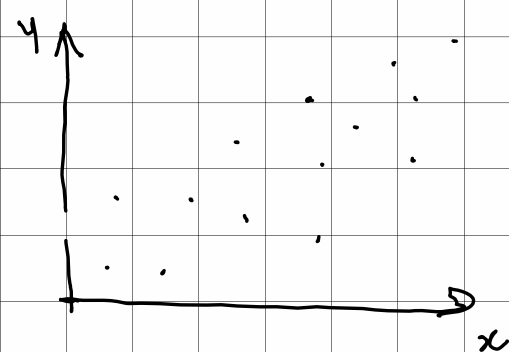
Curso insper

Regressão Linear

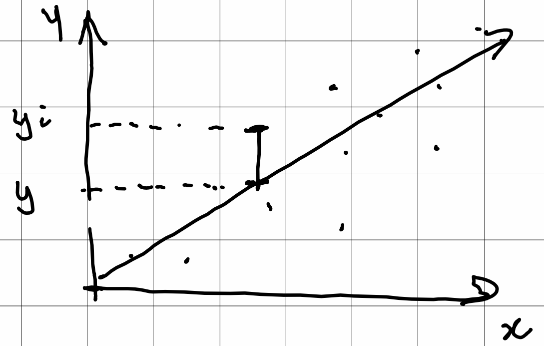
Considere o seguinte conjunto de dados:

- $x_1, x_2, \ldots, x_n$
- $y_1, y_2, \ldots, y_n$

Considere que esses dados se distribuem em um gráfico de dispersão da seguinte maneira:



- Queremos saber se os valores de x predizem de alguma maneira y.



- Suponhamos que a relação entre x e $\underline{y}$ é dada pela seguinte expressão:

$$y = a.x + b$$

- Como não é possível escolher a e b de modo que o comportamento seja reproduzido com exatidão, delimitaremos como critério encontrar (a, b) tal que

$$\sum_{i=0}^{N-1} | y_i - \underbrace{(a.x_i + b)}_{y} |$$

seja o menor possível.

Assim,

$$= \sum_{i=0}^{N-1} \{ y_i - (a x_i + b) \}^2$$

$$= \sum_{i=0}^{N-1} \{ y_i^2 - 2 y_i . (a x_i + b) + (a x_i + b)^2 \}$$

$$= \sum_{i=0}^{N-1} \Big(y_i^2 - 2 y_i a x_i - 2 y_i b$$

$$+ a^2 x_i^2 + 2 a x_i b + b^2 \Big)$$

$$= \sum_{i=0}^{N-1} \{ a^2 x_i^2 + a . (2 x_i b - 2 y_i x_i) + b^2 - 2 y_i b \}$$

$$= a^2 \sum_{i=0}^{N-1} x_i^2 + a \sum_{i=0}^{N-1} (2 x_i b - 2 y_i x_i)$$

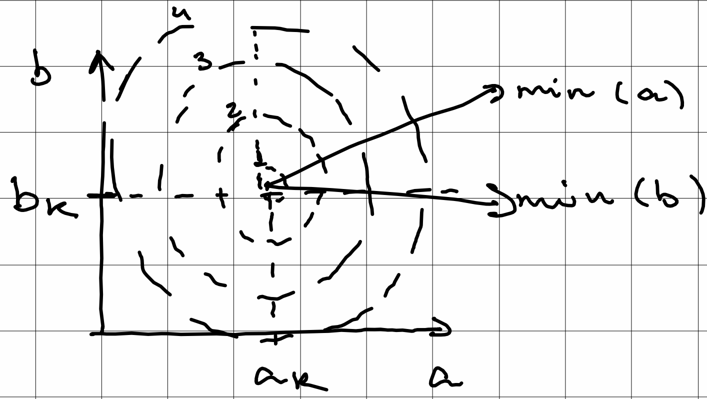
$$2b \sum_{i=0}^{N-1} y_i + N b^2$$

Da mesma forma:

$$= \sum_{i=0}^{N-1} \left(y_i^2 - \overline{2 y_i a x_i} - \overline{2 y_i b} + \overline{a^2 x_i^2} + \overline{2 a x_i b} + b^2 \right)$$

$$= \sum_{i=0}^{N-1} \left\{ b^2 + b\,(2 a x_i - 2 y_i) + a \cdot (x_i^2 - 2 y_i x_i) + y_i^2 \right\}$$

Portanto, para ambos os parâmetros, a soma dos quadrados constitui uma equação de 2º grau com mínimo admissível (já que o sinal do termo de maior potência em ambos os casos é positivo).



