



Douglas Diniz Landim

**ANÁLISE DE DEMANDA VIA INTELIGÊNCIA
ARTIFICIAL NO RESTAURANTE UNIVERSITÁRIO DO
INSTITUTO DE CIÊNCIA E TECNOLOGIA DA UNIFESP**

São José dos Campos, SP

Douglas Diniz Landim

ANÁLISE DE DEMANDA VIA INTELIGÊNCIA ARTIFICIAL NO RESTAURANTE UNIVERSITÁRIO DO INSTITUTO DE CIÊNCIA E TECNOLOGIA DA UNIFESP

Trabalho de conclusão de curso apresentado ao
Instituto de Ciência e Tecnologia – UNIFESP,
como parte das atividades para obtenção do tí-
tulo de Bacharel em Ciência da Computação.

Universidade Federal de São Paulo – UNIFESP

Instituto de Ciência e Tecnologia

Bacharelado em Ciência da Computação

Orientador: Prof. Dr. Vinícius Veloso de Melo

São José dos Campos, SP

Dezembro de 2018

Douglas Diniz Landim

ANÁLISE DE DEMANDA VIA INTELIGÊNCIA ARTIFICIAL NO RESTAURANTE UNIVERSITÁRIO DO INSTITUTO DE CIÊNCIA E TECNOLOGIA DA UNIFESP

Trabalho de conclusão de curso apresentado ao
Instituto de Ciência e Tecnologia – UNIFESP,
como parte das atividades para obtenção do tí-
tulo de Bacharel em Ciência da Computação.

Trabalho para apresentar em DEZEMBRO/2018:

Prof. Dr. Vinícius Veloso de Melo
Orientador

Professor
Convidado 1

Professor
Convidado 2

Professor
Convidado 3

São José dos Campos, SP
Dezembro de 2018

Este trabalho é dedicado aos meus pais que apoiaram e sacrificaram esforços para me manter ativo nessa jornada, a todos os professores que me somaram conhecimentos, oportunidades e esperanças indo além de suas rotinas e agendas em prol do ensino, e principalmente à todos que me motivaram me oferecendo desafios para que eu pudesse enfrentá-los superando meus próprios limites

Agradecimentos

Minha jornada pela graduação foi marcada por muita persistência, dificuldades e fracassos. Agradeço primeiramente a Deus por me dar fé e alimentar minha persistência e esperança. Apesar de todo o conteúdo técnico das mais de 40 disciplinas do meu curso, o que mais me agregou aprendizado foi o ambiente desafiador desta universidade; que somado à muitas dificuldades pessoais, acidentes, contra-tempos de saúde, profissão e família; constituiu o conjunto perfeito de desafios que me derrubaram muitas vezes e me fizeram ser uma pessoa melhor ao me levantar, mais convicto e perseverante a cada nova tentativa de conquistar minhas aprovações. Agradeço à minha família por sempre me apoiar dando tudo de si, a meu professor por me aceitar orientar e me motivar sempre me dando atenção como um amigo nas conversas no fim da aula durante o caminho até o estacionamento da faculdade, nas reuniões e chats online até nos finais de semana, aos amigos universitários, e a todos os professores que me acompanharam e ofereceram desafios em todos esses anos na UNIFESP.

*Mesmo desacreditado e ignorado por todos, não posso desistir, pois para mim, vencer é nunca
desistir.
(Albert Einstein)*

Resumo

O presente trabalho tem como objetivo o estudo de métodos para a previsão de vendas do restaurante universitário da Unifesp para evitar super-projeção de demanda com consequência de desperdício de alimentos, ou subprojeção com consequência de docentes ou discentes sem refeições. Em uma investigação anterior, realizada como trabalho de disciplina, o autor deste trabalho empregou métodos estatísticos para a construção do modelo de previsão. Neste trabalho, será empregado um método de aprendizado de máquina, mais especificamente, uma rede neural artificial do tipo Perceptron de Múltiplas Camadas. Na investigação anterior, verificou-se que a temperatura da região onde se localiza o campus do restaurante foi analisada por recorrência via *bootstrap* como um fator que exerce impacto sobre as vendas do restaurante em certos períodos do semestre. Assim, neste trabalho de conclusão de curso serão obtidas novas variáveis e um intervalo maior de amostragem na análise da previsão de demanda.

Palavras-chave: Redes Neurais Artificiais, Previsão de demanda, Aprendizado de Máquina, Inteligência Artificial, Perceptron Múltiplas camadas.

Abstract

This work has the objective of studying methods to forecast meals demand on Unifesp's university restaurant to avoid overprojection of demand as a result of food wastage, or underproduction with the consequence of teachers or students without meals. In an earlier research, carried out as a discipline work, the author of this paper used statistical methods for the construction of the prediction model. In this work, a machine learning method will be employed, more specifically, an artificial neural network of the Perceptron type of Multiple Layers. In the previous investigation, it was verified that the temperature of the region where the restaurant's campus is located was analyzed by recurrence via bootstrap as a factor that has an impact on restaurant sales in certain periods of the semester. Thus, in this work of conclusion of course will be obtained new variables and a greater interval of sampling in the analysis of the forecast of demand..

Keywords: Artificial Neural Networks, Demand Prediction, Machine Learning, Artificial intelligence, Perceptron Multiple layers.

Lista de ilustrações

Figura 1 – Tipos de métodos quantitativos Retirado de (JUNIOR, 2007).	28
Figura 2 – Métodos de previsão de demanda (JUNIOR, 2007).	29
Figura 3 – DADOS DE DEMANDA DE SAZONALIDADE DIÁRIA, EM FUNÇÃO DO HORÁRIO. Retirado de (RUAS, 2012).	30
Figura 4 – Neurônio Biológico, retirado de https://pt.khanacademy.org/science/biology/human-biology/neuron-nervous-system/v/anatomy-of-a-neuron	34
Figura 5 – Neurônio Artificial, Retirado de http://redesneuraisartificiais.blogspot.com/2010/10/o-primeiro-modelo-de-um-neuronio-criado.html	34
Figura 6 – Neurônio Artificial - Operação AND, Retirado de http://redesneuraisartificiais.blogspot.com/2010/10/o-primeiro-modelo-de-um-neuronio-criado.html	35
Figura 7 – Neurônio Artificial Perceptron, retirado de (JUNIOR, 2007) ()	36
Figura 8 – Funções de Ativação retirado de (MARIANO, 2014)	37
Figura 9 – Problemas Linearmente Separáveis e Não Separáveis. Retirado de (MARIANO, 2014).	44
Figura 10 – Problemas Linearmente Não Separáveis Múltiplas Soluções.	44
Figura 11 – Rede de perceptrons com múltiplas camadas. Retirado de (ALMEIDA, 2013)	45
Figura 12 – Ponto ótimo de parada da validação cruzada. Retirado de (MARIANO, 2014)	47
Figura 13 – Fases de treino da MLP-Back-Propagation. Retirado de (ALMEIDA, 2013)	48
Figura 14 – Rede Neural Perceptron de Múltiplas Camadas. Retirado de (LOPES, 2008).	48
Figura 15 – Rede Neural Perceptron de Múltiplas Camadas. Retirado de (ROCHA JOSÉ CELSO; MATOS, 2011)	49
Figura 16 – Apresentação da i-ésima observação do conjunto de treino à rede	51
Figura 17 – Cálculo do valor estimado de saída da rede.	51
Figura 18 – Aprendizado de um neurônio na camada de saída	52
Figura 19 – Reajuste dos pesos de saída	52
Figura 20 – Reajuste dos pesos de camadas ocultas	53
Figura 21 – Proposta de RNA, Figura gerada a partir da ferramenta online: http://alexlenail.me/NN-SVG/index.html	62
Figura 22 – Entradas de dia da semana em cofatores. Retirado de (LOPES, 2008).	67

Lista de tabelas

Tabela 1 – Representação de dados para regressão múltipla.	31
Tabela 2 – Operação Lógica AND	35
Tabela 3 – Plano de atividades para o TCC II	61
Tabela 4 – Tabela de dados do R.U.	68

Lista de abreviaturas e siglas

ICT	Instituto Central de Tecnologia
R.U.	Restaurante Universitário
UNIFESP	Universidade Federal de São Paulo
UFV	Universidade Federal de Viçosa
UNESP	Universidade Estadual Paulista Júlio de Mesquita Filho
BDMEP	Banco de Dados Meteorológicos para Ensino e Pesquisa
RNA	Rede Neural Artificial
PMC	Perceptron de Múltiplas Camadas

Sumário

1	Introdução	21
1.1	Contextualização e Motivação	21
1.2	Definição do problema	23
1.3	Justificativas	23
1.4	Objetivos	23
1.4.1	Objetivo geral	23
1.4.2	Objetivos específicos	23
1.5	Metodologia	24
1.6	Organização do documento	25
2	Fundamentação Teórica	27
2.1	Análises Estatísticas	27
2.1.1	Análise Exploratória dos Dados	27
2.1.2	Métodos de Previsão	27
2.1.3	Métodos de previsão de Demanda	28
2.1.4	Séries Temporais	28
2.1.5	Componentes Temporais	29
2.1.6	Previsão com Regressão Linear Múltipla	30
2.1.7	Algoritmo de Regressão Linear Múltipla	32
2.2	Inteligencia Artificial	32
2.2.1	Neurônio Artificial	33
2.2.2	Redes Neurais Artificiais	35
2.2.3	Funcionamento e treino do neurônio perceptron	38
2.2.4	Limite de um Perceptron	43
2.2.5	Rede Perceptrons Múltiplas Camadas - MLP	44
2.2.6	Perceptrons Múltiplas Camadas com Retro Propagação de Erro	46
2.2.7	Treino da MLP - BackPropagation para 1 sinal de saída	49
2.2.8	Algoritmo de treino e validação backpropagation	54
3	Trabalhos relacionados	57
3.0.1	ANÁLISES EM RESTAURANTES UNIVERSITÁRIOS	57
3.0.2	ANÁLISES EM OUTROS CENÁRIOS	58
3.0.3	TESE COMPARATIVA DE DIVERSOS MÉTODOS DE PREVISÃO DE DEMANDA.	58
4	Plano de atividades para o TCC II	61

4.1	Dados	61
4.1.1	Dados de Consumo e de Data.	62
4.1.2	Dados Climáticos	64
4.1.3	Estruturação dos Dados	65
5	Conclusão	69
	Referências	71

1 Introdução

1.1 Contextualização e Motivação

A previsão da demanda é o fator principal da eficiência de qualquer modelo de processamento do tipo entrada-saída, onde a sua saída deve atender uma demanda não determinística. É necessário prever a demanda para projetar e aperfeiçoar o processamento e a entrada deste modelo. Portanto, a previsão de demanda é um ponto de extrema importância para qualquer empresa e, sendo adequada, permite o ajuste de todo o mecanismo de operações da empresa para atendê-la com a melhor eficiência possível, maximizando lucros, minimizando perdas e, principalmente, atendendo todas as necessidades do cliente.

Todo restaurante universitário enfrenta problemas de previsão de demanda de refeições e prejuízos com a falta de vendas e ou o descarte de refeições não vendidas. Um dos grandes problemas enfrentados hoje no mundo é a elevação dos preços dos alimentos. O valor, além de monetário, é moral pois a alimentação é o recurso primitivo de base da humanidade que hoje ainda enfrenta um difícil acesso a este em muitas regiões carentes. O descarte de alimentos, provocado por suas limitações e durabilidade, gera não somente prejuízos monetários, mas ambientais e morais. Isso tem causado preocupações para a população em geral e também para empresas como restaurantes que sofrem diretamente os reflexos da variação no preço dos alimentos e na demanda. Atualmente, o restaurante universitário do ICT - UNIFESP não possui um sistema que ajude na gestão de compras dos alimentos e sofre dos problemas acima relatados.

No restaurante universitário do ICT - UNIFESP, as refeições são fornecidas de segunda a sexta-feira. O caso particular de restaurantes universitários envolve um fluxo de demanda influenciado pelo dia da semana, visto que a demanda é influenciada pela quantia de alunos presentes na universidade, que por sua vez é influenciada pela grade de aulas determinada semestralmente por dia da semana. O caso de análise para este projeto foi motivado após informações de relevantes desperdício de alimentos.

Devido às condições burocráticas no ambiente do restaurante que compreendem fidelidade de contrato e exclusividade, pois o restaurante se encontra em localização que o faz ser o único provedor de refeições ao público do campus, a escolha dos parâmetros não será influenciada por muitos fatores externos como concorrência, acessibilidade do ponto, entre outros.

Outro ponto importante é a obtenção dos valores de venda; não foram escolhidas as vendas diretas do ponto de venda de tickets de refeição, e sim os dados de coleta da entrada do restaurante, que demonstram a real movimentação de público no restaurante em determinado dia.

O estudo da relação de vendas, temperatura, outras variáveis climáticas e do ambiente, já é comum em outros cenários; dentre eles, o de maior destaque é na demanda de energia elétrica. Os cenários de vendas de alimentos perecíveis ganha também destaque apesar de se encontrar investimentos maiores de previsão de demanda na indústria de produção de energia elétrica. O objetivo, tanto no cenário deste trabalho - o restaurante universitário, como em outros cenários é o mesmo: atender toda a demanda de consumo e evitar transtorno a qualquer consumidor pela falta do produto desejado e evitar prejuízos de produção não consumida. Tais prejuízos impactam não só o fornecedor, mas o consumidor, visto que um fornecimento de produto e serviço com um bom planejamento de demanda poupa recursos ao produtor. Tais recursos podem ser investidos em melhor qualidade de produto e menor preço ao consumidor caso seja obtido um modelo de previsão viável para evitar sobrestimação de demanda com consequência de desperdício de alimentos, ou subestimação com consequência de docentes ou discentes sem refeições.

Tal problema tem sido impactante e frequente para o restaurante que informa que em alguns dias no mês passa por sobrestimação e desperdício superior a 30% do total produzido.

O modelo a ser analisado neste trabalho é o comportamento dos consumidores de um restaurante universitário (R.U), o qual precisa projetar sua compra de insumos e alocação de recursos na entrada de seu modelo de negócio, e projetar sua saída, que é a produção de refeições em quantidade numérica e inteira distribuída em função do tempo em dias. O consumo feito por alunos não se comporta de maneira determinística, já que este consumo é facultativo. Portanto, nem todos os alunos que estão na universidade se alimentarão no R.U. todos os dias.

De acordo com o contrato presente entre o R.U e a universidade, o mesmo deve atender totalmente à demanda do público, sendo multado se caso algum consumidor fique sem alimentação. Porém, este mesmo contrato não trata refeições que não são consumidas; logo, o R.U. deve lidar integralmente o prejuízo de refeições produzidas acima da demanda de consumo. O que o R.U. faz com essas sobras está fora do escopo deste trabalho.

Tais refeições fornecidas aos alunos são, em parte, subsidiadas pela universidade. O período de dados obtido para este trabalho compreende até Agosto de 2018 e segue um modelo de contrato antigo no qual o restaurante recebe R\$2,50 do aluno e R\$9,14 da universidade totalizando R\$11,64 por refeição.

De agosto em diante, iniciando no segundo semestre de 2018, a universidade subsidia R\$5,44 da refeição do aluno e o mesmo R\$2,50, o R.U. recebe o total de R\$7,94. Logo, esta previsão de demanda corresponde também aos interesses da administração do campus local, que periodicamente deve realizar uma alocação de recursos financeiros para subsidiar todas estas refeições consumidas.

