INSTITUTO FEDERAL DE EDUCAÇÃO,

CIÊNCIA E TECNOLOGIA DE MINAS GERAIS

MARCO AURÉLIO MONTEIRO LIMA

**TOPICOS ESPECIAS EM ENGENHARIA DE COMPUTAÇÃO:**

TREINAMENTO DE SÉRIE TEMPORAL COM MLP

BAMBUÍ

2017

**Análise do Algoritmo de treinamento e previsão**

Nesse trabalho foi utilizado o dataset da estação meteorológica de Bambuí, referentes ao ano de 2011 a 2016.

O objetivo proposto é desenvolver uma rede neural que utilize os dados referentes de 2015 para prever a temperatura referente a 2016.

Primeiramente foi feita uma análise do dataset para verificação e remoção de possíveis outliers.

Os dados do dataset foram carregados utilizando a função carrega dataset. Que lê um arquivo de texto separando os dados com “;”. Essa função lê a sequência de símbolos e quando encontra um “;” pula para próxima coluna.

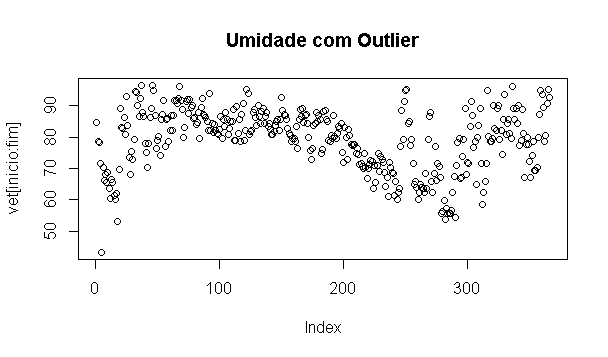
**Ex**: Estacao;Data;Hora;TempMaxima;UmidadeRelativaMedia; Cabeçalho

83582;01/01/2011;0000;24.4;89.5; 1 linha com dados

O cabeçalho é lido e a primeira sequência de símbolos da primeira linha é associado com o primeiro cabeçalho e assim por diante.

Primeiro foi analisado os valores da umidade para correção de outliers.

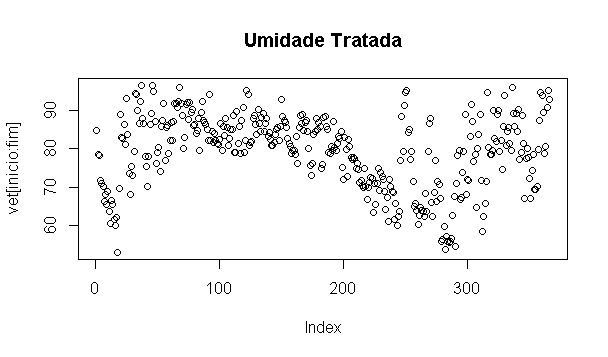
Figura 1: Gráfico 2d representando a umidade sem correção de outliers;



O gráfico acima mostra a umidade relativa sem tratamento de outliers.

Agora com os ajuste de outliers podemos ver na figura 2 logo abaixo, que os erros não são visíveis.

Figura 2: Gráfico 2d representando a umidade com correção de outliers;



Como não foi possível a identificação de outliers vamos plotar outros tipos de gráficos.

Figura 3: Gráfico histograma representando a umidade sem correção de outliers;