Curvas de nível: quanto menor o nível, menor os valores.
Chamemos nossa soma de quadrados de função S.
Se fixarmos um b, chamando-o de b_K, temos uma função de segundo grau, o mesmo com a_K (a fixado).

Sabemos que o ponto onde a derivada de uma função quadrática é zero é um mínimo ou um máximo e que ele é único. Portanto:

$$\frac{\partial S}{\partial a} \cdot \hat{i} + \frac{\partial S}{\partial b} \cdot \hat{j} = 0 \Rightarrow \nabla S = 0$$

ou seja, a solução do nosso problema é onde o gradiente de nossa função S é ZERO.

Portanto, encontremos $\frac{\partial S}{\partial a}$ e $\frac{\partial S}{\partial b}$:

$$S = a^2 \sum_{i=0}^{N-1} x_i^2 + a \sum_{i=0}^{N-1} (2 x_i b - 2 y_i x_i)$$

$$2b \sum_{i=0}^{N-1} y_i + N b^2$$

$$\frac{\partial S}{\partial a} = 2a \sum_{i=0}^{N-1} x_i^2 + \sum_{i=0}^{N-1} (2 x_i b - 2 y_i x_i)$$

$$S = \sum_{i=0}^{N-1} \left(b^2 + b(2ax_i - 2y_i) + a \cdot (x_i^2 - 2y_i x_i) + y_i^2 \right)$$

$$\frac{\partial S}{\partial b} = 2bN + \sum_{i=0}^{N-1} (2ax_i - 2y_i)$$

Botemos nossas duas derivadas parciais juntas :

$$\frac{\partial S}{\partial a} = 2a \sum_{i=0}^{N-1} x_i^2 + \sum_{i=0}^{N-1} (2x_i b - 2y_i x_i)$$

$$\frac{\partial S}{\partial b} = 2bN + \sum_{i=0}^{N-1} (2a x_i - 2y_i)$$

$$\frac{\partial S}{\partial a} = 2a \underbrace{\sum_{i=0}^{N-1} x_i^2}_{\gamma} + 2b \underbrace{\sum_{i=0}^{N-1} x_i}_{\delta} - 2 \underbrace{\sum_{i=0}^{N-1} y_i x_i}_{\eta}$$

$$\frac{\partial S}{\partial b} = 2bN + 2a \underbrace{\sum_{i=0}^{N-1} x_i}_{\delta} - 2 \underbrace{\sum_{i=0}^{N-1} y_i}_{\varphi}$$

$$\frac{\partial S}{\partial a} = 2a \underbrace{\sum_{i=0}^{N-1} x_i^2}_{\gamma} + 2b \underbrace{\sum_{i=0}^{N-1} x_i}_{\delta} - 2 \underbrace{\sum_{i=0}^{N-1} y_i x_i}_{\eta}$$

$$\frac{\partial S}{\partial b} = 2bN + 2a \underbrace{\sum_{i=0}^{N-1} x_i}_{\delta} - 2 \underbrace{\sum_{i=0}^{N-1} y_i}_{\varphi}$$

Portanto

$$\begin{cases} \cancel{2} a\gamma + \cancel{2} b\delta = \cancel{2}\eta \\ \cancel{2} bN + \cancel{2} a\delta = \cancel{2}\varphi \end{cases}$$

$$\begin{cases} a\gamma + b\delta = \eta \\ a\delta + bN = \varphi \end{cases}$$

é tudo que precisamos resolver para obtermos nossos parâmetros a e b.

Regressão Logística

Diferente do caso anterior, agora queremos predizer uma variável categórica.

Consideremos o caso mais simples:
dado um valor preditivo x, qual a classe y?