É necessário então entender e descobrir quais elementos influenciam este consumo humano, de que forma estes elementos exercem tal influência e em qual intensidade ela ocorre.

E para explorar tais elementos, é necessário explorar como este consumo ocorreu historicamente a fim de se encontrar padrões de comportamento, que no cenário deste trabalho acontece com informações distribuídas através do tempo.

Neste capítulo o principal objetivo é definir todos os conceitos que serão necessários para o entendimento total deste trabalho, analisando a forma que os dados são entendidos, e as ferramentas com os quais possam ser trabalhados.

1.2 Definição do problema

O problema a definir neste trabalho é encontrar um modelo de previsão de demanda por meio de algoritmos de regressão via aprendizado de máquina.

1.3 Justificativas

As atuais abordagens de previsão do restaurante universitário, que envolvem análise exploratória semanal e dedução subjetiva pelos responsáveis pelo restaurante são falhas por não serem calculadas considerando-se uma visão ampla de todo um histórico de dados de grande amostragem. Além disso, não cruzam outras informações como dados climáticos, dados de calendário anual, feriados próximos, dentre outros. Essas falhas causam desperdício de alimentos por super-estimação de demanda e prejuízo de refeições descartadas, e falta de atendimento por sub-estimação de demanda de consumo.

1.4 Objetivos

1.4.1 Objetivo geral

Construir um modelo utilizando uma Rede Neural Artificial para a previsão da demanda de refeições do restaurante universitário do ICT-UNIFESP com menos de 10% de erro.

1.4.2 Objetivos específicos

- Construir dois modelos utilizando algoritmos de regressão, por métodos estatísticos e de aprendizado de máquina;
- Realizar análise e comparação da qualidade dos modelos;

1.5 Metodologia

Análise de regressão Por se tratar de um trabalho de previsão de demanda, este trabalho irá realizar a coleta e tratamento dos dados de consumo do restaurante universitário da Unifesp, a coleta de dados climáticos que influenciam em tal demanda, calcular os dados de calendário derivados das datas de coleta, por exemplo, o dia da semana, a estação do ano, o número do semestre (se par ou ímpar), horário, etc.

Após a estruturação dos dados, será realizada uma análise exploratória fundamentada por (JUNIOR, 2007), Capítulo 2.4, para previsões de demanda em geral. O modelo de dados de venda estruturados receberá o acréscimo de dados de variáveis climáticas, como possíveis fatores de influencia no consumo, conforme ocorre no trabalho de previsão de demanda de energia elétrica fundamentado por (ALMEIDA, 2013; RUAS, 2012; SILVA, 2010).

Os trabalhos de (JUNIOR, 2007) e (SILVA, 2010) fundamentam também uma classificação de análises dos dados estruturados indicando os métodos de previsão de demanda para este modelo quantitativo que são a Regressão Linear Múltipla e RNAs.

Assim, a primeira investigação usará a análise estatística de regressão linear múltipla fundamentada por (CLARICE G.B. DEMÉTRIO; ZOCCHI, 2011) para se obter um modelo de previsão de demanda do R.U. do ICT-UNIFESP.

Estudo de Aprendizado de máquina As análises de inteligência artificial, fundamentadas por trabalhos de previsão de demanda em R.U. da Universidade Federal de Viçosa, de acordo (LOPES, 2008) e na Universidade Estadual Paulista Júlio de Mesquita por (ROCHA JOSÉ CELSO; MATOS, 2011), concluem a segunda aplicação da técnica de redes neurais artificiais com o método de Perceptron de Múltiplas Camadas (PMC) fundamentado em (HAYKIN, 1994), incrementando-se na técnica a inclusão de variáveis climáticas utilizadas nos trabalhos de previsão de (ALMEIDA, 2013; RUAS, 2012; SILVA, 2010), modificando a topologia como em (LOPES, 2008) para receber tais dados.

Após a modificação da topologia, os dados serão estruturados com a retirada de amostras aleatórias de validação e será construído um comitê de PMCs usando a média aritmética como agregador para a estimação.

Análise dos resultados De acordo com (MARIANO, 2014), a comparação e a discussão dos resultados deve ser feita com base em medidas de avaliação de erro absoluto médio (EAM), erro quadrado médio (EQM) e raiz do erro quadrado médio (REQM). A verificação estatística das diferenças nas medidas obtidas com as técnicas por meio do teste de hipótese de Wilcoxon Rank Sum.

1.6 Organização do documento

Este trabalho está organizado da seguinte maneira: o Capítulo 2 possui os conceitos básicos para o entendimento do trabalho. No Capítulo 3, apresenta-se alguns trabalhos relacionados. O Capítulo 4 contém o plano de atividades para o TCC II. A conclusão é apresentada no Capítulo 5.

2 Fundamentação Teórica

2.1 Análises Estatísticas

2.1.1 Análise Exploratória dos Dados

(JUNIOR, 2007) Cita que dados coletados em modelos de previsão possuem informações que quando são projetadas graficamente evidenciam comportamentos que em alguns casos podem ser visualizados e generalizados de forma subjetiva pelos gestores dos dados. Em todos os casos, a análise exploratória é necessária para selecionar o melhor método de análise que se enquadra neste comportamento.

Somente a análise exploratória não é o suficiente e realizada nos intervalos ou critérios incorretos pode comprometer seriamente as conclusões do comportamento dos dados, e que por sua vez pode comprometer seriamente a decisão dos gestores responsáveis por estes dados, no cenário de uma previsão de demanda. Isto ocorre atualmente no cenário de previsão da demanda de refeições do ICT-UNIFESP, onde a universidade e o estabelecimento que fornece as refeições não tem nenhum modelo de previsão de demanda.

De acordo com o gestor da atual empresa que fornece refeições no ICT, a análise utilizada para se prever as refeições é observar dentro da semana o dia anterior de consumo. Em variações de 300 para 450 refeições aproximadamente, isso tem provocado um desperdício médio de 150 refeições diárias. Em geral, de acordo com o restaurante, todos os dias o mesmo trabalha com um erro e um descarte de 30% das refeições que são trazidas e consumidas ao campus. Estima-se então que no período de 2011 - 01/08/2018 os estabelecimentos tenham tido um prejuízo de R\$1.885.938,40, e de 30% de R\$78.386,85 no atual período de 01/08/2018 - 31/10/2018 totalizando o montante R\$1.964.325,25. Aproximadamente 2 milhões de reais em prejuízo acumulado desde 2011.

2.1.2 Métodos de Previsão

(JUNIOR, 2007) Realiza uma revisão bibliográfica extensa abordando principais métodos de previsão de consumo sazonais, no cenário de uma indústria cosmética. Tais métodos estatísticos de previsão se dividem em 2 ramificações, sendo quantitativos ou qualitativos. Métodos qualitativos fazem um julgamento dos dados expostos sem um sistema de processamento analítico para se produzir novos modelos ou dados, eles são úteis para sistemas de agrupamento, clusterização ou classificação de dados, sem fornecer novas informações numéricas ou modelos preditivos.

Métodos quantitativos que é o foco deste trabalho, são analíticos e se baseiam em um

modelo matemático para realizar previsões.

Para realizar tais previsões os métodos quantitativos necessitam de um histórico de dados, para analisar padrões em seu comportamento e predizer o futuro que irá agir dentro deste padrão.

Estes métodos se ramificam em 2 tipos, as séries temporais e os modelos causais.

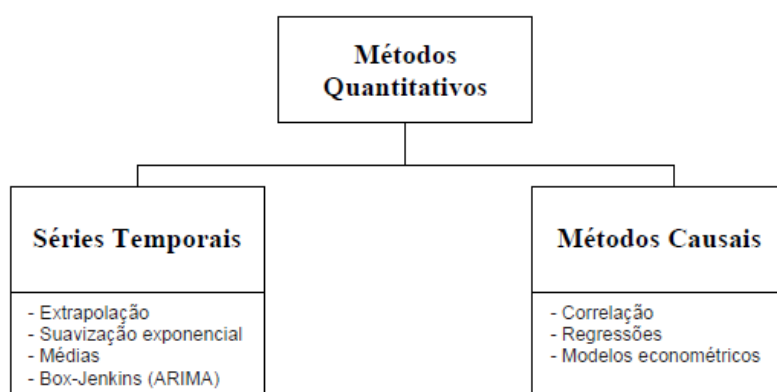


Figura 1 – Tipos de métodos quantitativos Retirado de (JUNIOR, 2007).

2.1.3 Métodos de previsão de Demanda

O autor supracitado também referencia métodos estatísticos especialmente selecionados para uma previsão de demanda, com a atenção de que alguns métodos qualitativos foram criteriosamente selecionados para prever uma demanda industrial, onde geralmente são previstas pelos métodos quantitativos.

O comportamento dos dados deste trabalho, apesar de ter uma distribuição de datas em função do tempo se classificando em um modelo de série temporal, assume-se a hipótese que tem tal comportamento impactado por relações causais com outras variáveis como recesso acadêmico, feriados, eventos, precipitações intensas que causam trânsito local e impactam na logística e frequência do público, entre outras variáveis de causas menos aparentes.

2.1.4 Séries Temporais

De acordo com (MORETTIN, 1987) uma série temporal é um conjunto de observações ordenadas em função do tempo, comumente iguais, apresentando uma dependência serial entre elas, sendo um dos objetivos do estudo de séries temporais, analisar e modelar essa dependência. Além disso, séries temporais são analisadas pelos seus principais movimentos, como tendência, sazonalidade e a componente aleatória, sendo a tendência e sazonalidade as propriedades que mais ganham destaque em pesquisas de previsão de demanda de carga elétrica. Os meios mais comuns de se analisar a componente sazonal são o Método de Regressão (paramétrico), Método de Médias Móveis (não paramétrico), e Método de Diferença Sazonal (sazonalidade

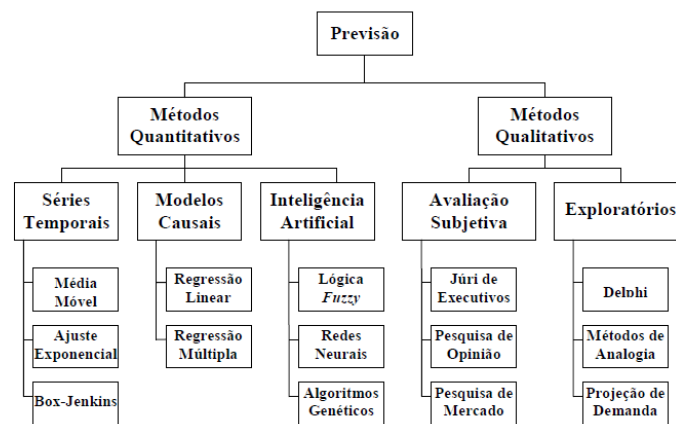


Figura 2 – Métodos de previsão de demanda (JUNIOR, 2007).

determinística). Neste trabalho, a distribuição dos dados do Restaurante se dá de forma paramétrica, ou seja, as vendas do restaurante se distribuem em função do tempo e com influências de parâmetros como o dia da semana por exemplo, sendo ideal as análises de regressão.

(ALMEIDA, 2013) também cita que para realizar a previsão de uma determinada série temporal é possível utilizar diferentes métodos. Pode-se classificá-los basicamente entre métodos estatísticos e baseados em inteligência artificial. Dentre os métodos estatísticos destacam-se:

- Regressão Linear Múltipla;
- Alisamento Exponencial;
- Média Móvel Integrada Auto-Regressiva (ARIMA)

Entre os métodos baseados em inteligência artificial tem-se:

- Redes Neurais Artificiais;
- Lógica Fuzzy.

2.1.5 Componentes Temporais

É importante observar que os métodos de previsão em séries temporais buscam uma redução da série temporal à um modelo estacionário e a decomposição da série em componentes de tendência, ciclo, sazonalidade (próxima figura) e termo aleatório. A tendência se entende pelo movimento persistente dos dados em uma direção, o ciclo indica o movimento oscilatório desta tendência, sazonalidade indica comportamento regular assumido de forma repetitiva e o termo aleatório se dá por movimentos irregulares presentes na série.

As técnicas de previsão para evidências sazonais, conforme usam métodos de regressão, que pode ser observados nos trabalhos de previsão de consumo de energia elétrica realizados

por (ALMEIDA, 2013), (RUAS, 2012), (SILVA, 2010) e de previsão de demanda de produtos cosméticos em (JUNIOR, 2007)

O cenário deste estudo analisa a frequência de alunos dentro do ICT - UNIFESP e consequentemente dentro do R.U tem sazonalidade anual a partir do momento em que a grade de disciplinas tender a ter sazonalidade anual.

Através da Figura percebe-se a sazonalidade diária, com picos de consumo próximo às 09:00, 14:00 e 20:00 horas, e baixo consumo durante a madrugada, entre 00:00 e 06:00 horas.

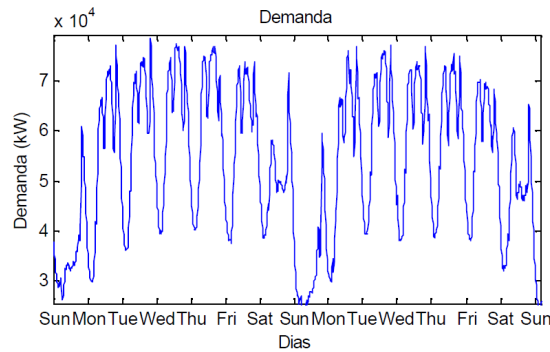


Figura 3 – DADOS DE DEMANDA DE SAZONALIDADE DIÁRIA, EM FUNÇÃO DO HORÁRIO. Retirado de (RUAS, 2012).

2.1.6 Previsão com Regressão Linear Múltipla

A teoria da regressão se iniciou no século XIX com Galton, que analisou a altura dos pais e dos filhos (X_i e Y_i) e buscou a influencia da altura do pai no filho, notando que a altura do filho tendia à media da altura do pai, e chamou esta técnica de regressão pois existe uma tendência dos dados regredirem à média.

(CLARICE G.B. DEMÉTRIO; ZOCCHI, 2011) Informa que em estudos de regressão aplica-se o método relacionar uma variável aleatória Y com uma variável X , e no caso da regressão múltipla com múltiplas variáveis X , representando causas diferentes que se combinam para a ocorrência de um valor Y .

No exemplo abaixo de regressão linear simples, podemos analisar o total de venda Y de um dia X , sendo relacionado nas variações $(\beta_0, \beta_1, \dots, \beta_k)$ de datas de X .

$$\text{Ou seja, } X, Y \rightarrow Y \simeq f(X)$$

Na regressão simples Y depende apenas de uma variável X , isto é, $Y \simeq f(\beta_0, \beta_1, \dots, \beta_k) + E_y$, Sendo E_y um erro aplicado à esta função para se chegar no valor Y . Logo a função $f(\cdot)$ pode ser linear nos parâmetros $(\beta_0, \beta_1, \dots, \beta_k)$, se a derivada da função em relação às derivadas dos parâmetros, corresponderem à $h(X)$, i , para i variando de 0 à k . Sendo $h(X)$, dependente apenas de X . $\frac{\partial f}{\partial \beta_i} = h(X)$, $i = 0, 1, \dots, k$

Na regressão múltipla, onde temos múltiplas relações causais X_1 à X_k , em ocorrências β_0 à β_k , temos a função: $Y \simeq f(X_1, X_2, \dots, X_k, \beta_0, \beta_1, \dots, \beta_k) + E_y$. Também sendo linear nos parâmetros se a derivada da função em relação às múltiplas derivadas dos parâmetros, corresponderem à $h(X_1, X_2, \dots, X_k)$, tendo $h(\cdot)$, dependente apenas de X_1, X_2, \dots, X_k . Em caso contrário, $f(\cdot)$ é uma função não linear dos parâmetros.

