

RELATÓRIO NAC 1

FIAP - INTELIGÊNCIA ARTIFICIAL E COMPUTACIONAL

Alexandre Russi Junior - RM: 78727
alexandrerrussijunior@gmail.com

April 11, 2020

1 Introdução

Serão apresentados dois problemas neste relatório, tais problemas serão denominados e numerados por *Água dos Alpes* (problema 1) e *Livros x Aulas x Notas* (problema 2). A água tem os pontos de ebulição diferentes em pressões barométricas diferentes, essa descrição caracteriza o problema 1. O problema 2 possui 40 observações e o objetivo é analisar as notas de cada aluno baseado na quantidade de livros lidos e na frequência nas aulas.

O objetivo deste relatório é analisar, com base em dados observados e algoritmos de Machine Learning, os resultados e previsões obtidas para cada problema apresentado.

Nesse contexto, na seção 2 revisaremos de maneira simples a teoria sobre Machine Learning e Regressão Linear por mínimos quadrados, na seção 3 apresentaremos experimentos sobre os dois problemas apresentados a partir de programação em linguagem Python e os resultados obtidos a partir do algoritmo desenvolvido, e a conclusão será feita na seção 4.

2 Revisão Teórica

Existem diversas aplicações de Machine Learning, uma das mais significantes é o uso em *data mining*. Machine Learning pode se dividir entre aprendizado não supervisionado, aprendizado supervisionado, e aprendizado por reforço [5]. Neste relatório, será utilizado o aprendizado supervisionado através do uso

de algoritmos de regressão linear.

O modelo regressão linear é uma poderosa ferramenta em análise de dados. A análise de regressão de mínimos quadrados ordinários (MQO) é o modelo estatístico mais usualmente empregado na ciência política contemporânea. Com a regressão é possível estimar o grau de associação entre Y , variável dependente e X_i , conjunto de variáveis independentes (explicativas). O objetivo é resumir a correlação entre X_i e Y em termos da direção (positiva ou negativa) e magnitude (fraca ou forte) dessa associação [1]. A notação para este modelo pode se dar por:

$$Y = \alpha + \beta_1 X_1 + \varepsilon \quad (1)$$

Y representa a variável dependente, ou seja, aquilo que queremos explicar/entender/predizer. X_1 , por sua vez, representa a variável independente, aquilo que o pesquisador acredita que pode ajudar a explicar/entender/predizer a variação de Y . O intercepto (α), também chamado de constante, representa o valor de Y quando X_1 assume valor zero [1].

3 Experimentos

Os algoritmos que serão desenvolvidos e comentados a seguir são utilizados para ler e analisar os dados observados de cada problema. Com o auxílio de bibliotecas abertas do Python faremos a regressão linear pelo método de mínimos quadrados.

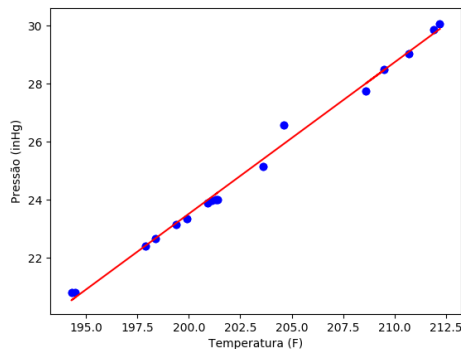


Figure 1: Ponto de ebulição da água em diferentes pressões barométricas. Pontos azuis são as observações realizadas, e a reta vermelha é a regressão linear calculada por mínimos quadrados.

3.1 Água dos Alpes

A água tem seu ponto de ebulição alterado de acordo com a pressão barométrica na qual está sujeita. Neste problema temos 17 dados de observações. Nestes dados temos como variáveis a temperatura de ebulição em graus Fahrenheit ($^{\circ}\text{F}$) e a pressão barométrica em polegadas de mercúrio (inHg). Com auxílio da biblioteca *sklearn* é possível calcular a regressão linear a partir do método de múltiplas variáveis, com base na Equação 1. Sendo assim, foi desenvolvido um algoritmo que lê os dados de temperatura e pressão, manipula os mesmos e calcula a reta da regressão linear (Figura 1).

Para testar algumas previsões no algoritmo desenvolvido, podemos analisar qual o valor de pressão para uma temperatura de 200°F . Testando isso no código, temos como resultado uma pressão barométrica de 23,5 inHg [4].

3.2 Livros x Aulas x Notas

No Problema 2, possuímos um total de 40 observações da relação entre número de livros lidos pelos alunos em um curso de estatística, a frequência nas aulas que representa o número de aulas que o aluno participou, e a nota que representa a nota final do curso.

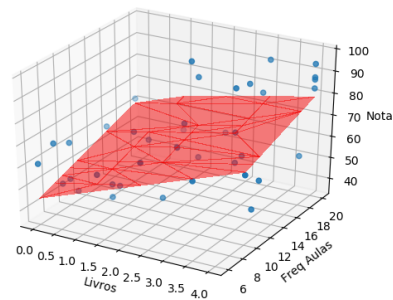


Figure 2: Número de livros lidos, aulas assistidas e notas finais de alunos. Pontos azuis são as observações realizadas, e a reta vermelha é a regressão linear calculada por mínimos quadrados.

Neste caso, temos um problema com multi-variáveis, portanto o resultado final após a regressão linear a partir dos dados não será uma reta em um eixo bidimensional X e Y, e sim um plano em uma representação tridimensional X, Y e Z (Figura 2).

A partir do algoritmo desenvolvido é possível realizar previsões das notas finais de cada aluno inserindo o número de livros lidos e aulas assistidas por ele. Com isso, podemos realizar uma previsão para um aluno que leu 4 livros e participou de 20 aulas. A nota final para tal previsão será de 79,2 [3].

4 Conclusão

O modelo de regressão linear é bem popular, a principal razão para tal popularidade tem sido a sua capacidade de prever e explicar variáveis métricas [2]. Neste relatório foi possível reconhecer essa capacidade de prever e explicar tais variáveis. Por esse motivo, os modelos estatísticos que utilizam a regressão linear por mínimos quadrados são importantes para o estudo de Machine Learning e Inteligência Artificial.

References

- [1] Dalson Figueiredo Filho, Felipe Nunes, Enivaldo Carvalho da Rocha, Manoel Leonardo Santos, Mariana Batista, and José Alexandre Silva Júnior. O que fazer e o que não fazer com a regressão: pressupostos e aplicações do modelo linear de mínimos quadrados ordinários (mqo). *Revista Política Hoje*, 20(1), 2011.
- [2] Joseph F Hair, William C Black, Barry J Babin, Rolph E Anderson, Ronald L Tatham, et al. *Multivariate data analysis*, volume 5. Prentice hall Upper Saddle River, NJ, 1998.
- [3] Alexandre Russi Junior. <https://bit.ly/2JZ7AbG>. GitHub Problema 2, 2020.
- [4] Alexandre Russi Junior. <https://bit.ly/3ee7xXo>. GitHub Problema 1, 2020.
- [5] Sotiris B Kotsiantis, Ioannis D Zaharakis, and Panayiotis E Pintelas. Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review*, 26(3):159–190, 2006.