11	19	-4	0
145	127	127	24.55 6	384	69	20	-6	0
146	114	112	18.60 1	440	55	18	12	0
147	130	119	10.13 1	205	7	19	-11	0
148	118	125	8.366	143	6	19	1	0
149	143	119	34.70 1	233	10	23	6	0
150	162	87	12.04 7	317	28	39	24	0
151	71	130	17.42 9	136	14	28	168	1
152	80	126	33.37 8	274	53	24	123	1
153	86	118	10.13 1	208	4	26	122	1
154	108	112	15.82 0	339	31	19	26	1
155	90	118	2.597	305	31	24	101	1
156	103	121	10.84 3	304	20	19	54	1
157	97	117	9.671	227	245	20	68	1
158	91	126	11.31 0	396	106	29	133	1
159	87	126	7.157	21	2	39	182	1
160	88	127	10.59 5	106	31	33	155	1
161	91	128	6.585	149	46	32	143	1
162	97	114	31.81 8	198	18	25	97	1
163	98	114	21.84 2	260	53	24	86	1
164	101	120	18.71 3	248	12	21	67	1
165	104	120	5.071	238	10	21	61	1
166	91	120	21.31 1	113	59	35	153	1
167	94	127	25.44 8	118	27	32	143	1
168	95	129	30.49 9	136	17	36	143	1
169	93	131	29.11 9	33	38	41	183	1
170	98	121	18.86 2	150	72	31	138	1

171	99	120	4.758	160	113	34	147	1
172	103	120	33.32 2	248	172	29	115	1
173	93	131	21.12 2	172	55	38	170	1
174	127	109	1.493	258	178	21	59	1
175	103	132	29.77 6	380	184	36	134	1
176	96	122	12.65 1	339	254	38	155	1
177	113	121	8.366	274	6	25	98	1
178	95	128	30.36 9	27	10	38	201	1
179	94	127	31.50 0	35	13	36	200	1
180	94	137	24.34 3	161	16	32	158	1
181	93	120	18.44 0	33	9	34	222	1
182	93	139	25.02 0	91	38	33	170	1
183	118	122	10.19 1	728	536	28	106	1
184	105	129	10.82 1	343	115	30	124	1
185	129	106	9.337	424	85	28	88	1
186	98	123	20.45 9	152	93	29	138	1
187	91	127	1.345	14	123	35	169	1
188	110	125	12.45 5	429	290	27	124	1
189	108	135	30.24 8	707	803	27	125	1
190	135	120	19.91 3	518	334	27	64	1
191	94	147	123	30	12	31	140	1
192	94	130	28.84 3	13	34	33	152	1
193	129	99	24.28 5	212	63	27	85	1
194	131	116	9.523	539	188	28	55	1
195	122	128	26.40 9	441	334	30	60	1
196	104	104	11.75 3	205	152	32	94	1
197	150	100	13.09 4	296	186	35	37	1

198	110	125	19.87 8	153	76	28	71	1
199	128	109	33.54 5	185	128	26	34	1
200	116	109	27.06 0	213	136	30	49	1

Tabela A.3 – Dados originais da *Strix Varia*.