Logo, $Y = f(X_1, X_2, \dots, X_k, \beta_0, \beta_1, \dots, \beta_k) + E_y$ é linear nos parâmetros se: $\frac{\partial f}{\partial \beta_i} = h(X_1, X_2, \dots, X_k)$

Representação dos dados Na tabela abaixo representa-se a estrutura de dados para regressão linear múltipla, com n observações da variável resposta Y e p variáveis explicativas. Assim, y_i é o valor da i -ésima resposta e x_{ij} na é o valor da variável x_j do conjunto de j variáveis correspondente à resposta y_i . Para que seja possível aplicar o método dos quadrados mínimos (explicado adiante) de resolução do sistema de variáveis X em relação à Y , é importante que se obtenha p observações Y , onde $n \geq (p + 1)$.

Tabela 1 – Representação de dados para regressão múltipla.

y	x_1	x_2	...	x_p
y_1	x_{11}	x_{12}	...	x_{1p}
y_2	x_{21}	x_{22}	...	x_{2p}
\vdots	\vdots	\vdots	...	\vdots
y_n	x_{n1}	x_{n2}	...	x_{np}

Satisfazendo o modelo final $Y_i = \beta_0 + \beta_1 * x_{i1} + \beta_2 * x_{i2} + \beta_p * x_{ip} + \epsilon_i$
 $i = 1, \dots, n$

Método dos Mínimos Quadrados O método dos mínimos quadrados, é uma técnica utilizada popularmente em estimações econômicas, que busca encontrar o melhor ajuste de um conjunto de dados conforme tabela e modelo acima, para minimizar a soma dos erros quadráticos ϵ . A técnica foi publicada por Adrien-Marie Legendre em 1809, em *Nouvelles méthodes pour la détermination des orbites des comètes*, realizando a hipótese de que a distribuição do ϵ seja aleatória e normal, embora tal dedução já existisse indiretamente nos teoremas de Gauss-Markov.

A minimização da soma dos erros quadráticos busca então minimizar a seguinte equação L :

$$L = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 * x_{i1} - \dots - \beta_p * x_{ip})^2$$

A técnica segue o princípio de ser linear nos parâmetros, já demonstrado acima conforme (CLARICE G.B. DEMÉTRIO; ZOCCHI, 2011). Logo deriva-se L em função dos parâmetros β .

$$\frac{\partial L}{\partial \beta_0} = -2 \sum_{i=1}^n [Y_i - \beta_0 - \beta_1 * x_{i1} - \dots - \beta_p * x_{ip}]$$

$$\frac{\partial L}{\partial \beta_j} = -2 \sum_{i=1}^n [Y_i - \beta_0 - \beta_1 * x_{i1} - \dots - \beta_p * x_{ip}] x_{ji}$$

Aplicando-se álgebra, igualando as derivadas parciais a zero e organizando o sistema em forma matricial, o modelo final pode ser dado por: $Y = X\beta + \varepsilon$.

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \quad \text{e} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix},$$

Logo,

$$\frac{\partial L}{\partial \beta} = -2X'Y + 2X'X\beta.$$

Aplicando-se álgebra matricial, substituindo o vetor de coeficientes β por $\hat{\beta}$ que é o objetivo da relação busca entre as variáveis X e Y : $\hat{\beta} = (X'X)^{-1}X'Y$

Portanto o modelo final é dado por $Y_{estimado} = X\hat{\beta}$.

2.1.7 Algoritmo de Regressão Linear Multipla

1. Estruturação dos dados em forma matricial, $Y_{1,1...n}$ e $X_{1...n,1...j}$, sendo $n \geq (j + 1)$
2. Obtenção do vetor de coeficientes $\hat{\beta}$ das variáveis X , através da operação $\hat{\beta} = (X'X)^{-1}X'Y$
3. Construção do modelo final por $Y_{estimado} = X\hat{\beta}$.

2.2 Inteligencia Artificial

A Inteligência artificial "A inteligência artificial é o ramo da ciência da computação que se ocupa do comportamento inteligente."(LUGER, 2004). O termo "Inteligencia Artificial" surgiu em meados de 1956 em uma conferência, chamada Dartmouth Summer Research Project on Artificial Intelligence, sediada nos Estados Unidos em Dartmouth College, Hanover, New Hampshire, a conferência foi organizada por John McCarthy e teve como proposta reunir matemáticos, cientistas da computação e pesquisadores buscando meios de como fazer com que as máquinas usem linguagem, abstrações de formulários e conceitos, para resolver problemas que eram reservados aos humanos, e que consigam melhorar o seu próprio desempenho.

Sistemas de inteligência artificial buscam então resolver funções e problemas que seres humanos conseguem resolver melhor do que máquinas convencionais, usando sua capacidade de abstração e aprendizagem com o erro.

2.2.1 Neurônio Artificial

História e inspiração da Inteligência Artificial Redes Neurais Artificiais são elementos de inteligência artificial da ciência da computação, inspirados no funcionamento do cérebro humano. São formadas por neurônios artificiais interconectados que são capazes de processar múltiplos valores de entradas, e reagir produzindo uma resposta relacionada à essas entradas. Como qualquer outro método de aprendizado de máquina, este modelo busca obter um aprendizado a partir dos dados de entrada recebidos, criando uma capacidade de generalização de problemas e assim buscam o objetivo principal de resolver novos problemas com este aprendizado.

Uma rede neural pode não produzir uma resposta esperada, resolvendo erroneamente um problema, assim como o cérebro humano tem limitações de aprendizado, que levam o homem a cometer falhas de decisões ou ações por um aprendizado mal treinado. Por isso as características fundamentais que tornam uma rede neural artificial em uma boa solução, é um bom planejamento de sua topologia, e método de treinamento, que serão explicados a seguir.

([BRYAN KOLB; ARIF, 2014](#)) Demonstra em sua pesquisa que a propriedade fundamental do cérebro é a sua capacidade de mudar com uma grande variedade de experiências, inclusive lesões que provocam perdas de neurônios, abordando princípios de plasticidade cerebral, mostra também que a capacidade de aprendizado de mamíferos que sofreram uma redução de neurônios causada por lesões, pode ser restaurada não somente pela recuperação desses neurônios perdidos, mas sim por novas conexões e sinapses entre outros neurônios. Ou seja, a rede sofre uma mudança de topologia.

([ZORZETTO, 2012](#)) Mostra em um experimento recente que o cérebro humano possui atualmente 86 bilhões de neurônios interconectados, e que sua capacidade de aprendizado e habilidades evolutivas também vêm aumentando em conjunto com seu número de neurônios e topologia de rede neural, que há 30 milhões de anos atrás tinha apenas 2,5 bilhões de neurônios e era apenas um animal arborícola quadrúpede.

Neurônio Biológico O modelo de neurônio artificial surge então, com a busca da inteligência artificial de reproduzir o comportamento de aprendizado humano, reproduzindo computacionalmente seu elemento biológico principal de aprendizagem: O neurônio biológico. Um neurônio biológico (Figura 05) processa informações recebidas por meio de seus dendritos e processa-as em seu corpo celular, tal reação à esses estímulos recebidos gera um sinal de saída como resposta aos estímulos, enviado através do axônio. Esse estímulo é repassado como sinal de entrada, através de outros neurônios por meio de seus dendritos, e o ciclo se repete em uma vasta rede. Controlando essas conexões, pontos de contato entre a resposta de um neurônio e a entrada de outro, existem as sinapses. Elas funcionam como agentes que permitem a interação acontecer ou inibem, e são acionadas por um conjunto somatório de estímulos. Se tal somatório de estímulos for satisfatório, elas permite a transmissão de sinal elétrico pelo axônio até

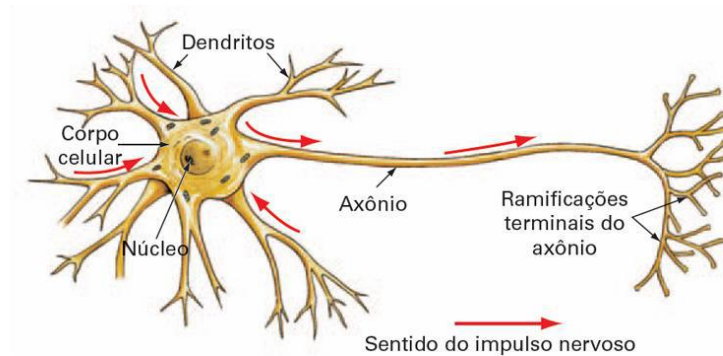


Figura 4 – Neurônio Biológico, retirado de <https://pt.khanacademy.org/science/biology/human-biology/neuron-nervous-system/v/anatomy-of-a-neuron>.

o dendrito de um neurônio vizinho, formando um ciclo de aprendizado em uma rede neural biológica.

Neurônio Artificial Warren McCulloch e Walter Pitts em 1942 observando o neurônio biológico iniciam a busca de um modelo computacional do mesmo. McCulloch era psiquiatra e neuroanatomista e passou cerca de 20 anos refletindo e estudando sobre a representação do sistema nervoso, em 1942 ele convidou Pitts, que era matemático, para fazer parte das suas pesquisas. Em 1943 lançaram o artigo "A Logical Calculus of the Ideas Immanent in Nervous Activity." chegando em um modelo matemático do neurônio artificial:

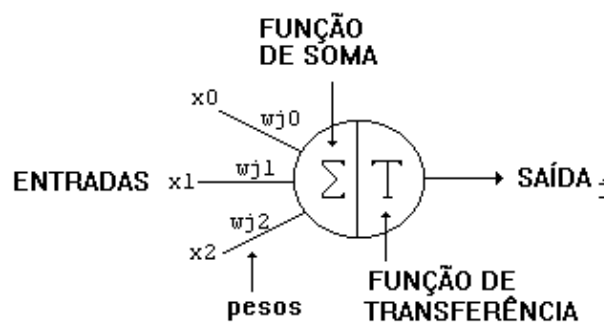


Figura 5 – Neurônio Artificial, Retirado de <http://redesneuraisartificiais.blogspot.com/2010/10/o-primeiro-modelo-de-um-neuronio-criado.html>

Assim como no neurônio biológico, este modelo reage à um vetor de entradas X_0, X_1, \dots, X_n onde tem suas sinapses representadas por pesos numéricos, existe uma soma ponderada dessas entradas que é controlada por uma função de transferência ou função de ativação, determinando se essa soma é maior que um valor numérico. Se essa soma for satisfatória o neurônio é ativado emitindo um valor de saída 1, caso contrário se emite um valor de saída 0. Todo o funcionamento deste modelo então é reduzido a responder se a soma recebida é maior que um valor numérico esperado.

Neste modelo simplista o neurônio consegue realizar operações lógicas. Na figura 07 os

pesos w_1 e w_2 e o limiar T estão ajustados para responder à operação AND que resumidamente à uma operação lógica que somente produz um valor positivo 1 na resposta de saída se ambos os valores de entrada forem 1.

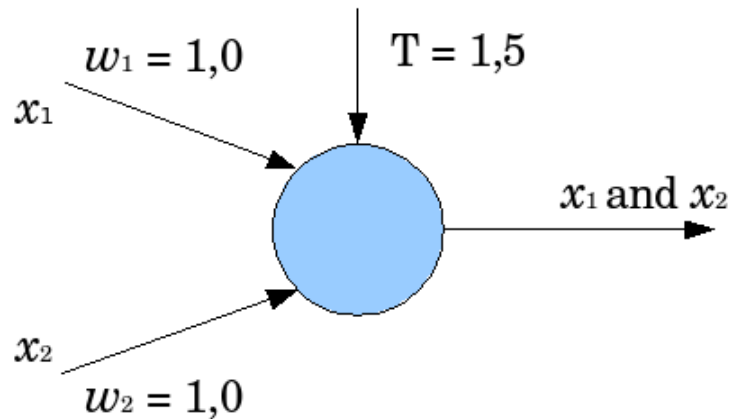


Figura 6 – Neurônio Artificial - Operação AND, Retirado de <http://redesneuraisartificiais.blogspot.com/2010/10/o-primeiro-modelo-de-um-neuronio-criado.html>

O neurônio irá comparar se a soma das entradas X multiplicadas pelos seus devidos pesos W , será menor que o limiar T . Ele somente irá produzir uma resposta positiva (1) se essa soma for maior. É fácil verificar então na imagem, que o resultado dessa soma só será positivo, se ambas as entradas forem 1, onde a soma será $2 > 1,5$.

Tabela 2 – Operação Lógica AND

X_1	X_2	Saída
0	0	0
0	1	0
1	0	0
1	1	1

2.2.2 Redes Neurais Artificiais

Perceptron O modelo de Warren McCulloch e Walter Pitts, apesar de conseguir simular o modelo de neurônio biológico e resolver algumas tarefas lógicas e matemáticas não atendia o objetivo principal da Inteligência Artificial: A capacidade de aprendizado. Para utilizar um modelo de neurônio artificial a fim da solução de um problema era necessário conhecer o ajuste dos pesos das entradas, e em uma tarefa de um cenário complexo e com muitas variáveis, ou pesos não perceptíveis ao valor esperado de saída, este ajuste não se torna trivial.

Em 1958, o cientista da computação Frank Rosenblatt desenvolveu o modelo de neurônio artificial Perceptron para solucionar este problema de ajuste de pesos. Basicamente neste novo modelo, os pesos das conexões são ajustados de forma autônoma com a introdução de pessoas associados e um valor bias, a fim de buscar um reconhecimento autônomo de padrões.

Tarefas de reconhecimento de padrões simples com separações lineares os seres humanos conseguem realizar de forma trivial mas ainda era um desafio para tal problema se resolvido por uma máquina.

O perceptron (MARIANO, 2014) possui apenas uma camada de entrada e saída, a saída utiliza como função de ativação a função degrau u que define quando o neurônio emitirá o sinal lógico 1 ou quando emitirá o sinal lógico 0.

O sinal de saída do perceptron entende-se então por: $y = \delta(\sum_{i=1}^n X_i W_i + b)$

- X_i - sinais de entrada do neurônio;
- W_i - pesos sinápticos do neurônio;
- b - bias ou limiar de ativação;
- $\delta(\cdot)$ - função de ativação;
- y - sinal de saída do neurônio.

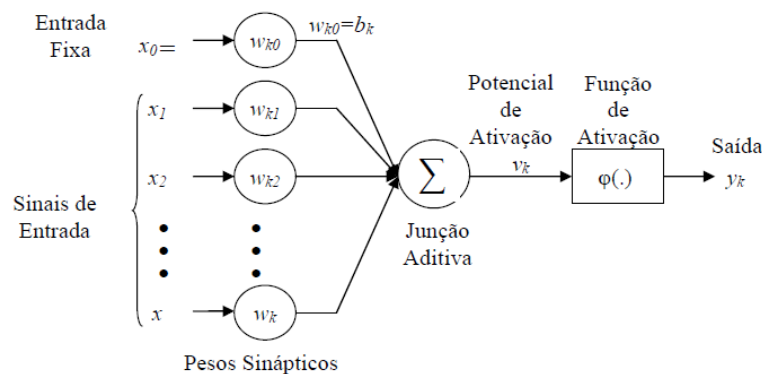
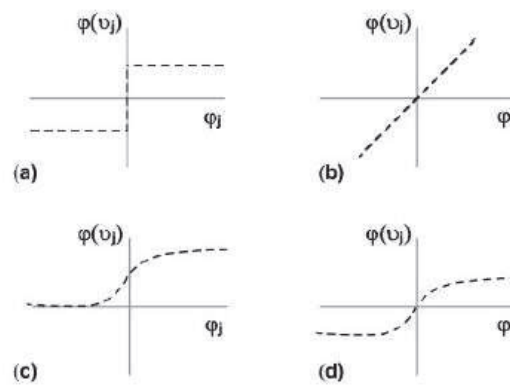


Figura 7 – Neurônio Artificial Perceptron, retirado de (JUNIOR, 2007) ()

Em algumas literaturas o bias pode ser reduzido a um peso W_0 com entrada X_0 fixo em 1 no neurônio, a representação gráfica da topologia pode mudar, mas no somatório de saída, o cálculo continua o mesmo.

