Inteligência
Artificial e Big
Data
Aula 05

Prof. Me Daniel Vieira

SENAI

Agenda

1- Regressão Linear
2- Correlação
3- Método dos quadrados mínimos
4 - Etapas Machine Learning
5 - Estudo de caso

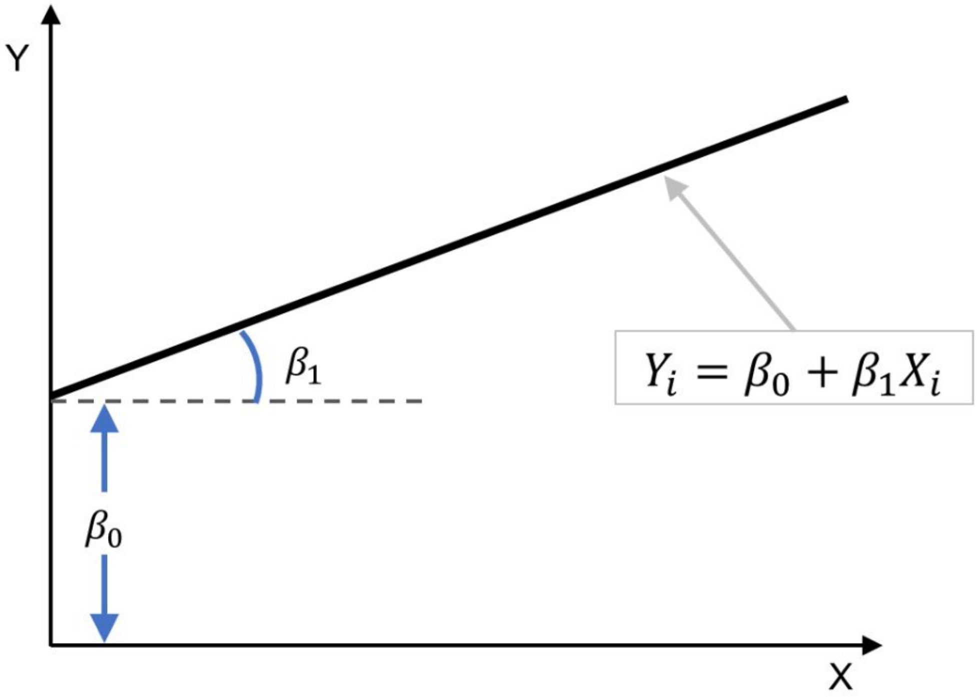
Equação Regressão Linear

$$Y = \beta_0 + \beta_1 X + e$$

Variável reposta	Intecepto	Coeficiente angular	Variável explicativa	Erro

B0 - é o valor de Y quando Xi é zero, já o B1 é o coeficiente angular e nós informa a taxa de variação e quão inclinada a reta está

Equação Regressão Linear

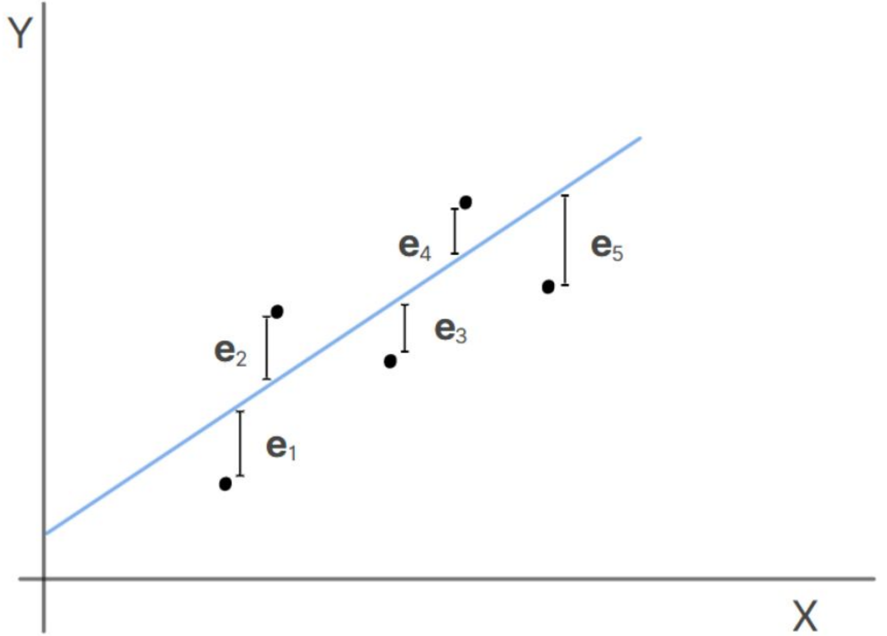


Método dos quadrados mínimos

$$\hat{\beta}_1 = \frac{n\sum XiYi - \sum Xi \sum Yi}{n\sum Xi^2 - (\sum Xi)^2} \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

