

# Particularidades de Métodos Lineares e Não-Lineares de Classificação

Daniel Lemos Simões  
398985  
daniellem@alu.ufc.br

Gustavo Filipe do Nascimento  
402889  
gusnas@alu.ufc.br

Pedro Cercelino Matos  
399325  
pedrocercelino@alu.ufc.br

Clailton Almeida Lopes  
400091  
clailtonx2@gmail.com

Lucas Martins de Oliveira  
398900  
eng.lucas@alu.ufc.br

**Resumo**—Neste artigo, foi estudado os métodos lineares e não-lineares de classificação. O primeiro método utilizado foi o de regressão logística, e em seguida, o de análise de discriminante linear, em ambos foram gerados as matrizes de confusão, finalizando os métodos lineares e iniciando os métodos não-lineares. Primeiramente, utilizou-se a análise discriminante quadrática, seguida pelo método dos K vizinhos mais próximos e por fim uma rede neural. Os resultados obtidos foram satisfatórios, os métodos lineares performaram melhor do que os não-lineares, devido principalmente ao tamanho do conjunto de dados.

**Index Terms**—classification models, grant application

## I. INTRODUÇÃO

Com diversas recessões ao redor do mundo, a previsão de dados se torna relevante, principalmente quando se trata de concessões pública, o aproveitamento das verbas deve ser feito da melhor maneira possível.

Algumas empresas privadas fazem uso de verbas públicas provenientes de concessões públicas, porém essas concessões são sempre um mistério, nunca se sabe o quanto irá receber, e se irá receber, por isso, as previsões são extremamente bem vindas, pois assim é possível que todo o planejamento financeiro seja feito com maior precisão, evitando futuros problemas.

Primeiramente, fora feito o aprendizado via modelos lineares sendo esses logistic classification e linear discriminant analysis, logo em seguida, utilizou-se de modelos não lineares: quadratic discriminant analysis, neural network e k-nearest neighbors.

Após a utilização dos métodos supracitados, foi feita a comparação dos resultados entre os dados provenientes de modelos lineares e os de modelos não-lineares.

## II. MÉTODOS

### A. Conjunto de dados

O conjunto de dados abordado neste artigo trata do problema percebido pela Universidade de Melbourne, em que jovens de todo o mundo perdem um tempo valioso em que

poderiam estar fazendo outras atividades, realizando pedidos de bolsas de pesquisa, que posteriormente acabam tendo seus pedidos rejeitados.

Com isto, a universidade disponibilizou um conjunto de dados contendo recursos que incluem variáveis que representam o tamanho da bolsa, a área geral de estudo e informações não identificadas sobre os pesquisadores que estão solicitando a bolsa.

Os dados estão divididos em um conjunto de treino formado por 1882 variáveis preditoras e 8190 solicitações de subsídio enquanto que o conjunto de teste é formado pelos mesmos 1882 preditores e 518 amostras, além disso, foi fornecido um conjunto reduzido de recursos, devido a alta correlação entre os preditores do conjunto inicial.

### B. Métodos Lineares de Classificação

No estudo dos modelos de classificação lineares, utilizamos dois métodos, são eles: Regressão Logística e a Análise do Discriminante Linear.

#### • Regressão Logística

A regressão logística é uma técnica de análise de regressão semelhante a regressão linear, porém voltada para problemas de classificação. Este modelo tem por objetivo prever o resultado de variáveis categóricas, no caso do conjunto de dados abordado estamos interessados em prever o resultado do preditor "Class", que pode assumir os valores *successfull* ou *unsuccessfull*.