A função de ativação δ se apresenta de forma linear ou não linear, determinando a saída de um neurônio a partir do seu potencial de ativação. As funções de ativação podem ser observadas na figura 09, sendo a de maior popularidade de utilização, a sigmoide.

As funções de ativação dá a capacidade do perceptron quando conectado em rede (explicado no próximo capítulo) de resolver problemas lineares e não lineares, agregando adaptação e improviso ao resolver programas que não estão contidos em seus dados de alimentação.



Tipos de função de ativação: (a) função degrau, (b) função linear, (c) função sigmoide, (d) função tangente hiperbólica

Figura 8 – Funções de Ativação retirado de (MARIANO, 2014)

Perceptron - Treino Supervisionado (ALMEIDA, 2013) cita que o processo de aprendizado do perceptron pode ocorrer de forma supervisionada quando o neurônio deve aprender a relacionar um conjunto observado de variáveis à um valor observado de saída deste mesmo conjunto. O neurônio recebe os sinais de entrada X_i e produz uma saída Y_i através do combinador linear e função δ de ativação; compara essa saída Y_i com a observação i obtida do conjunto de dados (ponto chave do processo de treino supervisionado) e por fim essa comparação irá gerar um erro e .

Critério de parada De acordo com algum critério adotado em cada contexto de aplicação do perceptron, este erro pode ser aceitado e o neurônio mantém o valor de seus pesos W_i que impactam na saída Y_i desejada. A aprendizagem do neurônio também pode atingir um critério de parada após N épocas de treinamento, diversos critérios de paradas podem ser combinados.

Época de treinamento Se o critério de parada não for aprovado, uma nova época de treinamento (ou repetição de interação) é iniciada mesmos valores X_i e Y_i passados, os valores de pesos W_i são reajustados com a taxa de aprendizagem, buscando o objetivo de se obter um erro menor.

Taxa de aprendizagem Esse reajuste de pesos W_i denomina-se taxa de aprendizagem, α que pode ter valores de escolha livre ao contexto de aplicação do perceptron, para reajustar estes pesos W_i . Assim que a taxa de aprendizagem ajusta os pesos, uma nova época N+1 de treinamento está se iniciando buscando novamente um erro menor. Ressaltando então que o critério de parada pode ser acionado e interromper o reinício do processo, se for estipulado um limite para o valor de N combinado ou não com um limite para o erro.

2.2.3 Funcionamento e treino do neurônio perceptron

Em (ALMEIDA, 2013) verifica-se de forma procedural, o processo do treino supervisionado do perceptron, com m observações $Y = (Y_i, Y_{i+1}, \dots, Y_m)$ com n entradas X relacionadas à cada Y_m , ocorre da seguinte maneira: Matematicamente o neurônio é representado por um vetor global de n pesos sinápticos (W_0, W_1, \dots, W_n) , e um valor de bias b . O processo do algoritmo abaixo ajustará esses pesos e bias.

1. Coleta do valor supervisionado inicial Y_0 , que é a primeira observação histórica do conjunto de dados.
2. Coleta do vetor de variáveis $(X_{00}, X_{01}, \dots, X_{0n})$ relacionados à saída supervisionada Y_i .
3. Inicialização dos pesos sinápticos (W_0, W_1, \dots, W_n) com valores aleatoriamente pequenos.
4. Definição de um valor real entre 0 e 1 para a taxa de aprendizado α .
5. Definição do valor inicial de bias b .
6. Definição da função de ativação δ
7. Definição do limite de épocas de treino l_e , se não houver, definir l_e para -1.
8. LAÇO: Percorrer e atribuir em (Y_d) de Y_0 até Y_m , inicializando o contador de épocas $c_e = 0$, enquanto $CP(e)$ for verdadeiro E/OU (l_e) for menor que c_e :
9. Nova coleta do vetor de variáveis $(X_{d0}, X_{d1}, \dots, X_{dn})$ relacionados à saída supervisionada Y_d .
10. Incrementar 1 em c_e .
11. Cálculo do sinal $u = \sum_{i=1}^n X_{di} * W_i + b$
12. Aplicação de valor de U na função de ativação δ obtendo $Y_u = \delta(u)$
13. Obtenção do erro $e = Y_d - Y_u$
14. Aplicação do critério de parada CP em relação ao erro e ; $CP(e)$
15. SE $CP(e)$ for APROVADO, o algoritmo continua, SE REPROVADO, a aprendizagem é feita ajustando $b = b + (\alpha * (e))$, e ajustando o vetor $W_i = W_i + (\alpha * e * X_{di})$.

Funcionamento do perceptron para logica AND .

Este exemplo aborda a aplicação do problema do neurônio de Warren McCulloch e Walter Pitts, o aprendizado para a tabela AND, que busca fornecer uma resposta verdadeira apenas se a primeira E todas as demais entradas forem verdadeiras.

A utilização do perceptron com o modelo de aprendizado supervisionado, busca então conseguir de forma autônoma os valores de pesos W_i e desta forma aprender o funcionamento da logica E, sem que seja necessário a tarefa humana de ajustar o mesmo.

linha	X_0	X_1	Y
0	0	0	0
1	0	1	0
2	1	0	0
3	1	1	1

Para este aprendizado vamos definir o critério de parada de 0 para o erro. Definir o valor 0 para o *bias*. 1 para α e função $\delta(u) = \text{degrau}$ (1, se $u \geq 0$, 0 se $u < 0$). O número de épocas terá o mesmo número de valores supervisionado, então o treino executará 1 vez para cada observação.

linha $d = 0$

1. Coleta do valor supervisionado $Y_d = Y_0 = 0$.
2. Coleta do vetor de variáveis ($X_{00} = 0, X_{01} = 0$).
3. Inicialização dos todos os pesos com valores aleatoriamente pequenos: $W_0 = 3, W_1 = 3$
4. Definição da taxa de aprendizado $\alpha = 1$.
5. Definição do valor de bias $b = 3$.
6. Definição da função de ativação $\delta = \text{degrau}$
7. Cálculo do sinal $u = x_{00} * w_0 + x_{01} * w_1 + b = (0 * 3 + 0 * 3 + 0) = 0$
8. Aplicação de valor de u na função de ativação δ obtendo $Y_u, Y_u = \delta(0) = 1$
9. Obtenção do erro $e = Y_0 - Y_u = 0 - 1 = -1$
10. Aplicação do critério de parada CP em relação ao erro e REPROVADO, pois o erro esperado é 0.
11. Como CP foi reprovado, a aprendizagem é realizada $b = b + (\alpha * e)$ e $W_i = W_i + (\alpha * e) * X_{0i}$

Aprendizagem:

1. $b = b + (\alpha * e) = 0 + 1 * (-1) = -1$
2. $W_0 = W_0 + (\alpha * e) * X_{00} = 3 + 1 * (-1) * 0 = 3$
3. $W_1 = W_1 + (\alpha * e) * X_{01} = 3 + 1 * (-1) * 0 = 3$

Reinício dos pesos reajustados pela aprendizagem e apresentação da próxima linha ao neurônio: linha $d = 1$.

linha	X_0	X_1	Y
1	0	1	0

1. Coleta do valor supervisionado $Y_d = Y_1 = 0$.
2. Cálculo do sinal $u = X_{10} * w_0 + X_{10} * W_1 + b = (0 * 3 + 1 * 3 - 1) = 0$
3. Aplicação de valor de u na função de ativação δ obtendo $Y_u, Y_u = \delta(0) = 1$
4. Obtenção do erro $e = Y_d - Y_u = 0 - 1 = -1$
5. Aplicação do critério de parada CP em relação ao erro e reprovado, pois o erro esperado é 0.
6. Como CP foi reprovado, a aprendizagem é realizada $b = b + (\alpha * e)$ e $W_i = W_i + (\alpha * e) * X_{1i}$

Aprendizagem:

1. $b = b + (\alpha * e) = -1 + 1 * (-1) = -2$
2. $W_0 = W_0 + (\alpha * e) * X_{10} = 3 + 1 * (-1) * 0 = 3$
3. $W_1 = W_1 + (\alpha * e) * X_{11} = 3 + 1 * (-1) * 1 = 2$

Reinício dos pesos reajustados pela aprendizagem e apresentação da próxima linha ao neurônio: linha $d = 2$

linha	X_0	X_1	Y
2	1	0	0

1. Coleta do valor supervisionado $Y_d = Y_2 = 0$.
2. Cálculo do sinal $u = X_{20} * w_0 + X_{20} * W_1 + b = (1 * 3 + 0 * 2 - 2) = 1$
3. Aplicação de valor de u na função de ativação δ obtendo $Y_u, Y_u = \delta(1) = 1$
4. Obtenção do erro $e = Y_d - Y_u = 0 - 1 = -1$
5. Aplicação do critério de parada CP em relação ao erro e reprovado, pois o erro esperado é 0.
6. Como CP foi reprovado, a aprendizagem é realizada $b = b + (\alpha * e)$ e $W_i = W_i + (\alpha * e) * X_{2i}$

Aprendizagem:

1. $b = b + (\alpha * e) = -2 + 1 * (-1) = -3$
2. $W_0 = W_0 + (\alpha * e) * X_{20} = 3 + 1 * (-1) * 1 = 2$
3. $W_1 = W_1 + (\alpha * e) * X_{21} = 2 + 1 * (-1) * 0 = 2$

Reinício dos pesos reajustados pela aprendizagem e apresentação da próxima linha ao neurônio: linha $d = 3$

linha	X_0	X_1	Y
3	1	1	1

1. Coleta do valor supervisionado $Y_d = Y_3 = 1$.
2. Cálculo do sinal $u = X_{30} * w_0 + X_{30} * W_1 + b = (1 * 2 + 1 * 2 - 3) = 1$
3. Aplicação de valor de u na função de ativação δ obtendo $Y_d, Y_u = \delta(1) = 1$
4. Obtenção do erro $e = Y_d - Y_u = 1 - 1 = 0$
5. Aplicação do critério de parada CP em relação ao erro e_n APROVADO, então conclui-se que os pesos sinápticos estão corretos.

Pesos finais e bias aprovados pelo processo de treino:

1. $b = -3$
2. $W_1 = 2$
3. $W_2 = 2$

Fim do processo de treino.

Validação do treino Após todo o processo de treino e apresentação dos conjuntos, é necessário realizar um novo ciclo no perceptron, para ver se o neurônio responde corretamente ao modelo de dados inicial.

Pesos finais e bias aprovados pelo processo de treino:

1. $b = -3$
2. $W_1 = 2$
3. $W_2 = 2$

linha $d = 0$

linha	X_0	X_1	Y
0	0	0	0

1. Coleta do valor supervisionado $Y_d = Y_0 = 0$.
2. Cálculo do sinal $u = X_{00} * w_0 + X_{01} * W_1 + b = (0 * 2 + 0 * 2 - 3) = -3$
3. Aplicação de valor de U na função de ativação δ obtendo Y_u , $Y_u = \delta(-3) = 0$
4. Obtenção do erro $e = Y_i - Y_u = 0 - 0 = 0$
5. Aplicação do critério de parada CP em relação ao erro e APROVADO, então o neurônio respondeu corretamente às entradas.

linha $d = 1$

linha d	X_0	X_1	Y
1	0	1	0

1. Coleta do valor supervisionado $Y_d = Y_1 = 0$.
2. Cálculo do sinal $u = X_{10} * w_0 + X_{11} * W_1 + b = (0 * 2 + 1 * 2 - 3) = -1$
3. Aplicação de valor de u na função de ativação δ obtendo Y_u , $Y_u = \delta(-1) = 0$
4. Obtenção do erro $e = Y_d - Y_u = 0 - 0 = 0$
5. Aplicação do critério de parada CP em relação ao erro e APROVADO, então o neurônio respondeu corretamente às entradas.

linha d	X_0	X_1	Y
2	1	0	0

linha $d = 2$

1. Coleta do valor supervisionado $Y_d = Y_2 = 0$.
2. Cálculo do sinal $u = X_{20} * w_0 + X_{21} * W_1 + b = (1 * 2 + 0 * 2 - 3) = -1$
3. Aplicação de valor de u na função de ativação δ obtendo Y_u , $Y_u = \delta(-1) = 0$
4. Obtenção do erro $e = Y_d - Y_u = 0 - 0 = 0$
5. Aplicação do critério de parada CP em relação ao erro e APROVADO, então o neurônio respondeu corretamente às entradas.

linha $d = 3$

linha d	X_0	X_1	Y
3	1	1	1

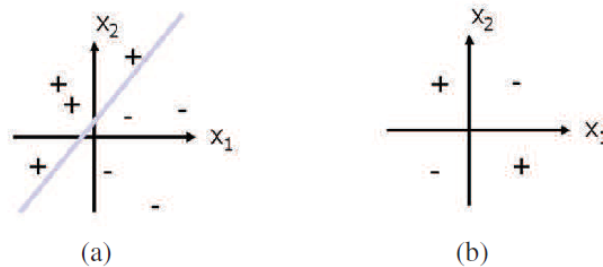
1. Coleta do valor supervisionado $Y_d = Y_3 = 1$.
2. Cálculo do sinal $u = X_{30} * w_0 + X_{31} * W_1 + b = (1 * 2 + 1 * 2 - 3) = 1$
3. Aplicação de valor de u na função de ativação δ obtendo Y_u , $Y_u = \delta(1) = 1$
4. Obtenção do erro $e = Y_d - Y_u = 1 - 1 = 0$
5. Aplicação do critério de parada CP em relação ao erro e APROVADO, então o neurônio respondeu corretamente às entradas.

Como o neurônio não teve nenhum critério de parada REPROVADO em todas as interações de todas as linhas, podemos dizer que o mesmo está devidamente treinado e validado para a solução do problema.

2.2.4 Limite de um Perceptron

É possível reparar que o processo de aprendizagem de um perceptron, se forma por uma combinação linear que é o valor u dos pesos e entradas x_i . Essa combinação pode ser aprovada ou reprovadas no processo repetitivo de treino controlado pela função δ , e pela aprovação do erro medido entre o valor da combinação linear e o valor esperado.

Então em 1969, Marvin Mjinsky e Seymour Papert realizaram uma publicação comprovando essa limitação de aprendizado à uma combinação linear, e provaram que o perceptron é limitado à resolução de problemas que são resolvidos ou classificados por apenas 1 linha, ou seja, problemas linearmente separáveis que consiste na mesma limitação do processo de regressão linear múltipla. Na figura 10, abaixo, o perceptron não pode solucionar o problema (b). E essa limitação desmotivou e parou os estudos com perceptrons.