Método dos quadrados mínimos

Gasto = 207.9 + 0.3Renda + erro



Método dos quadrados mínimos

A Regressão é uma técnica de análise de dados que prevê o valor de dados desconhecidos a partir de dados conhecidos

Equação da reta : $yi = \theta 1x1 + \theta 0 + Ei$
Erro quadrático médio = $1/N\ \Sigma(yi - yi)\hat{}2$

Método dos quadrados mínimos

Agora podemos ver a regressão como um problema de otimização. Devemos minimizar esta função de soma dos erros quadráticos. Isto é um problema conhecido na algebra linear como minímos quadrados e sua resolução dá as equações:

$$\theta_1 = correlacao(x, y) \times \frac{\sigma(y)}{\sigma(x)}$$

$$\theta_0 = \bar{y} - \alpha\bar{x}$$

Sendo $\bar{x}$ e $\bar{y}$ as médias dos valores de x e y, correlação(x, y) é a correlação entre os valores de x e y, e $\sigma(x)$, $\sigma(y)$ os desvios padrões de x e y.

Assim, vamos definir uma função para minimizar os valores de θ_1 e θ_0.

Correlação

Correlação Simples – quando se estuda o grau de relação entre duas variáveis, sendo uma dependente (Yi) e outra independente (Xi).
Correlação Múltipla – quando se estuda o grau de relação simultânea entre a variável dependente e duas ou mais variáveis independentes.
Correlação Parcial – no caso de uma relação múltipla, quando se estuda a relação pura entre duas variáveis, depois de eliminada estatisticamente a influência de outras variáveis independentes

Correlação

Coeficiente de Correlação (r) é a medida estatística que dimensiona o grau de relação entre duas ou mais variáveis.

$$r = \frac{\sum x_i y_i}{(n-1) * s_x . s_y}$$

Sendo: $x_i \Rightarrow$ desvios reduzidos da variável independente($x_i = X_i - \overline{X}$)

$y_i \Rightarrow$ desvios reduzidos da variável dependente ($y_i = Y_i - \overline{Y}$).

$n \Rightarrow$ número de valores observados.

s_x e $s_y \Rightarrow$ desvio padrão das respectivas variáveis.

Desvio padrão

Indica medida de dispersão dos dados

Desvio padrão populacional

$$DP = \sqrt{\frac{\sum_{i=1}^{n}(x_i - M_A)^2}{n}}$$

Desvio padrão amostral

$$s^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n - 1}$$

Correlação

Equação de Regressão da variável dependente (função linear).

$$y_i = \frac{r\, s_y \, . \, x_i}{s_{xi}} \quad \Rightarrow \quad Y_i - \overline{Y} = \frac{r\, s_y}{s_{xi}}(X_i - \overline{X})$$

Correlação

Equação de Regressão da variável dependente (função linear).

$$y_i = \frac{r\, s_y \,.\, x_i}{s_{xi}} \quad \Rightarrow \quad Y_i - \overline{Y} = \frac{r\, s_y}{s_{xi}}(X_i - \overline{X})$$

Exemplo: calcular o coeficiente de correlação entre altura (X_i) e peso (Y_i) de uma amostra de 10 estudantes universitários. Estimar o peso de um estudante com 196 cm de altura.

X_i	Y_2	X_i	y_i	$x_i^* y_i$	x_i^2	y_i^2	$\acute{Y}_i$
173	70	1	0,6	0,6	1	0,36	70,3
169	66	-3	-3,4	10,2	9	11,56	66,9
172	70	0	0,6	0,0	0	0,36	69,4
174	68	2	-1,4	- 2,8	4	1,96	71,1
165	64	- 7	-5,4	37,8	49	29,16	63,5
170	68	- 2	-1,4	2,8	4	1,96	67,7
171	72	- 1	2,6	- 2,6	1	6,76	68,6
168	65	- 4	-4,4	17,6	16	19,36	66,0
178	72	6	2,6	15,6	36	6,76	74,5
180	79	8	9,6	76,8	64	92,16	76,2
1720	694	0	0	156	184	170,40	694,0