	x1	x2	x3	x4	x5	x6	x7	y
1	0,8750	0,2955	0,2513	0,1351	0,0046	0,5789	0,1528	0
2	0,7697	0,3636	0,4243	0,1207	0,0432	0,5263	0,2425	0
3	0,9868	0,0227	0,2069	0,0941	0,1047	0,8421	0,3189	0
4	0,1382	0,6818	0,6017	0,4255	0,0137	0,1316	0,7010	0
5	0,8816	0,2500	0,2348	0,1170	0,0239	0,6053	0,1429	0
6	0,2961	1,0000	0,2449	1,0000	0,5097	0,2368	0,2226	0
7	0,1250	0,6970	0,6952	0,5082	0,6826	0,2105	0,7276	0
8	0,1184	0,4242	0,4984	0,0555	0,0319	0,2368	0,9468	0
9	0,0395	0,4697	0,2102	0,0914	0,1832	0,1316	0,9468	0
10	0,0132	0,8485	0,6358	0,4309	0,1684	0,0263	0,8439	0
11	0,9408	0,2273	0,6913	0,2047	0,0148	0,5000	0,0963	0
12	0,1184	0,4697	0,2599	0,1590	0,1365	0,1579	0,8837	0
13	0,1250	0,4242	0,0259	0,0548	0,0353	0,2368	0,9369	0
14	0,6974	0,3788	0,7023	0,0985	0,0023	0,3684	0,1628	0
15	0,7500	0,3182	0,7919	0,1294	0,0307	0,5263	0,2060	0
16	0,1908	0,6061	0,3377	0,3546	0,0330	0,1842	0,6777	0
17	0,0000	0,7576	0,0297	0,2867	0,0262	0,0000	0,8904	0
18	0,8224	0,3106	0,3769	0,1294	0,0057	0,6316	0,1860	0
19	0,3289	0,5379	0,3366	0,1855	0,0830	0,2105	0,4485	0
20	0,1382	0,4848	0,5355	0,0884	0,1468	0,4737	0,9502	0
21	0,8750	0,2348	0,9248	0,1879	0,0580	0,5263	0,0897	0
22	0,3553	0,3409	0,9535	0,0766	0,0023	0,2105	0,2990	0
23	0,7763	0,3030	0,3399	0,0894	0,0398	0,4211	0,1528	0
24	0,2763	0,7045	0,3274	0,2629	0,0876	0,2368	0,5449	0
25	0,1316	0,5985	0,9441	0,2716	0,0705	0,1579	0,7973	0
26	0,9605	0,2500	0,6281	0,2010	0,0011	0,5000	0,0532	0
27	0,6974	0,2576	0,9308	0,0985	0,0034	0,3158	0,1528	0
28	0,5789	0,2955	0,2552	0,1237	0,0330	0,2895	0,2159	0
29	0,2697	0,6212	0,6582	0,3237	0,0489	0,1842	0,5548	0
30	0,0789	0,6818	1,0000	0,3832	0,0523	0,1316	0,8106	0
31	0,3289	0,3788	0,1476	0,1556	0,0341	0,1579	0,1993	0
32	0,5921	0,4091	0,9260	0,1526	0,0125	0,3947	0,2060	0
33	0,0789	0,5076	0,9673	0,1432	0,0660	0,1842	0,8804	0
34	0,1579	0,5606	0,2406	0,2333	0,0853	0,2105	0,8007	0
35	0,8289	0,2803	0,9535	0,1281	0,0023	0,4474	0,1329	0
36	0,3750	0,4242	0,7929	0,1052	0,0205	0,2105	0,3389	0

37	0,2632	0,5758	0,1572	0,2595	0,0114	0,2105	0,6080	0
38	0,3750	0,4015	0,1521	0,1129	0,0205	0,2632	0,2691	0
39	0,3026	0,5833	0,1771	0,2185	0,0148	0,3158	0,5415	0
40	0,3421	0,4242	0,6631	0,1516	0,0216	0,2632	0,4850	0
41	0,1711	0,5682	0,0609	0,2390	0,0171	0,2368	0,7807	0
42	0,2237	0,7121	0,1199	0,3650	0,0319	0,1579	0,6213	0
43	1,0000	0,0000	0,8358	0,1136	0,0819	0,9211	0,3123	0
44	0,3224	0,4697	0,9845	0,2010	0,1149	0,1579	0,3654	0
45	0,8947	0,2500	0,4769	0,1375	0,0512	0,4211	0,0764	0
46	0,0921	0,3561	0,5084	0,0091	0,0216	0,2105	0,9767	0
47	0,8224	0,2955	0,8711	0,0726	0,1115	0,6316	0,2425	0
48	0,3355	0,6136	0,4461	0,7476	0,1672	0,0789	0,3455	0
49	0,3092	0,5985	0,0921	0,2424	0,0717	0,2895	0,4751	0
50	0,3158	0,6136	0,2193	0,3032	0,1274	0,1842	0,4718	0
51	0,4408	0,3788	0,6796	0,0534	0,0546	0,8421	0,7043	1
52	0,5395	0,3409	0,6951	0,0891	0,2526	0,6842	0,4884	1
53	0,3816	0,4167	0,4065	0,0417	0,0375	0,7632	0,7442	1
54	0,4408	0,4773	0,4781	0,0313	0,0216	0,7368	0,7542	1
55	0,3421	0,4167	0,2233	0,0824	0,0137	0,6053	0,6877	1
56	0,4539	0,4242	0,3523	0,0208	0,0830	1,0000	0,8306	1
57	0,5921	0,4773	0,4784	0,1597	0,0546	0,5789	0,5615	1
58	0,4605	0,4545	0,7159	0,1207	0,0239	0,3684	0,3721	1
59	0,6645	0,3030	0,1957	0,1788	0,1001	0,5526	0,4053	1
60	0,4079	0,3939	0,5025	0,0064	0,0046	0,8421	0,7841	1
61	0,3487	0,4545	0,2513	0,0471	0,0102	0,6579	0,7276	1
62	0,4474	0,3788	0,5101	0,0679	0,0171	0,5263	0,5482	1
63	0,6053	0,2348	0,1222	0,1516	0,3231	0,6053	0,4884	1
64	0,4474	0,4545	0,7378	0,0444	0,0660	0,8947	0,8106	1
65	0,6316	0,3030	0,3100	0,1765	0,0455	0,6053	0,5681	1
66	0,5000	0,3409	0,6929	0,0995	0,0091	0,4474	0,4385	1
67	0,4539	0,3788	0,0515	0,0356	0,0102	0,6842	0,6545	1
68	0,2697	0,4242	0,3027	0,1187	0,0080	0,4211	0,7342	1
69	0,6842	0,2879	0,6103	0,0494	0,1342	0,5000	0,4186	1
70	0,5921	0,2955	0,3769	0,1291	0,1843	0,6579	0,4817	1
71	0,4342	0,4848	0,0565	0,0276	0,0501	0,6842	0,8106	1
72	0,5855	0,3182	0,9089	0,0803	0,1570	0,6579	0,3854	1
73	0,4868	0,3864	0,2121	0,0659	0,0466	0,7632	0,6512	1
74	0,5066	0,3636	0,3043	0,1143	0,0523	0,4474	0,3787	1
75	0,6382	0,3182	0,9042	0,0988	0,6109	0,6053	0,5116	1
76	0,6053	0,3636	0,5355	0,0716	0,0353	0,4211	0,3754	1
77	0,7368	0,4091	0,1095	0,0914	0,2742	0,6316	0,3156	1
78	0,6118	0,2879	0,0956	0,0736	0,0739	0,3947	0,4153	1
79	0,3158	0,4848	0,2513	0,0645	0,0046	0,5526	0,6910	1
80	0,3553	0,3864	0,0949	0,0064	0,0421	0,7368	0,8505	1
81	0,4408	0,5076	0,4250	0,0413	0,0353	0,7368	0,7674	1
82	0,4079	0,4242	0,4354	0,0753	0,1581	0,6842	0,7010	1