Exemplos linearmente separável (a) e não separável linearmente (b)

Figura 9 – Problemas Linearmente Separáveis e Não Separáveis. Retirado de (MARIANO, 2014).

Em 1986 James McClelland e David Rumelhart proporam o método de desenvolvimento de uma rede neural de perceptrons, tangindo mais ainda a inspiração da solução no cérebro humano. O processo resume o treinamento do perceptron simples aplicado à um conjunto de perceptrons interligados, e assim solucionando problemas complexos que podem ser resolvidos com uma combinação de soluções.

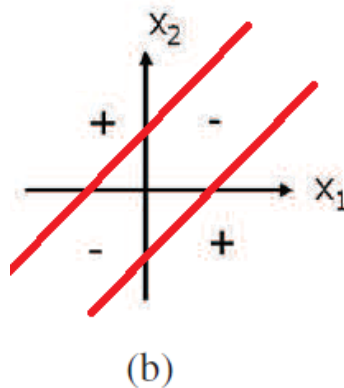


Figura 10 – Problemas Linearmente Não Separáveis Múltiplas Soluções.

2.2.5 Rede Perceptrons Múltiplas Camadas - MLP

A solução base para se combinar 2 ou mais perceptrons a fim de se resolver um problema com a combinação de 2 ou mais soluções lineares, é a utilização de um perceptron combinador de sinal de saída, já que cada perceptron pode ter múltiplas entradas e somente uma saída. Dessa forma as redes neurais vão formando colunas de perceptrons interconectados. Cada coluna é

denominada uma camada da rede neural. A última camada deve ter o número de perceptrons correspondente ao número de saídas desejadas. Na figura abaixo, encontra-se uma rede neural com duas camadas intermediárias e 3 saídas.

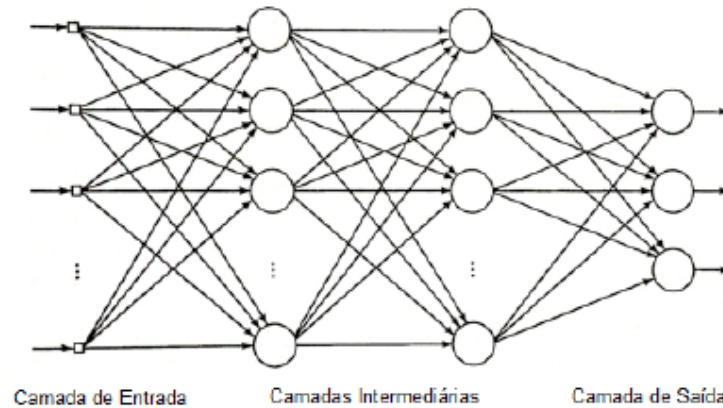


Figura 11 – Rede de perceptrons com múltiplas camadas. Retirado de (ALMEIDA, 2013)

Esta rede denominada MLP (Multilayer Perceptron) possui uma camada de entrada onde cada nó representa uma variável a ser considerada ao problema a ser analisado, e pelo menos uma camada intermediária.

Nesta camada intermediária os neurônios possuem geralmente uma função de ativação sigmoideal logística ou tangente hiperbólica, e conceitualmente no mínimo 1 neurônio desta primeira camada oculta deve receber no mínimo 2 conexões de entrada. E uma camada de saída, na borda à direita, com o número de neurônios correspondente ao número de soluções buscadas.

Treino e Validação da MLP O conjunto de dados de entrada na rede MLP supervisionada, deve ser dividido em 2 partes principais, Treinamento e Validação. É importante salientar que as observações de ambos os conjuntos devem originar do mesmo conjunto de dados para representar o mesmo problema, as observações Y_i relacionadas aos vetores i de variáveis $X_{i1}, X_{i2}, \dots, X_{ij}$ do conjunto de treino, devem ter a mesma estrutura, mesmo número j de variáveis X por vetor de entrada e devem representar o mesmo problema que as observações Y_i do conjunto de validação.

O conjunto de dados da validação e treino somados devem formar exatamente o conjunto original, sem informações excedentes ou em falta.

Os dados de treino e dados de validação, podem ser separados de forma aleatória, sendo os dados de treino os responsáveis pelos reajustes de pesos e capacidade de generalização da rede, e os dados de validação responsáveis pelo processo de validação pós-treino.

É importante adotar um bom critério de parada de treino, pois um treino prolongado tende a convergir em ajustes de pesos memorizados dos valores observados nos dados de treino

e isso causa o fenômeno de overfitting na rede neural, que é a perda da capacidade de generalização. Pesos sinápticos que sofrem processo prolongado de reajustes acaba "viciando" a rede neural para reconhecer apenas os dados de treino.

Parada do erro mínimo: O critério de parada do erro mínimo encerra o treinamento da RNA quando a mesma obtém um erro menor que o mínimo estipulado para o valor observado, este é o critério mais simples, adotado nos casos onde existe um limiar de erro já determinado pelo problema, entre o valor observado e o valor a se prever.

Parada por número de épocas: Um critério de parada pode ser limitado também ao número de épocas de treino. A determinação deste número de épocas pode ocorrer por tentativa e erro, visto que haverá a convergência de um número pequeno para uma baixa capacidade de aprendizado, e a convergência de um número grande para o processo de overfitting. Logo é necessário realizar experimentos que validem número de épocas fora desses intervalos de convergência.

Parada por validação cruzada: Por fim a validação cruzada é a técnica que se utiliza os dados dos conjuntos de validação e treino de forma cruzada, neste processo então, os dados de treino são utilizados no processo iterativo de aprendizagem, e no fim deste processo o conjunto é validado com os dados de validação, obtendo-se um novo erro de validação. A medição do erro de validação passa por um processo de avaliação em função do número de épocas, a fim de se detectar um ponto de número de épocas onde o erro quadrático médio da amostra de validação sofre uma curva de crescimento após um limiar de decréscimo.

É notório observar que o erro quadrático médio das amostras de treinamento sofrerá um decréscimo em função do aumento do número de épocas de treinamento, convergindo ao processo de overfitting, ou memorização da rede.

O ponto de parada no limite inferior do erro quadrático médio da amostra de validação, será o ponto ótimo de parada de treinamento.

O treinamento de uma rede MLP para previsões de demanda de restaurantes universitários que obteve sucesso em (LOPES, 2008) e (ROCHA JOSÉ CELSO; MATOS, 2011) é a retro propagação de erro.

2.2.6 Perceptrons Múltiplas Camadas com Retro Propagação de Erro

A rede Perceptron Múltiplas Camadas com Retro Propagação de Erro (do inglês, M.L.P Backpropagation), faz o reajuste dos pesos sinápticos dos neurônios através de duas fases:

Feed-forward Nesta primeira fase de treino os sinais $X_{i0}, X_{i1}, \dots, X_{in}$ com sua respectiva saída Y_i do conjunto de dados são apresentados à todos os neurônios da primeira camada. O processo de propagação do sinal de saída de cada neurônio segue o princípio do neurônio ar-

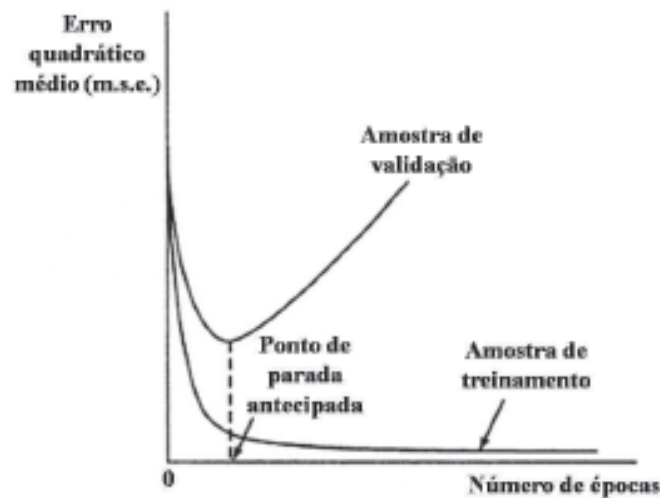


Figura 12 – Ponto ótimo de parada da validação cruzada. Retirado de (MARIANO, 2014)

tificial apresentado na seção anterior, que envia o sinal de saída como um sinal de entrada ao neurônio seguinte.

Feed-backward Nesta fase é obtido um valor de erro da camada de saída. Este erro é utilizado na equação de reajuste do peso sináptico das conexões dos neurônios da camada de saída com o sinal de saída dos neurônios da última camada oculta. E depois este erro é propagado realizando outra equação de reajuste de peso sináptico das conexões dos neurônios nas camadas anteriores, no sentido contrário em direção à camada de entrada. Isto permite que os pesos sináptico de todas as camadas de intermediárias neste processo tenham seus pesos ajustados.

MLP Backpropagation em R.U em trabalhos relacionados Em (LOPES, 2008) A rede neural perceptron de múltiplas camadas é utilizada para tratar a previsão de demanda do R.U da UFV, utilizando apenas como variáveis quantitativas as 5 ultimas observações anteriores ao dia a se analisar, e como variáveis qualitativas o dia da semana variando de segunda a sexta, em valores binários.

Em (ROCHA JOSÉ CELSO; MATOS, 2011) o trabalho realizado com neurônios artificiais para prever a demanda do R.U da UNESP, envolve apenas uma única camada de entrada e uma segunda camada para saída, porém com uma diversidade maior de variáveis de entrada correlacionadas com o consumo do restaurante.

Ambos os modelos possuem topologia de uma camada oculta para entrada dos dados, e uma camada de saída. que de acordo com (RAGA A. DE P.; CARVALHO, 2000) cita que através de uma análise de Cybenko, uma camada intermediária é o suficiente para aproximar qualquer função contínua e 2 camadas intermediárias são suficientes para aproximar qualquer

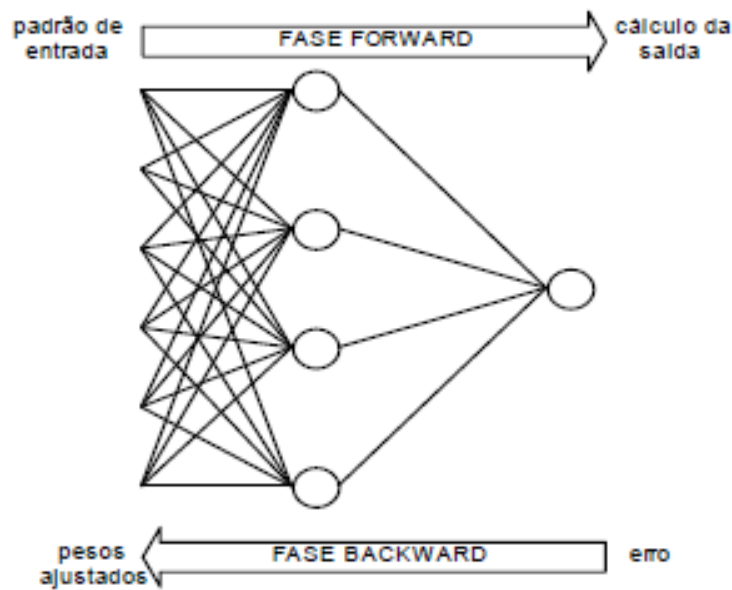


Figura 13 – Fases de treino da MLP-Back-Propagation. Retirado de (ALMEIDA, 2013)

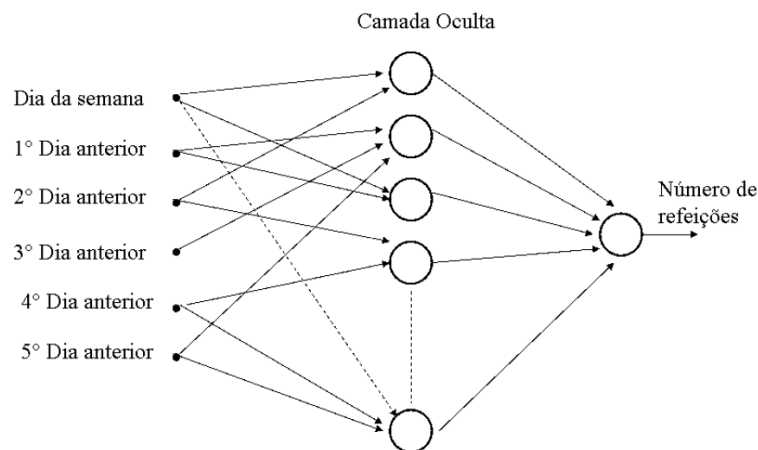


Figura 14 – Rede Neural Perceptron de Múltiplas Camadas. Retirado de (LOPES, 2008).

função matemática, e devendo ser observado o fato de que em alguns casos, a utilização de 2 ou mais camadas pode facilitar o treinamento da rede, porém a utilização de um grande número de camadas intermediárias ou ocultas, é inviável, pois em cada uma delas a estimativa do erro se trata de uma estimativa da estimativa do erro da camada anterior, e este cascadeamento de estimativas pode se tornar menos preciso à medida que cresce.

(MARIANO, 2014) cita que o número de neurônios na primeira camada oculta é proporcional à dimensão do espaço de observação. Logo em modelos preditivos de demanda, supracitados, observa-se que no mínimo, na primeira camada oculta, é utilizado um número de neurônios igual ao número de variáveis que influenciam no dado preditivo.

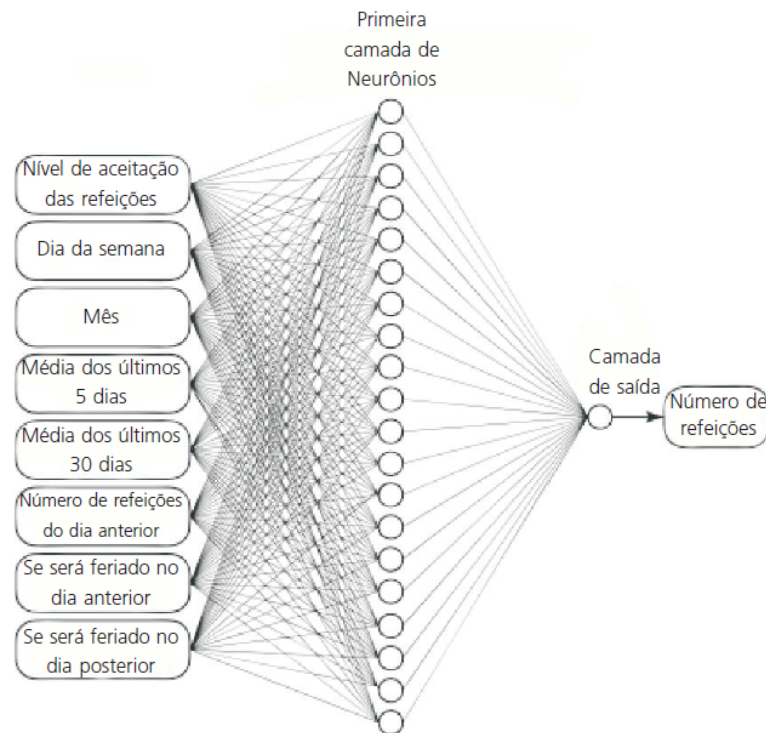


Figura 15 – Rede Neural Perceptron de Múltiplas Camadas. Retirado de (ROCHA JOSÉ CELSO; MATOS, 2011)

Parâmetros de treino da MLP Backpropagation A definição da função δ de ativação interfere na linearidade do modelo a ser analisado, sendo a função sigmoide a mais popular na literatura.