$y \in \{0, 1\}$

$x \in \mathbb{R}$

Nesse caso, podemos utilizar o que chamamos de <u>função logística</u> p/ modelar o valor de:

$$\mathbb{P}(Y=1) = \underbrace{\frac{1}{1+e^{-(ax+b)}}}_{\text{função logística}}$$

Chamemos ela de $g(x)$. Então,

$\lim\limits_{x \to \infty} g(x) = 1 \qquad \lim\limits_{x \to -\infty} g(x) = 0 \qquad$ se $a > 0$

$\lim\limits_{x \to \infty} g(x) = 0 \qquad \lim\limits_{x \to -\infty} g(x) = 1 \qquad$ se $a < 0$

Agora, perceba

$$\mathbb{P}(Y=1) = \frac{1}{1+e^{-(ax+b)}}$$

$$\frac{\mathbb{P}(Y=1)}{\mathbb{P}(Y=0)} = \frac{\mathbb{P}(Y=1)}{1-\mathbb{P}(Y=1)} = \frac{\left.\frac{1}{1+e^{-(ax+b)}}\right\} A1}{1 - \frac{1}{1+e^{-(ax+b)}}},$$

...

$$= \frac{A1}{\frac{\cancel{1}+e^{-(ax+b)} - \cancel{1}}{1+e^{-(ax+b)}}} = \frac{A1}{\frac{e^{-(ax+b)}}{1+e^{-(ax+b)}}}$$

$$= \frac{1}{\cancel{1 + e^{-(ax+b)}}} \cdot \frac{\cancel{1 + e^{-(ax+B)}}}{e^{-(ax+b)}}$$

$$\frac{\mathbb{P}(Y=1)}{\mathbb{P}(Y=0)} = \frac{1}{e^{-(ax+b)}} = e^{ax+b}$$

Agora, aplicando o logaritimo de base euler:

$$\log_e\left(\frac{\mathbb{P}(Y=1)}{\mathbb{P}(Y=0)}\right) = \log_e\left(e^{ax+b}\right)$$

$$= (ax+b) \cdot \log_e(e)$$

$$= ax + b$$

Assim, voltamos ao problema da regressão linear simples.

Então, descobrindo a e b, digamos que $ax + b = k$.

$$\log_e \left(\frac{\mathbb{P}(Y=1)}{1-\mathbb{P}(Y=1)} \right) = k$$

$$\frac{\mathbb{P}(Y=1)}{1-\mathbb{P}(Y=1)} = e^k$$

$$\mathbb{P}(Y=1) = e^k \cdot \{1-\mathbb{P}(Y=1)\}$$

$$\mathbb{P}(Y=1) = e^k - e^k \cdot \mathbb{P}(Y=1)$$

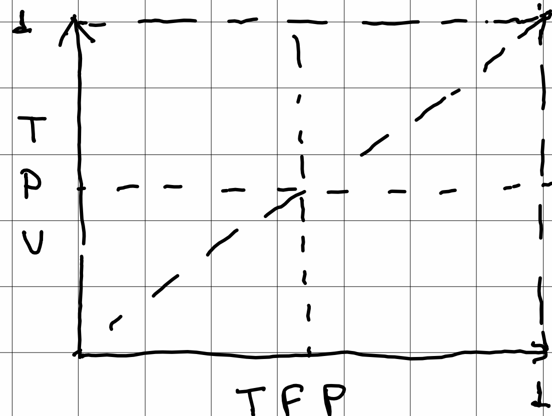
$$\mathbb{P}(Y=1) + e^k \mathbb{P}(Y=1) = e^k$$

$$\mathbb{P}(Y=1) \cdot (1+e^k) = e^k$$

$$\boxed{\mathbb{P}(Y=1) = \frac{e^k}{1+e^k}}$$

ROC: Receiver operating Characteristics

É uma forma de avaliar classificadores. Constrói-se um gráfico:



$$TFP = \frac{FP}{N}$$

$$TPV = \frac{VP}{P}$$

Classificadores ruins ficam na diagonal

TFP = taxa de falsos positivos
TPV = taxa de positivos verdadeiros

Bom classificador = canto superior esquerdo
Bom classificador se invertido = canto inferior direito.