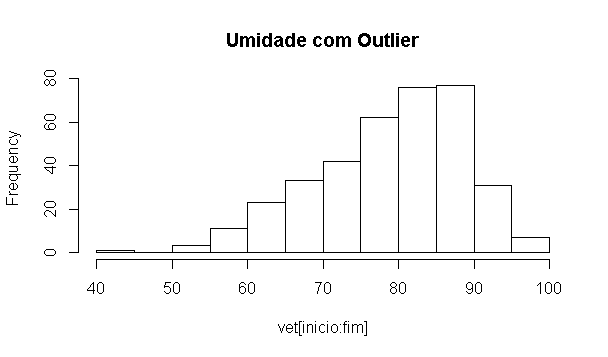
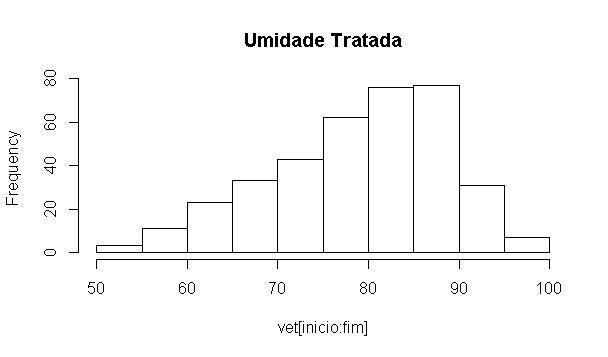


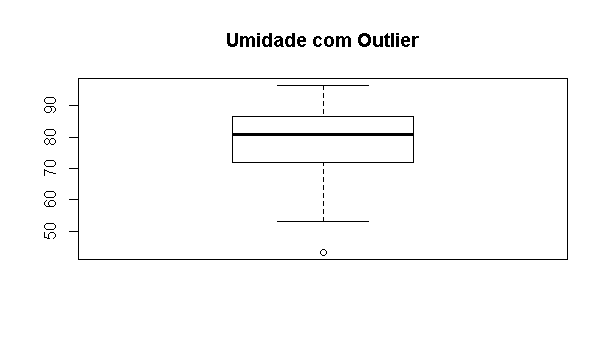
Figura 4: Gráfico histograma representando a umidade com correção de outliers;



Na figura 4 é possível perceber uma melhor visualização dos erros comparando coma figura 3. Por exemplo na figura 3 havia um vazio entre 40 e 50.

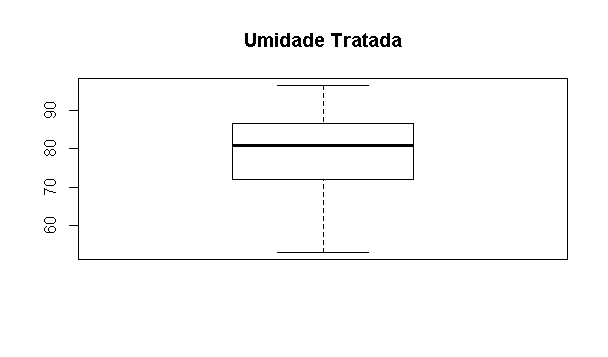
Plotando agora gráficos do tipo boxplot será possível visualizar ainda melhor esses outliers.

Figura 5: Gráfico boxplot representando a umidade sem correção de outliers;



O ponto bem abaixo do box representa o outlier da série umidade.

Figura 6: Gráfico boxplot representando a umidade com correção de outliers;



Agora analisando a série de temperatura podemos ver os mesmos plots utilizados na série umidade.

Figura 7: Gráfico 2d representando a temperatura sem correção de outliers;

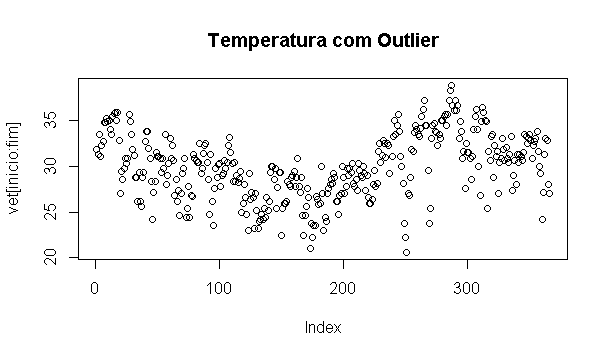
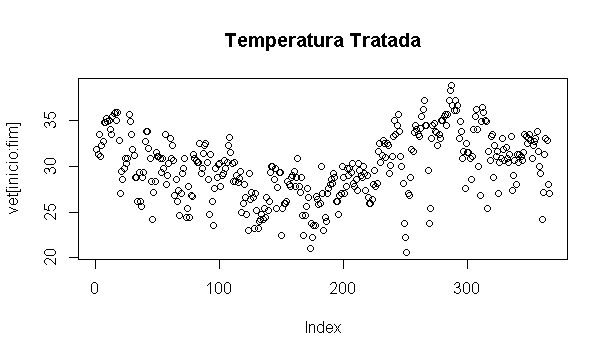


