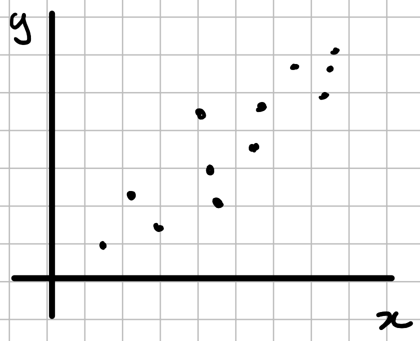
## Curso Insper - Regressão e Classificação.

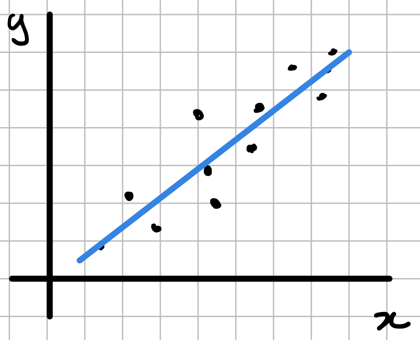
Considere o seguinte conjunto de dados

- $x_0, \dots, x_{N-1}$
- $y_0, \dots, y_{N-1}$

O gráfico de dispersão dos dados tem o seguinte formato:



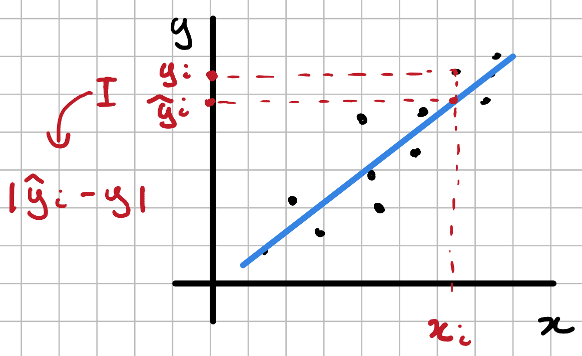

Queremos saber se os valores de $x$ predizem de "alguma forma" os valores de $y$.




Supomhamos que $x$ e $y$ apresentam a seguinte relação:

$$y = a.x + b$$

Como não é possível encontrar os parâmetros $a$ e $b$ de modo que o comportamento seja reproduzido com exatidão, definiremos um critério razoável:




$a$ e $b$ são escolhidos tal que:

$$\hat{a}, \hat{b} = \arg \min_{a, b} \sum_{i=0}^{N-1} \{ y_i - \hat{y}(x_i) \}^2$$

Onde:

$\hat{a}, \hat{b}$ = são os parâmetros que procuramos

$\hat{y}(x_i) = \hat{a}.x_i + \hat{b}$

$\arg\min(\ )$ = argumentos que minimizam a função à direita.

$\{\ \}$ : parentese de nível (notação própria)

Assim, analisemos a função do nosso critério:

$$\sum_{n=0}^{N-1} \{ y_i - \hat{y}(x_i) \}^2$$

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

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

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

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

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

Portanto, p/ b fixado, temos uma função de grau 2 de a.

Da mesma forma:

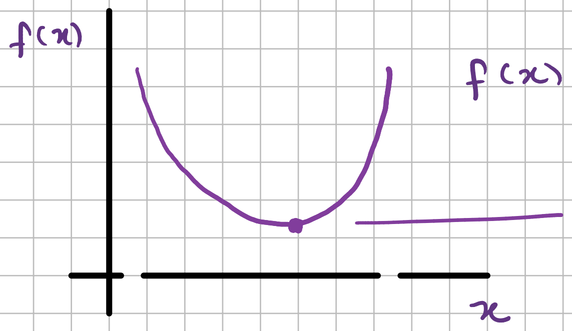
$$= \sum_{n=0}^{N-1} \{ y_i^2 - 2 y_i a x_i - 2.y_i b + a^2.x_i^2 + 2.a x_i.b + b^2 \}$$