A taxa de aprendizagem, define a velocidade de reajuste dos pesos, podendo variar de 0 a 1. Ressalta-se que uma taxa próxima de 1 provoca picos oscilatórios na taxa de aprendizado, e taxa próxima de 0 provoca lentidão da convergência de aprendizagem. Valores comuns de utilização ficam entre 0,2 e 0,8.

2.2.7 Treino da MLP - BackPropagation para 1 sinal de saída

No caso do problema de predição de demanda onde se busca apenas um valor de saída, que é a previsão de vendas em relação às variáveis de entrada, o cálculo de reajuste dos pesos da camada de saída é reduzido à apenas 1 neurônio na camada de saída, porém o somatório no cálculo do erro pode envolver mais neurônios de saída em outras aplicações.

Um vetor i do conjunto de dados, de tamanho n de variáveis de entrada $X_{i_0}(j), X_{i_1}(j), \dots, X_{i_n}(j)$, que são os sinais de entrada, são apresentados à rede, relacionados à um valor supervisionado Y_i .

Em resumo o treino da rede perceptron com backpropagation obtém o sinal de saída no último neurônio s da rede, através da propagação dos sinais de saída dos neurônios anteriores

da rede, feito pela aplicação da combinação linear dos sinais de entrada com os pesos sinápticos em uma função de ativação. É calculado o erro quadrático e deste sinal de saída em s com o valor Y_i , e a regra do Gradiente Descendente com base neste erro é utilizada para o reajuste dos pesos sinápticos em todas as conexões de todos os neurônios. O processo é denominado regra delta generalizada, Δ

Neste capítulo, será denominado a função de ativação do neurônio por $F(.)$ para uma melhor coerência, devido ao símbolo δ utilizado para esta função nas seção anterior sobre o neurônio artificial ser amplamente utilizado para outros fins na regra de reajuste dos pesos sinápticos da rede MLP. Esta função na MLP - Backpropagation deve ser uma função não-linear diferenciável em todos os pontos, popularmente é usada a função sigmoide.

O erro quadrático do treino de uma rede MLP com Backpropagation, com 1 sinal de saída, no neurônio de saída s , em uma i -ésima iteração de treino em uma determinada época k será:

$$e(i) = \frac{1}{2} * ((Y_i - Y_{i\mu}(s))^2)$$

onde $Y_{i\mu}(s)$ é o sinal de saída deste neurônio de saída, e Y_i é o valor de saída da observação do conjunto de dados.

O principal objetivo de todo o método backpropagation conforme o avanço de épocas, é o treino em uma determinada época reduzir a média de erros quadráticos do conjunto de validação ($m.s.e$) em n iterações de validação desta época:

$m.s.e = \frac{1}{n} \sum_{k=1}^n e(k)$ É importante observar que 1 época consiste no par treino e validação. Logo após o treino, os erros $e(k)$ serão obtidos novamente através das entradas Y_i do conjunto de validação, e respostas propagadas $Y_{i\mu}$ na camada de saída da rede já treinada, para então o $m.s.e$ ser calculado.

O valor $m.s.e$ será o critério de parada do treinamento de toda a rede e deve ser observado a cada época de treinamento, assim que este valor atingir um ponto ótimo, conforme figura deste ponto ótimo exibida anteriormente, o critério de parada será verdadeiro e o treinamento deve parar, pois a partir desse ponto a rede converge à um overfitting (divergindo de uma capacidade de generalização, e memorizando os dados de treino, podendo ter sua estimativa eficiente somente no conjunto de dados de treino).

Ressaltando que toda saída de todo neurônio j da rede, com n entradas, é calculada através da equação:

$$Y_{i\mu}(j) = F(\mu)$$

Onde $\mu = (\sum_{i=1}^n (X_{i_n}(j) * W_{i_n}(j) + b)$ é o potencial de ativação do neurônio j e F é a função de ativação do neurônio.

Fase Feedforward Nesta fase, em determinada época (k), em um vetor de n variáveis X_i correspondentes à um total de venda Y_i , todos os neurônios j propagam o sinal de saída através do calculo de $Y_{i\mu}(j) = F(\mu(j))$. No cálculo de μ os pesos sinápticos de cada neurônio j

conectado à uma entrada $X_{i_n}(j)$ são $W_{i_n}(j)$. A saída estimada pela rede será $Y_{i_\mu}(s)$.

Se a iteração i da época k for a primeira, todos os pesos $W_{i_n,j}$ em todos os neurônios, são inicializados com valores aleatórios pequenos.

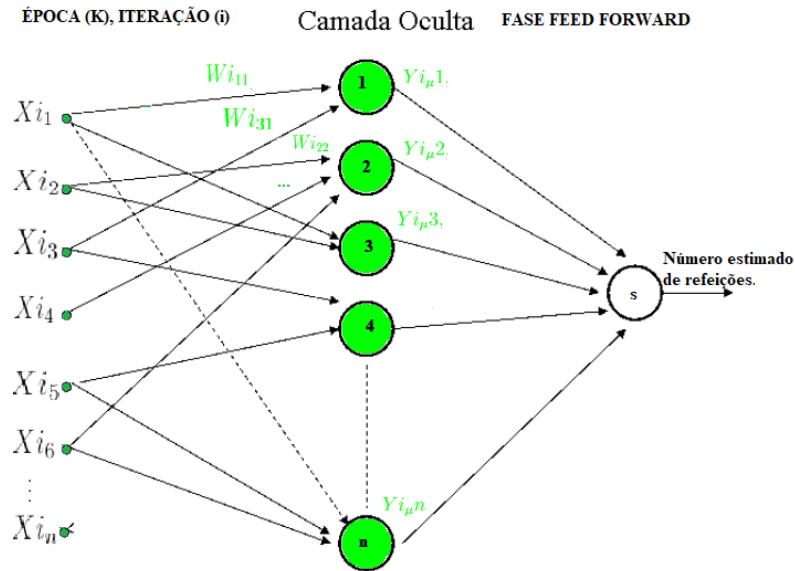


Figura 16 – Apresentação da i -ésima observação do conjunto de treino à rede

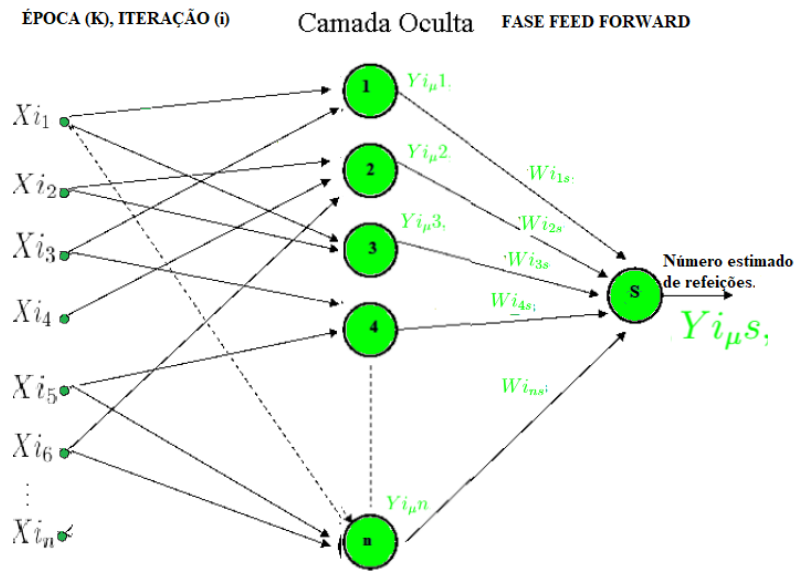


Figura 17 – Cálculo do valor estimado de saída da rede.

O valor de $Y_{i_\mu}(j)$. dos neurônios da primeira camada oculta, são obtidos de forma que $X_{i_n}(j)$ são os sinais de entrada das variáveis de 1 à n .

O valor de $Y_{i_\mu}(j)$. dos neurônios das próximas camadas, inclusive a de saída, utiliza o sinal de saída Y_{i_μ} dos neurônios da camada anterior conectados à j como um sinal de entrada X_{i_n} .

Fase Feedbackward O Processo de reajuste dos pesos sinápticos com o objetivo de atingir a minimalização do erro quadrático da validação, é feito pelo método do Gradiente Descendente, reajustando os pesos sinápticos $W_{i_n}(j)$ correspondente à cada neurônio j da rede, para um valor $\Delta W_{i_n}(j)$, da seguinte forma:

$$\Delta W_{i_n}(j) = \eta * \delta_j * X_{i_n}(j)$$

Regra δ para neurônio de saída Exclusivamente para a camada de saída, δ é calculado por

$$\delta_s = (Y_i - Y_{i_\mu}(s)) * F'(\mu(s)).$$

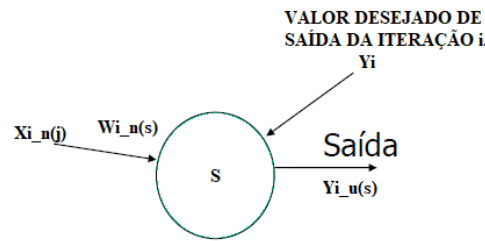


Figura 18 – Aprendizado de um neurônio na camada de saída

O valor dos n pesos $W_{i_n}(s)$ do neurônio de saída s com n entradas vindas das saídas de neurônios j das camadas anteriores, é obtido por $\Delta W_{i_n}(s) = \eta * \delta_s * Y_{i_\mu}(j)$

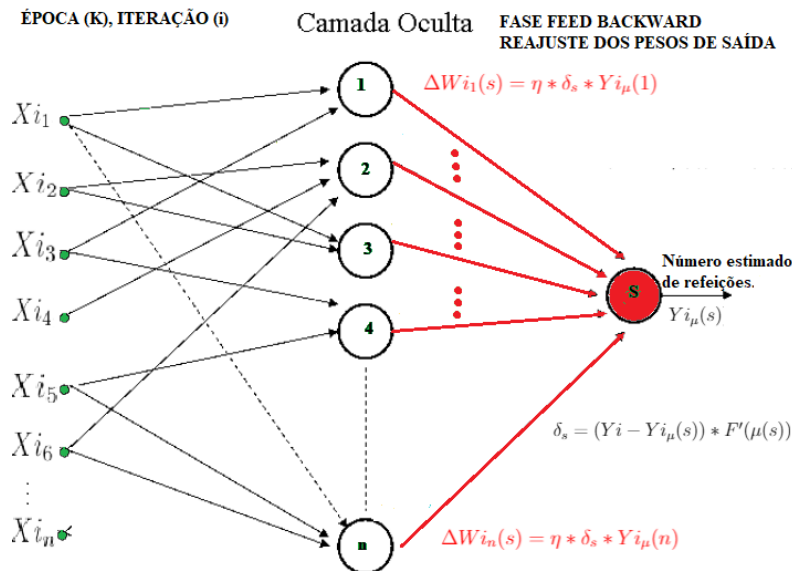


Figura 19 – Reajuste dos pesos de saída

Regra δ para neurônios de camadas ocultas Nesta etapa, os neurônios n da camada oculta obterão seu δ através da seguinte equação usando o δ dos m neurônios da camada poste-

rior que se conectam à ele:

$$\delta_n = F'(\mu(n)) * (\sum \delta_m * W_{im})$$

Na rede neural de 1 camada oculta e 1 neurônio de saída, a soma será reduzida à apenas 1 neurônio, ficando então a regra δ da seguinte forma:

$$\delta_n = F'(\mu(n)) * (\delta_s * W_{ns})$$

Por fim os pesos das n entradas $W_{in}(n)$ conectadas ao neurônio n terão seus pesos reajustados para a próxima iteração da seguinte forma:

$$\Delta W(i+1)_n(n) = \eta * \delta_n * X_{in}(n)$$

Após o reajuste chegar na primeira camada oculta, um novo vetor de observações i é apresentado à rede repetindo, iniciando novamente a fase feed-forward e repetindo o ciclo até o fim das entradas.

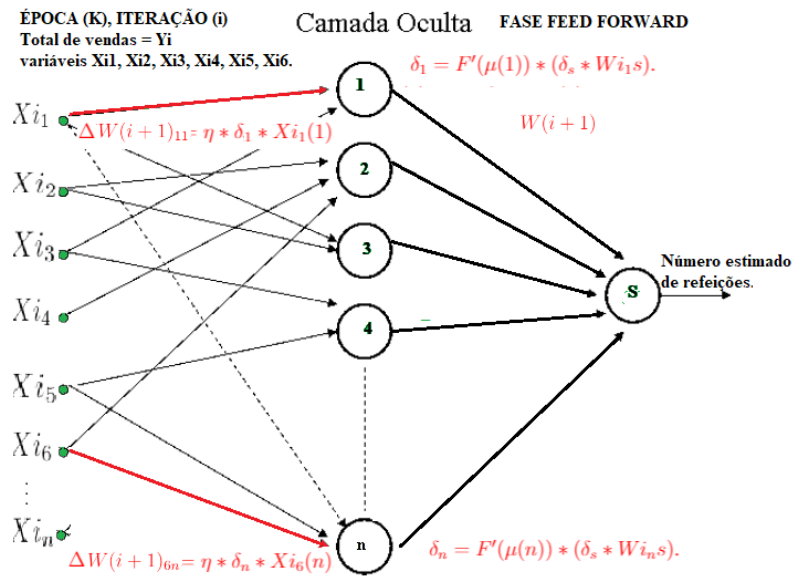


Figura 20 – Reajuste dos pesos de camadas ocultas

Validação de treino Após se encerrar uma época, que percorreu todas as i -ésimas entradas do conjunto de treino, um conjunto de dados de validação, previamente separado, é apresentado à rede. A validação ocorre só com a fase feedforward, obtendo-se os erros quadráticos da camada de saída com o dado de validação observado. Então é calculado o *m.s.e*, obtendo-se a média de todos os i -ésimos erros quadráticos do conjunto de validação. O valor *m.s.e* deve ser acompanhado até que se atinja um ponto ótimo, ou seja, um limite inferior após k épocas.

Quando este limite for atingido em uma época k_o , todos os parâmetros da rede dentro da época k_o serão os parâmetros desejados do modelo final da rede.

Quando a curva do *m.s.e* estiver sendo realizada, não se saberá o ponto ótimo de parada, até que ele comece a ser superado, logo é importante arbitrar um número l para salvar os parâmetros das l épocas anteriores.

2.2.8 Algoritmo de treino e validação backpropagation

Parâmetros de entrada:

- 'rna'[][] , Grafo de nós do tipo neurônio, da rede neural, onde rna[m][n] corresponde ao nó "n" da camada "m".
- valor real η , Taxa de aprendizado .
- Yt[][] e Yv[], Matrizes de valores reais Yt do conjunto de treino e Yv do conjunto de validação, onde Y[0][0] corresponde ao primeiro valor supervisionado, e Y[0][m] corresponde ao último parâmetro X relacionado ao valor supervisionado.
- L, Limite do contador de épocas, a cada novo erro mínimo o contador *m.s.e* é zerado, após o limite L do aumento de erros do *m.s.e* o treino é interrompido.

O principal critério de parada do treino, será o contador L. Durante o treino os valores de *m.s.e* sofrerão uma curva com similaridade parabólica, concavidade apontada para cima, conforme Figura 12 da seção 2.3.5. Busca-se então obter o valor mínimo desta curva. O treino trabalhará com 2 cópias do grafo rna[[]]. Sendo rnaAtual[[]] atual, e rnaMinimo[[]]. A cada novo valor mínimo encontrado, o rnaMinimo[[]] receberá o rnaAtual[[]], ao fim do treino quando o limite L de incremento de erros for superado, o treino será interrompido retornado o grafo rnaMinimo[[]]

1. Instancie rnaAtual[[]] e rnaMinimo[[]].
2. Instancie um contador inteiro de erros ce=0.
3. Copie rna[[]] para rnaAtual[[]] e rnaMinimo[[]].
4. Inicialize uma variável inteira, contador de épocas, ck = 0.
5. Inicialize uma variável real, errok.
6. Insira uma cópia de rna[[]] para rnaAtual[[]]
7. LAÇO: conte i1, Enquanto ce<L faça:
8. -LAÇO: conte i2, Percorra it[] em Yt[i2].