Embora utilize uma regressão linear para prever os dados da saída, a regressão logística aceita apenas entradas dicotômicas (ou seja, valores binários) e é obtida através da função de ligação probabilística, a sigmóide, que possui este nome pois tem um formato semelhante a um "S". A função sigmóide é calculada segundo a equação abaixo:

$$p = \frac{1}{1 + \exp[-(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)]} \quad (1)$$

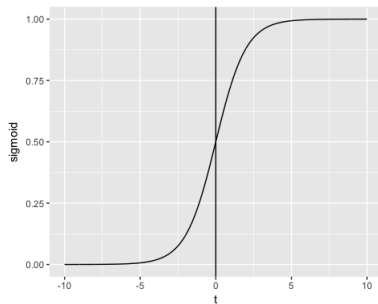


Figura 1: Exemplo de função sigmoide

A figura 1 mostra um exemplo de função sigmoide, este tipo de visualização só é possível quando se tem apenas um parâmetro para representar, que não é o caso do conjunto de dados abordado neste artigo, mas é importante perceber como ao utilizar esta função, pode-se aplicar um limiar para se obter saídas binárias.

Após realizar a regressão logística no conjunto de treino, levando em consideração apenas os preditores do conjunto reduzido, obtemos o modelo logístico, e então podemos construir o modelo preditivo de regressão linear no conjunto de teste. Ao aplicar o limiar de 0.5 para binarizar a saída em *successful* ou *unsuccessful*, temos os valores preditos e os reais para a classe preditor "Class". O desempenho da predição, pode ser avaliada através da matriz de confusão a seguir:

	successful	unsuccessful
successful	148	43
unsuccessful	41	286

Tabela I: Matriz de confusão obtida

A matriz de confusão fornece os dados sobre o desempenho do modelo, em sua diagonal obtemos os acertos da predição e na diagonal oposta os erros. A fração geral dos acertos é então calculada como a razão da soma da diagonal principal da matriz de confusão pela soma de todas as amostras. Para o modelo de regressão logística, a fração geral de acertos foi de 0.8478.

- **Análise do Discriminante Linear**

A análise do discriminante linear (LDA) é uma técnica de redução de dimensionalidade semelhante ao PCA (*Principal Component Analysis*), com a diferença que o método LDA é de aprendizagem supervisionada, diferentemente da técnica do PCA em que a saída não é utilizada para ajustar o modelo.

A LDA surgiu como um método linear muito útil para lidar com preditores com mais de duas variáveis

de classe para se prever, corrigindo a imprecisão que uma regressão logística pode obter neste caso, além disso, visa reduzir a dimensão dos dados, mantendo uma grande variabilidade nos dados transformados.

O LDA faz algumas suposições simplificadoras sobre seus dados:

- 1) Que seus dados são gaussianos, ou seja, que cada variável é modelada como uma curva gaussiana quando plotada.
- 2) Que cada atributo tem a mesma variação, ou seja, que os valores de cada variável variam em torno da média pela mesma quantidade, em média.

Com essas premissas, o modelo LDA estima a média e a variação de seus dados para cada classe. O passo a passo para calcular a LDA é o seguinte:

- 1) Calcular a classe interna entre matrizes de dispersão de classe.
- 2) Calcular os autovetores e os autovalores correspondentes para as matrizes de dispersão.
- 3) Classificar os autovalores e selecionar os k-valores.
- 4) Criar uma nova matriz contendo autovetores que mapeiam para os autovalores de k.
- 5) Obter os componentes da LDA, retirando o produto escalar dos dados e da matriz de autovetores.

Após calcular o modelo LDA, estamos interessados em realizar previsões com o conjunto de teste. A LDA faz previsões estimando a probabilidade de um novo conjunto de entradas pertencer a cada classe, neste caso as classes *successful* ou *unsuccessful*.

O modelo utiliza o teorema de Bayes para estimar a probabilidade da classe de saída. O Teorema de Bayes é uma fórmula matemática usada para o cálculo da probabilidade de um evento dado que outro evento já ocorreu, o que é chamado de probabilidade condicional.