83	0,2697	0,4697	0,6458	0,1086	0,0319	0,5000	0,7674	1
84	0,4803	0,4394	0,7703	0,1257	0,0239	0,3684	0,3754	1
85	0,5132	0,3864	0,2150	0,1146	0,0216	0,4474	0,3754	1
86	0,5263	0,3712	0,3826	0,1361	0,0648	0,4474	0,3488	1
87	0,4276	0,3485	0,0370	0,0632	0,0569	0,5263	0,5714	1
88	0,4539	0,3258	0,2087	0,0034	0,0091	0,8947	0,8505	1
89	0,4474	0,3258	0,0000	0,0000	0,0000	0,7105	0,7110	1
90	0,6053	0,2955	0,8011	0,0881	0,1001	0,5526	0,5382	1
91	0,6382	0,2727	0,4852	0,0585	0,2412	0,5263	0,4817	1
92	0,5066	0,3939	0,8812	0,3005	1,0000	0,9211	0,6545	1
93	0,6447	0,3030	0,3696	0,1086	0,0250	0,4737	0,3621	1
94	0,6842	0,2879	0,0845	0,0588	0,5597	0,5000	0,4186	1
95	0,5395	0,3636	0,4400	0,1388	0,0148	0,3684	0,3189	1
96	0,6776	0,3788	0,7023	0,1701	0,0796	0,6842	0,4884	1
97	0,4934	0,3939	0,2772	0,1963	0,2173	0,8947	0,6744	1
98	0,2368	0,3939	0,2973	0,0363	0,0455	0,3947	0,9070	1
99	0,6513	0,3258	0,2365	0,1328	0,2924	0,4474	0,2924	1
100	0,5066	0,3561	0,4451	0,0353	0,0273	0,6316	0,6179	1
101	0,5855	0,2652	0,2406	0,0854	0,0125	0,1842	0,0000	0
102	0,6382	0,2121	0,7023	0,1496	0,0466	0,2105	0,0565	0
103	0,5263	0,3030	0,4709	0,0595	0,0080	0,2105	0,1262	0
104	0,2829	0,4167	0,0779	0,1849	0,0159	0,1053	0,3123	0
105	0,4145	0,3939	0,0423	0,2417	0,2241	0,1842	0,1993	0
106	0,2566	0,5076	0,3167	0,2420	0,0535	0,0789	0,4053	0
107	0,2500	0,6667	0,4567	0,6514	0,2048	0,1316	0,4983	0
108	0,0789	0,4318	0,4204	0,0182	0,0512	0,2895	1,0000	0
109	0,1711	0,5606	0,5105	0,2339	0,0330	0,2105	0,7575	0
110	0,9474	0,2576	0,8794	0,2013	0,0114	0,5263	0,0930	0
111	0,9013	0,0379	0,1398	0,1805	0,2742	0,7368	0,2525	0
112	0,8026	0,2803	0,6396	0,0534	0,0159	0,3684	0,1196	0
113	0,8026	0,2803	0,7023	0,1187	0,0125	0,4474	0,1495	0
114	0,7895	0,2803	0,0000	0,1217	0,0000	0,4737	0,1528	0
115	0,7171	0,2652	0,0145	0,0571	0,0091	0,3421	0,1362	0
116	0,6250	0,3712	0,2197	0,0289	0,0205	0,3421	0,1728	0
117	0,6513	0,3106	0,1648	0,1240	0,0046	0,3684	0,2126	0
118	0,6382	0,2652	0,6224	0,1469	0,0319	0,3421	0,2060	0
119	0,3947	0,2955	0,4848	0,0955	0,0444	0,2105	0,2724	0
120	0,0329	0,5303	0,0911	0,3449	0,3003	0,0263	0,8571	0
121	0,0263	0,6970	0,1509	0,2659	0,3879	0,0263	0,8571	0
122	0,0329	0,6970	0,7119	0,2420	0,0603	0,0789	0,8804	0
123	0,0789	0,6667	0,8321	0,2867	0,0273	0,1053	0,8040	0
124	0,1382	0,6061	0,5733	0,3066	0,0171	0,1579	0,7708	0
125	0,1776	0,6212	0,3195	0,3879	0,0296	0,1579	0,6877	0
126	0,1974	0,6212	0,6796	0,3748	0,0250	0,1579	0,6678	0
127	0,2368	0,5682	0,7509	0,2413	0,0933	0,2632	0,6777	0
128	0,2303	0,6061	0,4891	0,2518	0,0683	0,2368	0,6611	0