9. —Se $i2=0$, inicialize os pesos de $K[atual]$ com valores aleatoriamente pequenos.
10. —Execute a fase feedforward apresentando $it[]$ em $rnaAtual[][]$ e guarde o erro em $errok$.
11. —Execute a fase feedbackward reajustando os pesos de $rnaAtual[][]$ com $errok$.
12. —Inicialize um vetor temporario $ev[]$ de tamanho $Yv[].tamanho$, a quantidade de observações de treino.
13. —LAÇO: Conte $i3$, Percorra $iv[]$ em $Yv[i3]$.
14. —Execute a fase feedforward apresentando $iv[]$ em $rnaAtual[][]$ e insira o erro quadrático em $ev[i]$.
15. —Obtenha a média dos valores quadráticos do vetor $ev[]$ e insira em $mse[i1]$
16. —Destrua $ev[]$.
17. —Se $mse[i1] < mse[i1 - 1]$, $rnaMinimo[][]$ recebe $rnaAtual[][]$, $ce = 0$.
18. —incremente ck , $ck=ck+1$.
19. Retorne $rnaMinimo[][]$

3 Trabalhos relacionados

3.0.1 ANÁLISES EM RESTAURANTES UNIVERSITÁRIOS

No estudo estatístico feito por (LANDIM, 2016), foi analisada a correlação entre a temperatura e o consumo nos dias de vendas do restaurante universitário do campus ICT da Unifesp, sendo que os dados continham uma amostra das vendas do segundo semestre de 2016. Devido ao baixo volume de ocorrências, os dados foram submetidos à reamostragem via bootstrap. De acordo com os gráficos das amostras, identificou-se que a correlação mostrada nos gráficos da primeira metade do semestre e do período total do semestre formaram distribuições bimodais. Porém, na segunda metade do semestre formou-se uma distribuição unimodal. Portanto, concluiu-se que outras variáveis e outros modelos de análises podem ser utilizados para tal análise de demanda.

(LOPES, 2008) faz o mesmo estudo deste cenário do ICT UNIFESP aplicado na Universidade Federal de Viçosa (UFV). Neste estudo, os dados utilizados foram somente o histórico de vendas do restaurante universitário, e nenhuma variável de ambiente foi coletada como temperatura, precipitação, número de alunos matriculados, etc. O algoritmo utilizado foi o Traincgp (Conjugate gradient backpropagation with Polak-Ribiere updates) no software Matlab. Este algoritmo não envolve o cálculo das derivadas segundas das variáveis e converge ao mínimo da função quadrática em um número finito de iterações como cita o autor. Foram então considerados para cada nó da rede neural, o dia da semana (como segunda, terça, quarta, quinta e sexta) e cada camada dessa rede utilizando os 5 dias anteriores para cada nó (as 5 segundas anteriores, 5 terças anteriores e assim sucessivamente) e, por fim, obtido um modelo pela rede que apresentou erro máximo de 3.

(ROCHA JOSÉ CELSO; MATOS, 2011) também realiza o estudo de demanda no restaurante universitário da Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP), novamente com os métodos de redes neurais artificiais com backpropagation, e utilizando apenas como fonte de dados o histórico numérico das vendas realizadas, e outras variáveis intermediárias obtidas a partir deste como médias de subconjunto de observações (médias de segundas-feiras), e a única variável de ambiente coletada foi o número de feriados próximos à observação de venda. No estudo do total de dias analisados, verifica-se que em 73% (187 dias), o método de média simples propiciou um maior erro em relação à RNA, que por sua vez ocasionou um erro maior nos 23% (69 dias) restantes. Em se tratando de menor desperdício, observa-se que a RNA apresenta erros maiores que 50 refeições em 13 dias, enquanto o método da média simples apresenta erros maiores que 50 refeições em 58 dias, concluindo-se então que o método de

RNA foi bem mais eficiente do que o cálculo de média simples utilizado pela administração do restaurante universitário.

3.0.2 ANÁLISES EM OUTROS CENÁRIOS

([RUAS, 2012](#)) fazem uma análise de previsão de demanda de energia elétrica no estado do paran , entre os anos de 2004 e 2006, utilizando redes neurais artificiais e m quinas de vetores de suporte. Apesar de n o ser o mesmo exemplo do cen rio do restaurante universit rio do ICT UNIFESP, temos na distribui  o dos dados de consumo coletados como uma s rie temporal, e foi utilizado nesta pesquisa uma rede parcialmente recorrente de Elman, que permite a previs o de um passo de tempo   frente. Para que seja poss vel realizar a previs o v rios pontos   frente,   necess rio utilizar os valores j  previstos, ou seja, a sa da da rede, como entradas da mesma. Logo, como o cen rio da demanda do RU no ICT UNIFESP se apresenta como uma s rie temporal e tentaremos prever a demanda um passo   frente, ser  interessante a este trabalho utilizar e comparar as t cnicas previstas nesta refer ncia.

([ALMEIDA, 2013](#)) analisa um cen rio semelhante de demanda de energia el trica, por m utilizando-se t cnicas de previs o de demanda com Rede Neural Artificial do tipo Multilayer Perceptron combinado com l gica fuzzy que permite colocar vari veis de temperatura (entre outras) em um conjunto de regras que impactam no problema. Esta an lise   interessante pois   semelhante   refer ncia 2 que faz um estudo com redes neurais perceptron no mesmo tema deste trabalho de conclus o de curso, mas com a adi  o da l gica fuzzy que permite utilizar outras vari veis de ambiente impactantes, que n o sejam somente o hist rico num rico do consumo.

([SILVA, 2010](#)) tamb m aplica t cnicas de redes neurais para previs o de demanda de energia el trica, com o estudo de vari veis clim ticas, por m atrav s de um modelo de MAPA SOM - (Self-Organizing Map) que   um tipo de rede neural desenvolvido para reconhecimento de padr es. Apesar de ser um modelo n o supervisionado, o modelo   ideal para organizar as principais vari veis impactantes e descart veis na previs o. O mapa som utilizado pelo autor apresenta os dados associados aos seus neur nios de forma que padr es similares encontram-se em neur nios cont guos, tendo uma organiza  o topol gica. Deste modo   poss vel se extrair rela  es abstratas entre as vari veis do vetor de dados atrav s da sua posi  o nos mapas componentes, que por meio de uma escala de cores mostram a quantidade de uma vari vel espec fica em cada neur nio do mapa.

3.0.3 TESE COMPARATIVA DE DIVERSOS M TODOS DE PREVIS O DE DEMANDA.

([JUNIOR, 2007](#)) realiza um trabalho de compara  o entre diversas t cnicas de M todos estoc sticos e Rede Neural Artificial (RNA), para a previs o da demanda de produtos cosm ticos,

distribuídos em séries temporais. Entre eles Redes Neurais tipo feedforward, com o algoritmo backpropagation que foi o principal foco no trabalho de previsão de RU na Universidade Federal de Viçosa e na Universidade Estadual Paulista Júlio de Mesquita Filho. Será realizado a análise dessas comparações e seleção de principais algoritmos que possam ser aplicados e comparados neste trabalho.

4 Plano de atividades para o TCC II

Para a continuação deste trabalho o cronograma abaixo deve ser seguido.

Tabela 3 – Plano de atividades para o TCC II

Atividades	Março	Abril	Maio	Junho	Julho
1	✓				
2	✓				
3		✓			
4		✓	✓		
5		✓	✓	✓	
6				✓	
7				✓	✓

1. Carregamento, limpeza e pré-processamento do conjunto de dados do R.U do ano de 2017. Será usada a ferramenta Matlab;
2. Construção de modelo de previsão, conforme figura seguinte, via regressao linear multipla utilizando-se os dados de 2017 para prever os de 2018; A medicao de qualidade do modelo sera feita por meio das seguintes medidas: o Erro Absoluto Medio (EAM), Erro Quadrado Medio (EQM) e Raiz do Erro Quadrado Medio (REQM);
3. Implementação e teste da Rede Neural Artificial do tipo Perceptron Multi-camadas com algoritmo de backpropagation para treinamento;
4. Realizar os experimentos para se medir o desempenho preditivo da RNA;
5. Implementar a abordagem de comitê de RNAs via bootstrap (diferentes amostragem dos dados para cada rede) e bagging (predição final sendo a média das predições das redes);
6. Realizar os experimentos para se medir o desempenho preditivo da RNA com o comitê;
7. Escrever o TCC 2

4.1 Dados

Abstraindo o histórico de vendas do restaurante, temos 2 tipos de dados principais. A venda podendo ser uma observação Y , e a data da venda sendo uma observação X com múltiplas variáveis. Se comportando de modo que a função de X implica em Y .

$$X, Y \rightarrow Y \simeq f(X)$$

$$Demanda : Y \simeq Data : f(X)$$

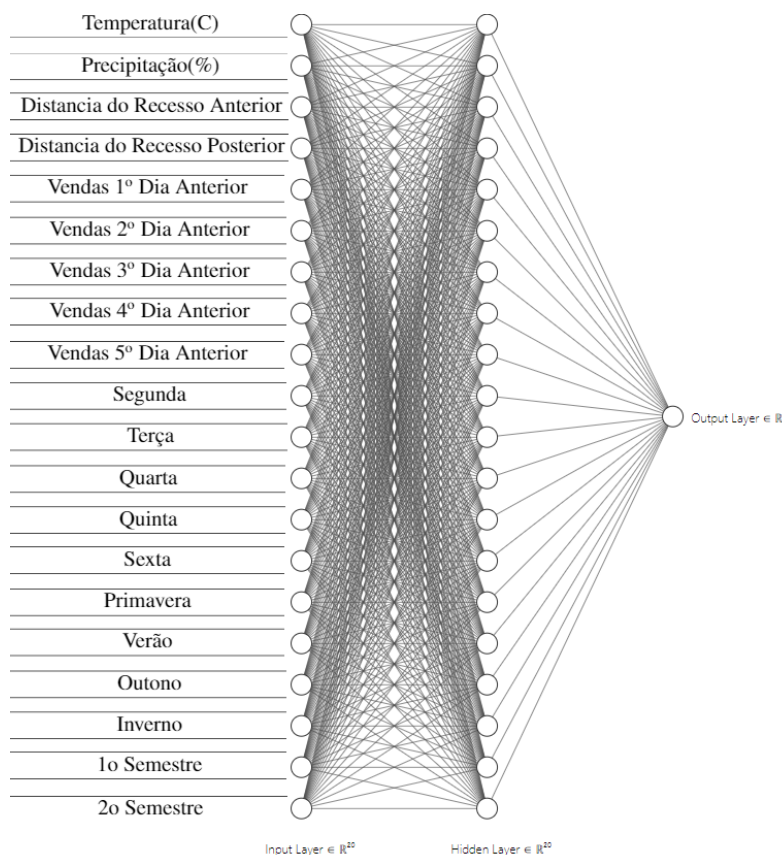


Figura 21 – Proposta de RNA, Figura gerada a partir da ferramenta online: <http://alexlenail.me/NN-SVG/index.html>

A observação principal a ser analisada neste trabalho é o número de vendas do R.U do ICT - UNIFESP, em valor inteiro, obtido de um dia letivo, em um determinado período sendo este o almoço ou jantar.

O tempo, onde a informação principal se distribui, é em formato de data, e contém apenas 1 valor da venda total de almoços nesta data, e da venda total de jantãs nesta data.

4.1.1 Dados de Consumo e de Data.

Os dados históricos de consumo no restaurante foram retirados do atual sistema banco de dados de refeições subsidiadas do Hospital São Paulo, que gerencia os dados dos refeitórios de todas as unidades da Unifesp. É importante ressaltar que tais dados são apenas de funcionários, docentes e discentes da instituição. Eventuais compras de refeições realizadas por visitantes que não possuem vínculo com a instituição não são registradas pelo sistema.

Outro fato importante é que estes dados registrados no banco, são registrados no momento em que a refeição é realizada, pelos terminal da entrada do refeitório e não no momento que é vendida pelo caixa do refeitório. Caso fossem registrados no momento da venda no caixa, a análise estatística do consumo poderia sofrer um viés errôneo em relação à

frequência de pessoas no refeitório.

Os alunos são os que realizam o maior volume de consumo, e são os que tem um padrão de frequência à universidade de modo estocástico, que tende a ter uma sazonalidade anual de frequência na universidade desde 2017, quando a instituição padronizou a grade horária de disciplinas. Além de ser o objetivo principal da previsão por serem os únicos que têm as refeições subsidiadas, pagando um custo de apenas R\$2,50 por refeição, onde entende-se que este custo pode ter mais forte influência na decisão do aluno realizar uma refeição em comparação à outro discente ou pessoa com vínculo na instituição. Por isso os dados que serão analisados neste trabalho, serão somente da frequência de alunos, e terão como início o ano de 2017.

Apenas alguns funcionários autorizados tem acesso ao banco de dados do sistema de refeições da instituição, entre eles o fiscal de contrato do restaurante universitário. Para obter tais dados neste trabalho, foi necessário obter uma autorização com a direção do campus ICT - UNIFESP e em seguida solicitar a exportação dos dados ao fiscal. Os parâmetros dos quais o mesmo consegue realizar a exportação de dados são diversos, porém foi solicitado o formato de valores separados por vírgula que compreendem o ano de 2017 a 2018.

O modelo exportado segue o seguinte da seguinte forma abaixo, contendo o exemplo de 2 datas:

```

1 "CONSULTA POR PERÍODO " ,,,,,,, "CAMPUS: SÃO JOSÉ DOS CAMPOS " ,,, "RU:
  TODOS " ,,, "VINCULO: ALUNO " ,,, "PERÍODO DE: 01/01/2017 A 31/10/2018 " ,,,
  DATA,VENDAS CAFÉ,VENDAS ALMOÇO,VENDAS JANTAR,VENDAS
  REFEIÇÃO*,TOTAL VENDAS,ENTR. CAFÉ,ENTR. ALMOÇO,ENTR.
  JANTAR,TOTAL ENTR. REFEIÇÃO*,TOTAL ENTRADA
  (31/10/2018),0,395,0,395,395,0,362,0,362,362
  (30/10/2018),0,667,0,667,667,0,437,256,693,693

```

Logo, no exemplo da segunda linha deste modelo de dados, serão analisados o primeiro valor "(30/10/2018)" como a primeira variável de entrada que corresponde à data, o oitavo valor "437" que corresponde à variável de saída do total de almoço, e ao nono valor "256" que corresponde à variável de saída do total de janta.