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \quad (2)$$

Da noção de probabilidade condicionada e em posse dos valores de  $\pi_k$ ,  $u_k$  e  $\sigma^2$  do conjunto de dados, podemos utilizar a função discriminante para a classe k:

$$\delta_k(x) = x \frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} + \log(\pi_k) \quad (3)$$

Ao realizar o método LDA, obtemos a predição para o conjunto de treino, o índice de acerto do modelo está exemplificado na tabela 2 através da matriz de confusão. Para o modelo de análise do discriminante linear, o desempenho da fração geral de acertos foi de 0.8494 do total das amostras testadas. Em ambos os métodos de

	successful	unsuccessful
successful	156	45
unsuccessful	33	284

Tabela II: Matriz de confusão obtida

classificação linear, as variáveis do conjunto de dados reduzido, foram os preditores escolhidos para ajustar o modelo.

### C. Métodos Não-Lineares de Classificação

Nesta análise foram utilizados três métodos não-lineares de classificação, sendo eles: Análise Discriminante Quadrática, *K Nearest Neighbors* (K Vizinhos Mais Próximos) e Redes Neurais.

#### • Análise Discriminante Quadrática

Trata-se de um variante da Análise do Discriminante Linear, pois semelhante ao citado, resulta da assunção de que as observações de cada classe são retiradas de uma Distribuição Gaussiana, e associando estimativas para parâmetros no Teorema de Bayes para a realização da predição. No entanto, na Análise Discriminante Quadrática é estimada uma diferente matriz de covariância para cada classe observada. Utilizando o Classificador de Bayes, é definida uma observação  $X = x$  para a classe, onde a função:

$$\delta_k(x) = -\frac{1}{2} \log |\Sigma_k| - \frac{1}{2} (x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k) + \log \pi_k$$

retorna o maior valor possível, sendo seu valor quadrático em  $x$ . A Análise Discriminante Quadrática vem a ser mais conveniente para amplos conjuntos de treino, onde a variância do classificador não é de grande importância. Caso contrário, a Análise Discriminante Linear viria a ser mais adequada.

#### • K Vizinhos Mais Próximos

O método classificador KNN (*K Nearest Neighbors*, ou K Vizinhos Mais Próximos) pode vir a ser vantajoso em relação aos métodos citados anteriormente, pois possui uma abordagem não paramétrica com relação aos dados, isto é, não há necessidade de suposições quanto à distribuição condicional de uma variável de saída, dada a variável de entrada.

A partir de uma amostra de teste  $x_0$  e um número inteiro positivo  $K$ , são identificados os  $K$  pontos mais próximos de  $x_0$  em um conjunto de dados de treino, representados por  $N_0$ . Em seguida, a probabilidade

condicional para a saída de uma determinada classe  $j$  é estimada através da fração de pontos em  $N_0$  cujo os valores de resposta são iguais a respectiva classe:

$$P(Y = j) | X = x_0 = \frac{1}{K} \sum_{i \in N_0} I(y_i = j)$$

$I(y_i = j)$  é uma variável indicadora que retorna o valor 1 se  $y_i \neq j$ , e 0 se  $y_i = j$ .

Por fim, o método KNN aplica a regra de Bayes e designa a classe para a amostra  $x_0$  com maior probabilidade.

Abaixo temos um exemplo de uma classificação KNN, com  $K = 3$ .

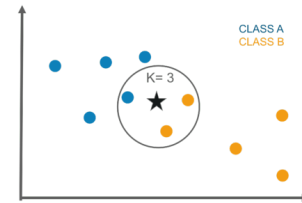


Figura 2: Exemplo de Classificação KNN, com  $K = 3$ .

Retirado de <https://www.edureka.co/blog/k-nearest-neighbors-algorithm/>, acesso em 21/11/19.

O valor ótimo de  $K$  irá variar de acordo com o conjunto de dados utilizado, no entanto, algoritmos de otimização como Algoritmos Genéticos, Evolução Diferencial e Enxame de Partículas poderiam ser utilizados para a escolha do valor de  $k$  que levaria a melhor performance. Há de se ter cuidado com esta etapa, pois a mesma poderia vir a reduzir desempenho de seu modelo.