129	0,2303	0,7273	0,4103	0,3119	0,0387	0,1579	0,6179	0
130	0,2697	0,6061	0,5188	0,2625	0,0398	0,2368	0,5748	0
131	0,3224	0,5303	0,0909	0,1496	0,1013	0,2632	0,4784	0
132	0,3355	0,5152	0,5540	0,1277	0,0023	0,2105	0,4053	0
133	0,5461	0,2576	0,7773	0,1392	0,0887	0,3421	0,2558	0
134	0,5921	0,2576	0,9974	0,1314	0,0341	0,3158	0,2359	0
135	0,5395	0,2348	0,4741	0,1160	0,0648	0,3158	0,2724	0
136	0,3882	0,4091	0,7788	0,1213	0,0159	0,2632	0,3389	0
137	0,3684	0,4167	0,2513	0,0988	0,0068	0,2368	0,3920	0
138	0,3553	0,4545	0,1047	0,1566	0,2560	0,2105	0,3821	0
139	0,2763	0,5606	0,5997	0,1852	0,0455	0,2632	0,6047	0
140	0,2171	0,5909	0,3961	0,2827	0,0671	0,2368	0,6744	0
141	0,1711	0,4242	0,4016	0,2272	0,1274	0,2632	0,8505	0
142	0,7829	0,2727	0,1695	0,1297	0,0762	0,6053	0,1960	0
143	0,7895	0,2121	0,3927	0,1708	0,2457	0,6579	0,1561	0
144	0,6316	0,3864	0,1401	0,0733	0,0125	0,3684	0,1894	0
145	0,6645	0,4167	0,6855	0,1250	0,0785	0,3947	0,1827	0
146	0,5789	0,3030	0,5193	0,1439	0,0626	0,3421	0,2425	0
147	0,6842	0,3561	0,2828	0,0649	0,0080	0,3684	0,1661	0
148	0,6053	0,4015	0,2336	0,0440	0,0068	0,3684	0,2060	0
149	0,7697	0,3561	0,9687	0,0743	0,0114	0,4737	0,2226	0
150	0,8947	0,1136	0,3363	0,1025	0,0319	0,8947	0,2824	0
151	0,2961	0,4394	0,4866	0,0417	0,0159	0,6053	0,7608	1
152	0,3553	0,4091	0,9318	0,0881	0,0603	0,5000	0,6113	1
153	0,3947	0,3485	0,2828	0,0659	0,0046	0,5526	0,6080	1
154	0,5395	0,3030	0,4417	0,1099	0,0353	0,3684	0,2890	1
155	0,4211	0,3485	0,0725	0,0985	0,0353	0,5000	0,5382	1
156	0,5066	0,3712	0,3027	0,0982	0,0228	0,3684	0,3821	1
157	0,4671	0,3409	0,2700	0,0723	0,2787	0,3947	0,4286	1
158	0,4276	0,4091	0,3158	0,1291	0,1206	0,6316	0,6445	1
159	0,4013	0,4091	0,1998	0,0030	0,0023	0,8947	0,8073	1
160	0,4079	0,4167	0,2958	0,0316	0,0353	0,7368	0,7176	1
161	0,4276	0,4242	0,1839	0,0461	0,0523	0,7105	0,6777	1
162	0,4671	0,3182	0,8883	0,0625	0,0205	0,5263	0,5249	1
163	0,4737	0,3182	0,6098	0,0834	0,0603	0,5000	0,4884	1
164	0,4934	0,3636	0,5224	0,0793	0,0137	0,4211	0,4252	1
165	0,5132	0,3636	0,1416	0,0760	0,0114	0,4211	0,4053	1
166	0,4276	0,3636	0,5949	0,0339	0,0671	0,7895	0,7110	1
167	0,4474	0,4167	0,7104	0,0356	0,0307	0,7105	0,6777	1
168	0,4539	0,4318	0,8514	0,0417	0,0193	0,8158	0,6777	1
169	0,4408	0,4470	0,8129	0,0071	0,0432	0,9474	0,8106	1
170	0,4737	0,3712	0,5266	0,0464	0,0819	0,6842	0,6611	1
171	0,4803	0,3636	0,1329	0,0497	0,1286	0,7632	0,6910	1
172	0,5066	0,3636	0,9302	0,0793	0,1957	0,6316	0,5847	1
173	0,4408	0,4470	0,5897	0,0538	0,0626	0,8684	0,7674	1
174	0,6645	0,2803	0,0417	0,0827	0,2025	0,4211	0,3987	1