Correlação

$$s_x^2 = \frac{\Sigma x_i^2}{n-1} = \frac{184}{9} = 20,4444 \quad \Rightarrow \quad s_x = 4,52$$

$$s_y^2 = \frac{\Sigma y_i^2}{n-1} = \frac{170,40}{9} = 18,9333 \quad \Rightarrow \quad s_y = 4,35$$

Coeficientes de Correlação

$$r = \frac{\Sigma\, x_i\, y_i}{(n\text{-}1).s_x s_y} = \frac{156}{9 \times 4,52 \times 4,35} = 0,88 \quad \Rightarrow \quad \text{Ótima correlação positiva.}$$

Correlação

Equação Regressão

$$Yi - \bar{Y} = r\frac{Sy}{Sx}(Xi - \overline{X})$$

$$Yi - 69{,}4 = \frac{0{,}88 * 4{,}35}{4{,}52}(Xi - \overline{172})$$

Yi=0,85* Xi-146,2 (+ 69,4)

Correlação

Estimar o peso de uma pessoa com 1,96 m = 196 cm

$Yi=0,85* Xi- 76,8$

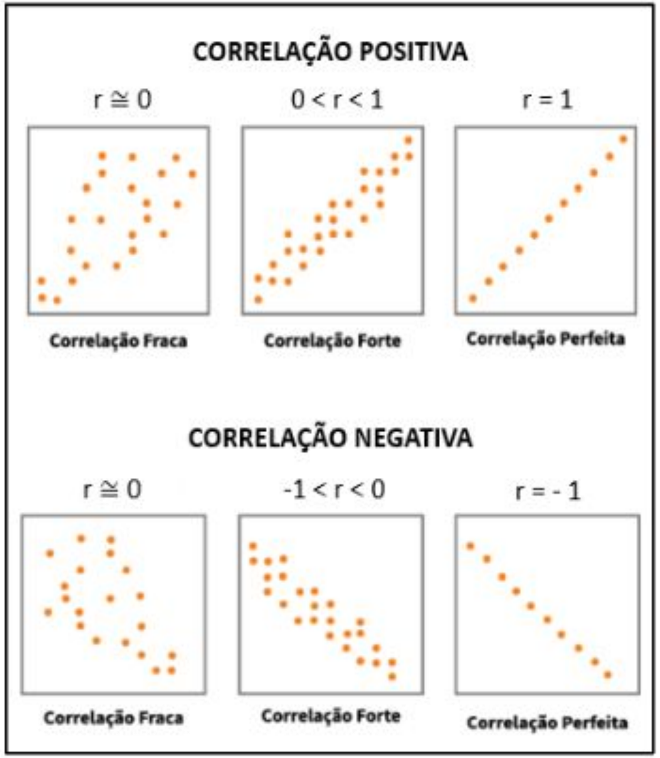
Substituindo 1,96 na equação

$Yi = 0,85 * 196 - 76,8$

$Yi = 166,6 - 76,8$

$Yi = 89,8$

Correlação



O tipo de correlação pode ser visualizado a partir do **Coeficiente de Correlação** (ou Coeficiente de Pearson - r). Os valores obtidos neste coeficiente relacionam-se da seguinte forma:

r = 0: Correlação nula ou inexistente entre variáveis.
r = 1: Correlação positiva entre variáveis.
r = -1: Correlação Negativa entre variáveis.

Correlação Positiva:
No qual as duas variáveis crescem no mesmo sentido. Ou seja, enquanto um aumenta, o outro também aumenta.

Correlação negativa:
No qual as duas variáveis variam em sentidos contrários. Ou seja, enquanto um aumenta, o outro diminui

Correlação Nula:
Não há interação entre variáveis.

Exercícios

1) Plotar o gráfico dos seguintes dados e calcular sua correlação:
Foram obtidos no Departamento de Nutrição, em certa Empresa, os seguintes dados sobre conversão alimentar em bovinos: Verificar se há correlação entre as variáveis, calculando o coeficiente de correlação e estimar a Conversão Alimentar para bovinos com idade de 8 anos.