Figura 8: Gráfico 2d representando a temperatura com correção de outliers;



Analogamente a umidade podemos ver que esse tipo de gráfico não é tão eficiente para visualização de outliers.

Figura 9: Gráfico histograma representando a temperatura sem correção de outliers;

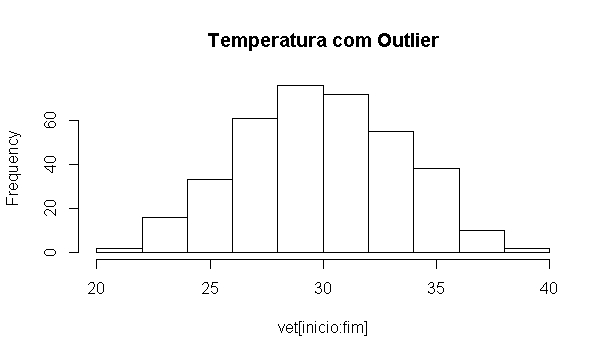
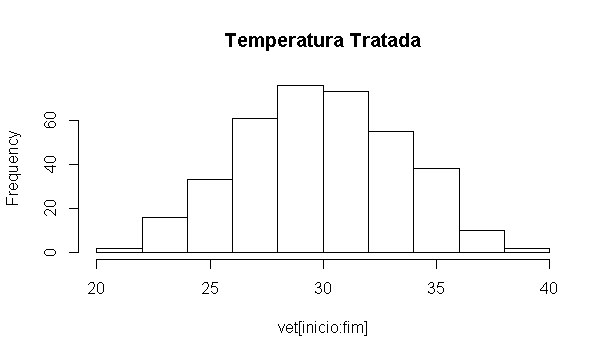


Figura 10: Gráfico histograma representando a temperatura com correção de outliers;



O Histograma também não foi tão eficiente na verificação de outliers. Não teve praticamente nenhuma alteração.

Figura 11: Gráfico boxplot representando a temperatura sem correção de outliers;

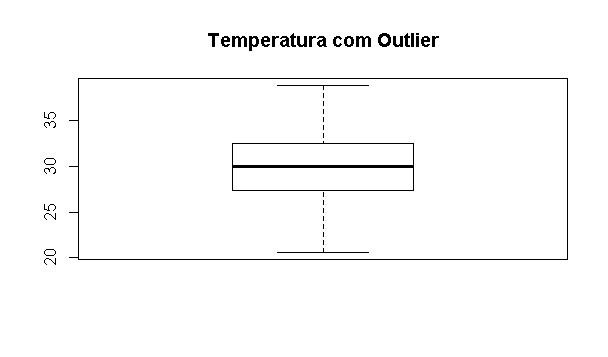
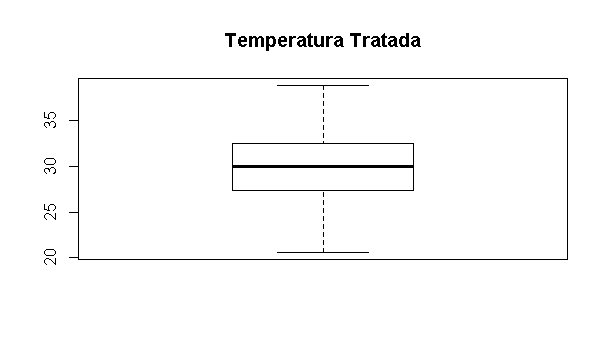