175	0,5066	0,4545	0,8312	0,1237	0,2093	0,8158	0,6478	1
176	0,4605	0,3788	0,3532	0,1099	0,2890	0,8684	0,7176	1
177	0,5724	0,3712	0,2336	0,0881	0,0068	0,5263	0,5282	1
178	0,4539	0,4242	0,8478	0,0050	0,0114	0,8684	0,8704	1
179	0,4474	0,4167	0,8794	0,0077	0,0148	0,8158	0,8671	1
180	0,4474	0,4924	0,6796	0,0501	0,0182	0,7105	0,7276	1
181	0,4408	0,3636	0,5148	0,0071	0,0102	0,7632	0,9402	1
182	0,4408	0,5076	0,6985	0,0266	0,0432	0,7368	0,7674	1
183	0,6053	0,3788	0,2845	0,2407	0,6098	0,6053	0,5548	1
184	0,5197	0,4318	0,3021	0,1113	0,1308	0,6579	0,6146	1
185	0,6776	0,2576	0,2607	0,1385	0,0967	0,6053	0,4950	1
186	0,4737	0,3864	0,5712	0,0471	0,1058	0,6316	0,6611	1
187	0,4276	0,4167	0,0376	0,0007	0,1399	0,7895	0,7641	1
188	0,5526	0,4015	0,3477	0,1402	0,3299	0,5789	0,6146	1
189	0,5395	0,4773	0,8444	0,2336	0,9135	0,5789	0,6179	1
190	0,7171	0,3636	0,5559	0,1701	0,3800	0,5789	0,4153	1
191	0,4474	0,5682	0,0035	0,0061	0,0137	0,6842	0,6678	1
192	0,4474	0,4394	0,8052	0,0003	0,0387	0,7368	0,7076	1
193	0,6776	0,2045	0,6780	0,0672	0,0717	0,5789	0,4850	1
194	0,6908	0,3333	0,2659	0,1771	0,2139	0,6053	0,3854	1
195	0,6316	0,4242	0,7373	0,1442	0,3800	0,6579	0,4020	1
196	0,5132	0,2424	0,3281	0,0649	0,1729	0,7105	0,5150	1
197	0,8158	0,2121	0,3656	0,0955	0,2116	0,7895	0,3256	1
198	0,5526	0,4015	0,5549	0,0474	0,0865	0,6053	0,4385	1
199	0,6711	0,2803	0,9365	0,0582	0,1456	0,5526	0,3156	1
200	0,4725	0,4074	0,4587	0,1392	0,0974	0,4678	0,5096	0

Tabela A.4 - Dados resultantes de *Strix Varia*