Idade	1	2	3	4	5	6	7
C. Alimentar	5,6	5,2	4,8	4,5	4,4	2,9	2,7

Exercícios

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
idade = np.array([1,2,3,4,5,6,7])
C = np.array([5.6,5.2,4.8,4.5,4.4,2.9,2.7])
```

Exercícios

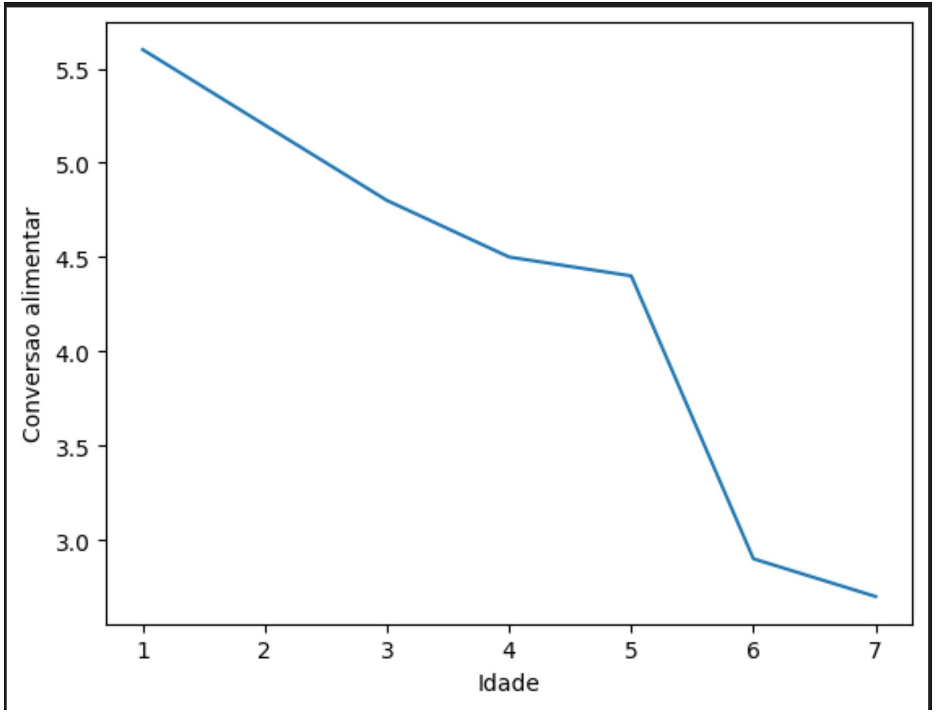
```
ex_1 = pd.DataFrame({
 "idade":  idade,"Conversao alimentar":C
})

print(ex_1.corr())
```

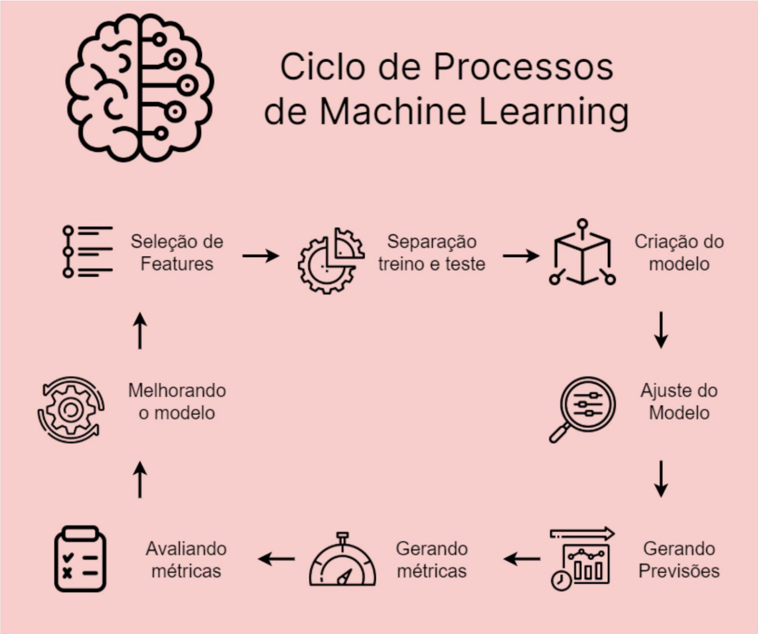
Exercícios

```
plt.plot(ex_1['idade'],ex_1 ['Conversao alimentar'])
plt.xlabel('Idade')
plt.ylabel('Conversao alimentar')
```

Exercícios



Etapas de Machine Learning



Estudo de caso

Você foi contratado por uma empresa para criar uma modelo de aprendizado de máquina para prever o preço dos imóveis com base na quantidade de andares, cômodos.

Obrigado!

Prof. Me Daniel Vieira

Email: danielvieira2006@gmail.com

Linkedin: Daniel Vieira

Instagram: Prof daniel.vieira95