Em todo os bancos de dados gerenciados pelo setor de tecnologia da informação da Unifesp, do campus ICT, o mesmo informou que foi constam 428620 refeições subsidiadas no banco de dados do sistema antigo, no período de 2011 à 2016, e 111454 refeições no período de 2017 a 01/08/2018 que fecham o modelo de contrato antigo. O calculo do custos médios cobrados pelo restaurante de R\$9,14 e R\$2,50 pelo aluno, estima que ambos gastaram respectivamente R\$4.936.276,36 e R\$1.350.185,00 enquanto os restaurantes que estiverem em contrato nesse período com a universidade, tiveram um faturamento bruto acumulado de R\$6286461,36

Já neste atual 2º semestre de 2018, iniciando em 01/08/2018 à 31/10/2018, o valor total subsidiado pela universidade é de R\$179019,52, R\$82270,00 pelos alunos e o faturamento bruto acumulado pelo restaurante é de R\$261289,52

É importante observar que na implementação computacional das análises de predição, o campo de data é utilizado para o controle da informação e união dos registros climáticos e de consumo, além ser utilizado para obtenção de dados derivados do tempo como o dia da semana, o período do semestre e a estação do ano. Já os campos de número total de refeições consumidas, devem constar como formatos quantitativos nas estruturas de dados dos algoritmos a serem implementados.

Informações derivadas da data não necessitam de coleta, como por exemplo dia da semana, pois podem ser obtidas por métodos de cálculos, sabendo-se apenas um dia da semana correspondente à um dia do ano, ou utilizando bibliotecas computacionais, no caso a biblioteca datetime para a linguagem python.

Para a obtenção de feriados e recessos do calendário acadêmico de um ano determinado a informação deve ser obtida através do link oficial da instituição, <http://www.unifesp.br/reitoria/prograd/pro-reitoria-de-graduacao/informacoes-institucionais/calendario-academico>.

A única forma de exportação desta informação é em pdf e em formato não tratável facilmente por algoritmos. Então os dias de recesso respectivos ao campus ICT-UNIFESP devem ser carregados manualmente nos códigos a serem implementados.

4.1.2 Dados Climáticos

Também serão analisadas variáveis climáticas juntos com os dados de consumo, de forma a analisar a influência de fatores externos como temperatura média ambiente e precipitação. Tais dados podem ser obtidos de forma gratuita pelo BDMEP - Banco de Dados Meteorológicos para Ensino e Pesquisa, pertencente à instituição pública INMET - Instituto Nacional de Meteorologia, pertencente ao MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO do Governo Brasileiro.

É necessário um cadastro no site <http://www.inmet.gov.br/portal/index.php?r=bdmep/bdmep> para a obtenção dos dados.

A instituição contém dados registrados de forma digital desde 1961 no país inteiro, os dados históricos referentes a períodos anteriores a 1961 ainda não estão em forma digital e, portanto, estão indisponíveis no BDMEP.

O BDMEP pode fornecer os registros de Precipitação(mm), Temp Máxima(°C), Temp Mínima(°C), Insolação(horas) Evaporação do Piche(mm), Temperatura Compensada Média(°C), Umidade Relativa Média(%), Velocidade Vento Média(mps).

Importante ressaltar que o BDMEP leva 90 dias para registrar cada nova data. Os dados utilizados para este trabalho foram da estação 83781 - São Paulo MIR de Santa - SP.

Observa-se que para futuros trabalhos no atual campus ICT-UNIFESP, a estação TAU-BATE - SP 83794, que no momento tem o banco de dados vazio, poderá ser utilizada.

Serão analisadas a temperatura máxima do dia que tem registro em momento próximo à frequência de refeições no horário de almoço, a temperatura mínima do dia que está próxima à janta, e a umidade relativa média que pode indicar presença de chuva no dia, que por sua vez causa diversos impactos ao público da região como a própria frequência do mesmo na instituição, principalmente no fator logístico.

Os dados se apresentam da seguinte forma:

```

1 Estacao;Data;Hora;Precipitacao;TempMaxima;TempMinima;Insolacao;Evaporacao
  Piche;Temp Comp Media;Umidade Relativa Media;Velocidade do Vento Media;
  83781;01/01/2017;0000;;31.7;;8.5;6.1;25.2;66;3.466667;
  83781;01/01/2017;1200;0;;18.1;;;;;
  83781;02/01/2017;0000;;31.3;;5.6;6.3;23.96;83;2.9;
  83781;02/01/2017;1200;16;;18.1;;;;; 83781;03/01/2017;0000;;31.3;;7;4.5;23.1;87;3;
  83781;03/01/2017;1200;3.7;;19.2;;;;;

```

Observa-se que cada data contem 2 linhas, sendo a primeira linha com o terceiro campo de valor "0000"apresentando a temperatura máxima do dia no quinto campo, e a segunda linha com o terceiro campo de valor "1200"contendo a temperatura mínima do dia no sexto campo. A umidade relativa media se encontra no décimo campo da primeira linha de cada data. Todas as datas apresentam sempre o mesmo padrão de 2 linhas.

4.1.3 Estruturação dos Dados

Neste trabalho temos múltiplas relações causais, como dados climáticos e dados derivados da data. Logo será usado um algoritmo de regressão linear múltipla na primeira estruturação de dados, seguido de algoritmos de aprendizado de máquinas que utilizarão a mesma estrutura, transportando o valor de cada parâmetro causal para cada variável X_i . O objetivo é encontrar a linearidade dessas relações, para que seja possível prever um valor de venda futura, dado os valores de observações futuras.

O modelo inicial de dados, para o exemplo das 2 primeiras datas letivas de 2018, seguem com os dados de entrada de consumo e em seguida os dados de entrada de clima:

Modelo obtido do histórico de dados do R.U.

Modelo obtido do histórico do BDMEP.

```
1 DATA,VENDAS CAFÉ,VENDAS ALMOÇO,VENDAS JANTAR,VENDAS
  REFEIÇÃO*,TOTAL VENDAS,ENTR. CAFÉ,ENTR. ALMOÇO,ENTR.
  JANTAR,TOTAL ENTR. REFEIÇÃO*,TOTAL ENTRADA;
  (26/02/2018),0,930,10,940,940,0,446,190,636,636;
  (27/02/2018),0,640,13,653,653,0,470,237,707,707;
```

```
1 83781;26/02/2018;0000;;27.2;;1.1;2.9;22.2;88;1.6; 83781;26/02/2018;1200;0;;19.6;;;;;
  83781;27/02/2018;0000;;28.5;;3.6;1.1;22.92;84;2.5;
  83781;27/02/2018;1200;73.1;;19;;;;;
```

Os dados de consumo no modelo reduzido de previsão de almoço, fica somente com a data da venda, e o total de refeições realizadas no almoço.

```
1 DATA,ENTR. ALMOÇO; (26/02/2018),446; (27/02/2018),470;
```

O modelo reduzido é novamente estruturado para cada linha conter as 5 ultimas observações de vendas anteriores, conforme modelo da figura 14.

No caso dos primeiros dias do semestre que não contêm todas as 5 vendas anteriores, o valor será preenchido com 0.

```
1 Observações Históricas
  DATA,ENTR. ALMOÇO, 1º Dia Anterior, 2º Dia Anterior, 3º Dia Anterior, 4º Dia
  Anterior, 5º Dia Anterior;
  (26/02/2018),446,0,0,0,0,0;
  (27/02/2018),470,446,0,0,0,0;
```

Na tabela de dados climáticos, a redução ocorre para o valor da data, na linha 0000, contendo o valor de temperatura máxima, e precipitação.

A informação de data, irá gerar informações derivadas de data como o valor em forma de fator do dia da semana (obtido através de métodos que retornam o dia da semana, dado uma data, numero do semestre sendo 1 para o primeiro, e 2 para o segundo).

Um vetor de feriados, recessos acadêmicos e finais de semana colhidos do link da prograd, será preenchido manualmente. Cada linha de observação fará um calculo de distância de data sobre este vetor, e o calculo retornará as 2 menores distâncias, sendo o primeiro a distância do recesso anterior mais próximo, e o segundo valor a distância do recesso posterior mais próximo.

1 26/02/2018;27.2;88; 27/02/2018;28.5;84;

O dia da semana seguirá o modelo binário de acordo com os trabalho de previsão de demanda em R.U realizados por (LOPES, 2008) e (ROCHA JOSÉ CELSO; MATOS, 2011).

Dia da semana	Valor				
Segunda-feira	1	0	0	0	0
Terça-feira	0	1	0	0	0
Quarta-feira	0	0	1	0	0
Quinta-feira	0	0	0	1	0
Sexta-feira	0	0	0	0	1

Figura 22 – Entradas de dia da semana em cofatores. Retirado de (LOPES, 2008).

Os dados estruturados ficarão de acordo com a Tabela 1, onde a coluna X_i receberá os dados da coluna $i + 1$ 3ª linha até a 2ª até a 15ª linha, e o valor Y_i receberá o dado da coluna $i + 1$ na 1ª linha:

Por fim a partir da tabela, os dados serão separados em forma de matrizes para aplicação nos métodos de regressão linear e aprendizado de máquina, conforme o exemplo abaixo.

$$Y = \begin{bmatrix} 446 \\ 470 \end{bmatrix} \quad X = \begin{bmatrix} 27.2 & 88 & 1 & 4 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 28.5 & 84 & 1 & 4 & 446 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Tabela 4 – Tabela de dados do R.U.

	26/02/2018	27/02/2018
VENDAS - Y_i	446	470
PARÂMETROS X_i		
Temperatura(C)	27.2	28.5
Precipitação(%)	88	84
Distancia do Recesso Anterior	1	2
Distancia do Recesso Posterior	4	3
Vendas 1º Dia Anterior	0	446
Vendas 2º Dia Anterior	0	0
Vendas 3º Dia Anterior	0	0
Vendas 4º Dia Anterior	0	0
Vendas 5º Dia Anterior	0	0
Segunda	1	0
Terça	0	1
Quarta	0	0
Quinta	0	0
Sexta	0	0
Primavera	0	0
Verão	1	1
Outono	0	0
Inverno	0	0
1o Semestre	1	0
2o Semestre	1	0

5

Conclusão

As análises diversas de previsão de demanda para o tema abordado requerem extensos métodos de implementação e estruturação de dados. Uma das etapas mais importantes do trabalho, é o método de coleta de dados. Muitos podem ser de acesso burocrático, ou de difícil busca. E são o requisito primordial para o início de qualquer análise.

A diversidade de métodos de aprendizado de máquina é imensurável, e dentro de apenas uma análise, que é o treino com retropropagação, pode-se montar infinitas topologias diferentes com base na estrutura dos dados coletados. As heurísticas sobre a definição de topologia apesar de diversas, não são determinísticas, e o processo requer análise exploratória, subjetiva e de tentativa e erro sobre o tema e problema a ser abordado.

Todavia foi notório a eficiência dos modelos de aprendizado de máquina em trabalhos relacionados à restaurantes universitários.

Como no ICT - Unifesp, não há qualquer modelo atual de previsão, e a falta de um modelo causa desperdício de alimentos e prejuízo ao restaurante, a abordagem dessa pesquisa e sua continuação com novos métodos após a conclusão deste trabalho torna-se viável.

Referências

- ALMEIDA, V. A. d. *Previsão De Carga Através De Modelos Neuro-fuzzy*. 91 p. Tese (Doutorado) — Universidade Federal do Rio de Janeiro, 2013. Disponível em: <<http://monografias.poli.ufrj.br/monografias/monopoli10007620.pdf>>. Citado 9 vezes nas páginas 13, 24, 29, 30, 37, 38, 45, 48 e 58.
- BRYAN KOLB; ARIF, M. Harnessing the power of neuroplasticity for intervention. In: *FRONTIERS IN HUMAN NEUROSCIENCE*. [S.l.: s.n.], 2014. p. 13. Citado na página 33.
- CLARICE G.B. DEMÉTRIO; ZOCCHI, S. S. *Modelos de Regressão*. [S.l.]: Departamento de Ciências Exatas, ESALQ, USP, 2011. Citado 3 vezes nas páginas 24, 30 e 31.
- HAYKIN, S. *Neural networks, a comprehensive foundation*. [S.l.]: Macmilan, 1994. Citado na página 24.
- JUNIOR, A. M. *Análise de métodos de previsão de demanda baseados em séries temporais em uma empresa do setor de perfumes e cosméticos*. Dissertação (Mestrado) — Pontifícia Universidade do Paraná. Curitiba, 2007. Citado 8 vezes nas páginas 13, 24, 27, 28, 29, 30, 36 e 58.
- LANDIM, D. D. *Análise de demanda no restaurante universitário do instituto de ciência e tecnologia da unifesp - universidade federal de são paulo. Projeto da Disciplina de Simulação de Sistemas*, Universidade Federal de São Paulo, v. 1, n. 1, p. 6, 2016. Disponível em: <<https://github.com/ddlandim/monografy-ann-demand-prediction/tree/master/referencias>>. Citado na página 57.
- LOPES, L. Modelos para previsão de demanda no restaurante universitário utilizando técnicas de redes neurais. *Congresso Nacional De Excelência Em Gestão. Universidade Federal De Viçosa.*, Universidade Federal De Viçosa, IV, n. 1, p. 14, 2008. Disponível em: <<http://www.novoscursos.ufv.br/projetos/ufv/gesep/www/wp-content/uploads/5.>> Citado 7 vezes nas páginas 13, 24, 46, 47, 48, 57 e 67.
- LUGER, G. F. *Inteligência Artificial. Estruturas e Estratégias para a Solução de Problemas Complexos*, 4ª ed. [S.l.]: Porto Alegre: Bookman, 2004. Citado na página 32.
- MARIANO, F. C. M. Q. *REdes Neurais na predição de valores energéticos de alimentos para aves*. 130 p. Tese (Doutorado) — Universidade Federal de Lavras, 2014. Citado 7 vezes nas páginas 13, 24, 36, 37, 44, 47 e 48.
- MORETTIN, P. A. *Análise de séries temporais*. [S.l.]: Editora Blucher, 1987. Citado na página 28.
- RAGA A. DE P.; CARVALHO, A. P. L. F. L. *Redes neurais artificiais: teoria e aplicações*. [S.l.]: Livros técnicos e científicos, 2000. Citado na página 47.
- ROCHA JOSÉ CELSO; MATOS, F. D. F. Utilização de redes neurais artificiais para a determinação do números de refeições diárias de um restaurante universitário. *Rev. Nutr., Campinas*, Universidade Estadual Paulista Júlio de Mesquita Filho., n. 1, p. 8, 2011. Citado 7 vezes nas páginas 13, 24, 46, 47, 49, 57 e 67.

RUAS, G. I. e. S. Previsão de demanda de energia elétrica utilizando redes neurais artificiais e support vector regression. *Copel Distribuição S/A (COPEL)*, Universidade Federal do Paraná(UFPR), v. 1, n. 1, p. 10, 2012. Disponível em: <http://www.dcc.fc.up.pt/~ines/enia07_html/pdf/27927.pdf>. Citado 4 vezes nas páginas 13, 24, 30 e 58.

SILVA, C. e. a. Metodologia para previsão de demanda de energia elétrica em curtíssimo prazo baseada em um modelo probabilístico discreto. *Desenvolvimento de Algoritmos e Software para Previsão de Carga na Área de Concessão da MuxEnergia - Horizonte de Curtíssimo Prazo*, Universidade Federal do Pampa - UNIPAMPA, v. 1, n. 1, p. 4, 2010. Disponível em: <http://alegrete.unipampa.edu.br/gesep/wp-content/uploads/2010-05/Previs%C3%A3o-de-demanda-utiilizando-m%C3%A9todo-de-Markov_Final.pdf>. Citado 3 vezes nas páginas 24, 30 e 58.

ZORZETTO, R. Recontagem de neurônios põe em xeque ideias da neurociência. *Revista Fapesp 192*, Fapesp, v. 1, n. 1, p. 6, 2012. Citado na página 33.