Figura 12: Gráfico boxplot representando a temperatura com correção de outliers;



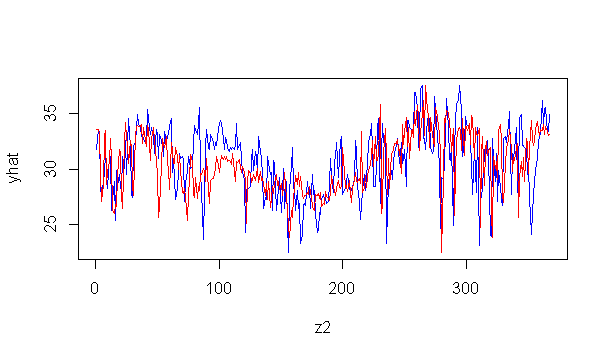
Aparentemente pode-se notar que a série temperatura não teve tantos outliers visíveis. Um dos motivos de toda essa análise e correção antes de começar a treinar a previsão da rede deve-se ao fato de estarmos trabalhando com uma base de dados real, e simplesmente um dia pode não ter sido colhido a temperatura e umidade. E realmente teve dias, que não houve coleta. Para correção desses erros, os dados nulos foram substituídos por dados referentes ao mesmo dia do ano anterior.

Com rede tratada foram padronizados os dados, da temperatura e da umidade, que foi nada mais do que colocar os valores referente aos mesmos numa proporção de 0 e 1.

Foi especificado para rede 10 neurônios e 15000 épocas. Para o treinamento da rede foi utilizado a biblioteca “RNNS” do R que traz a implementação pronta da rede MLP, que foi o algoritmo utilizado. Para a previsão foi utilizado apenas os dados do ano de 2015, umidade e temperatura, que anteriormente foram tratados,

A rede foi treinada utilizando a seguinte lógica, pegar a umidade e a temperatura do dia anterior ao que se quer prever, e assim prever a temperatura do dia.

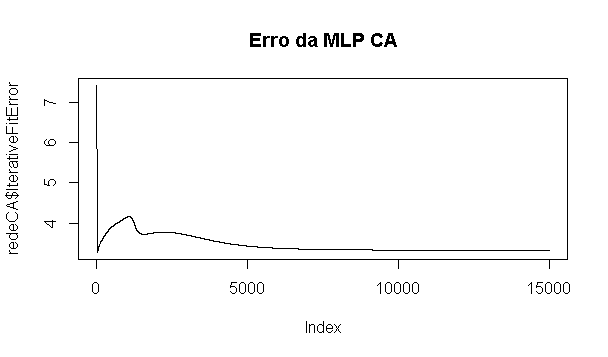
Figura 13: Gráfico em linhas representando a temperatura 2015 e 2016;



Analisando o gráfico da figura 13, pode-se notar um erro maior antes e depois do 100 e bem no final do gráfico que foi onde a rede não previu muito bem.

Gerando um gráfico de convergência do erro da mlp:

Figura 14: Gráfico em linha representando a convergência do erro da MLP;



Aparentemente a rede só gastaria um total de 10000 épocas para treinar a rede com uma taxa de erro considerável.

Uma grande dificuldade no treinamento foi o fato de separar os dados do dataset, eles haviam valores em nulos como eram muitos dados foram difíceis de visualizar entre outros.

Concluindo, a rede teve uma taxa de erro boa, não passando de 2,5. Abaixo alguns erros gerados de 2 execuções:

"ERRO da Previsão de 2016 (C°): 2.35225477300939"

"ERRO da Previsão de 2016 (C°): 2.35058493076434"

**Bibliografia**

JR, Oswaldo; MONTGOMERY, Eduard; ***Redes Neurais***: Fundamentos e Aplicações com Programas em C. Rio de Janeiro: Editora Ciência Moderna Ltda. 2007.