#### • Redes Neurais

O princípio por trás deste método busca emular o funcionamento de um conjunto de neurônios no cérebro humano. Esse tipo de rede é elaborada para cumprir tarefas que são fáceis para humanos, mas difíceis para o computador, como classificar objetos de estudo dentro de categorias pré-definidas. Também pode ser usada para reconhecimento de padrões, como identificação facial ou biometria. Por ser ótima para detectar padrões, ela pode ser usada para identificar anomalias que fujam do esperado.

Para isso, ele dispõe de uma camada inicial de inputs para a entrada das observações trabalhadas, camadas internas de neurônios que se interconectam passando seus resultados uns aos outros e propagando a informação em questão, além de uma camada final de output que retorna as predições executadas.

A cada uma das entradas é associado um peso que, em conjunto com o bias do modelo, formam as entradas para cada neurônio. Para ajudar o modelo a definir uma saída dentro das características do set que estiver sendo trabalhado, bem como criar não linearidade e dar significado a esse output, existe a função de ativação, que pode ser de vários tipos dependendo da necessidade.

Este método é indicado para casos onde se tem um dataset não-linear, bem como aqueles com grande número de entradas. Como exemplo, muitas aplicações são usadas para trabalhar imagens, em que pixels correspondem a entradas. Outra vantagem está no fato de redes neurais poderem ser usadas para tarefas de classificação, bem como de regressão.

Infelizmente, essa técnica exige considerável poder de processamento dos dispositivos que forem executá-la, sendo assim algoritmos mais simples como árvore de decisão ou SVM, em alguns casos, são melhores alternativas. Outra desvantagem se encontra na limitação desse método de não poder cuidar de datasets que possuem valores faltando em sua constituição.

Abaixo, um exemplo de rede neural com quatro inputs, cinco neurônios internos e uma saída:

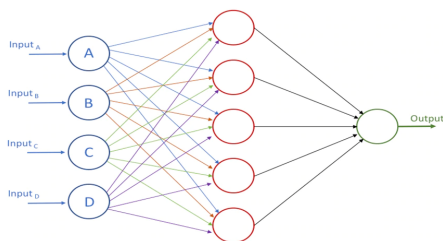


Figura 3: Exemplo de rede neural simples.

Retirado de <https://developer.oracle.com/databases/neural-network-machine-learning.html>, acesso em 21/11/19.

### III. RESULTADOS

A tabela abaixo, contém os resultados de acerto obtidos nos diferentes métodos de classificação abordados neste artigo:

Método Classificador	Acerto No Conjunto de Teste
Regressão Logística	84,78%
Análise do Discriminante Linear	84,94%
Análise Discriminante Quadrática	74,51%
K Vizinhos Mais Próximos (K = 6)	71,42%
Rede Neural	80,31%

É notado uma melhor performance nos métodos de classificação lineares referente aos não lineares.

### IV. CONCLUSÃO

Após análise dos métodos lineares e não-lineares, conclui-se que os métodos não-lineares não melhoraram o desempenho dos modelos de classificação. Pois os métodos lineares se saíram melhor em situações como por exemplo na análise discriminante quadrática, por que o conjunto de dados tem de ser muito grande para que o método seja efetivo, ao contrário da análise do discriminante linear que funciona de maneira adequada com conjuntos menores. O mesmo acontece com a rede neural em comparação com a regressão logística.

### REFERÊNCIAS

- [1] Kuhn, M., Johnson, K. (2013). *Applied Predictive Modeling* New York: Springer.
- [2] Michela Mulas. (2019). *Classification models*. Slides de Aula.
- [3] Linear Discriminant Analysis for Machine Learning (06/04/2016) Disponível em: *Linear Discriminant Analysis for Machine Learning* Acesso em: (23/11/2019)
- [4] K vizinhos mais próximos - KNN (17/03/2017) Disponível em: *K vizinhos mais próximos - KNN* Acesso em: (24/11/2019)