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

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

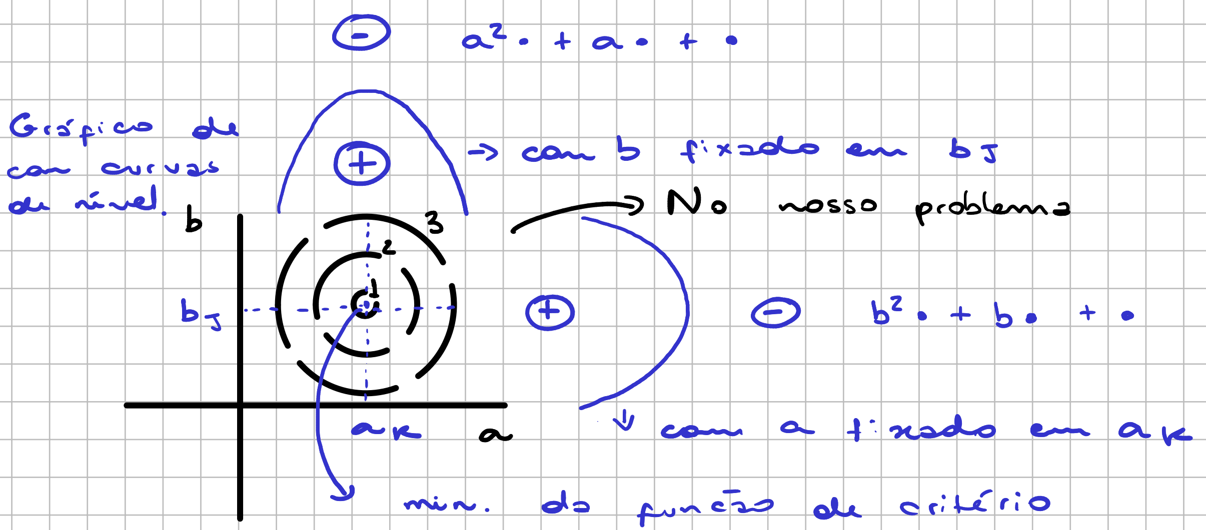
Portanto, p/ um a fixo, também temos uma função de grau 2 de b.

funções de grau 2, onde o termo de maior grau tem coeficiente positivo tem um ponto de derivada zero que corresponde ao mínimo da função:




$$f(x) = c_1 . x^2 + c_2 . x + c_3$$

$$\arg\min_{x} f(x) = x : \frac{df}{dx}(x) = 0$$




nível 1 → menores valores
nível 3 → maiores valores.

Chamemos nossa função critério de S, então:

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

Ou seja, procuramos o ponto onde o gradiente é igual à zero.

Retomemos a função S

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

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

$$\frac{\partial S}{\partial b} = 2a \underbrace{\sum_{n=0}^{N-1} x_i}_{\gamma} - 2 \underbrace{\sum_{n=0}^{N-1} y_i}_{\varphi} + 2Nb$$

Simplificando, acabamos com o sistema:

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

Agora basta resolvermos o sistema de equações.

# Regressão logística

Diferente do caso anterior, queremos predizer uma variável categórica.

Por exemplo...

Dado um valor $x$, qual a classe $y$?

$y = \{0, 1\}$

$x = \mathbb{R}$

Neste caso, podemos utilizar a função logística:

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

Perceba que:

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

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

Agora, note que:

$$\frac{\mathbb{P}(Y=1)}{\mathbb{P}(Y=0)} = \frac{\mathbb{P}(Y=1)}{1 - \mathbb{P}(Y=1)} = \frac{\left.\frac{1}{1 + e^{-(ax+b)}}\right\} A_1}{1 - \frac{1}{1 + e^{-(ax+b)}}}$$

$$= \frac{A_1}{\frac{\cancel{1} + e^{-(ax+b)} - \cancel{1}}{1 + e^{-(ax+b)}}}$$

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

Portanto,

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

Aplicando o logaritimo neperiano:

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

$$= (ax+b).\underbrace{\cancel{\log_e(e)}}_{1}$$

Assim voltamos p/ o problema da regressão.

Após descobrir $\hat{a}$ e $\hat{b}$, digamos que p/ um determinado $x_i$ temos que $\hat{a}x_i + \hat{b} = k$.

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

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

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

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

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

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

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

# Curvas ROC:

ROC: Receiver operating characteristics.

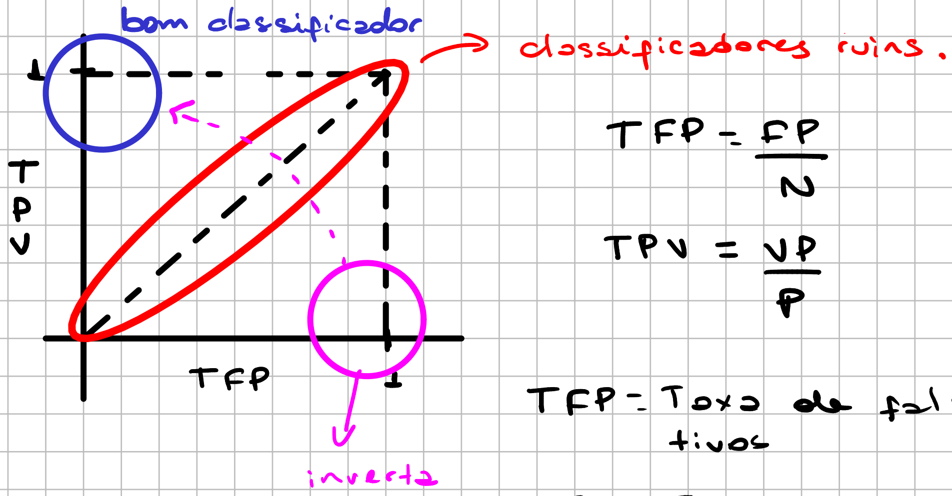
É uma maneira de avaliar classificadores. Constroi-se um gráfico:

VP = Verdadeiros positivos

FP = falso positivo

N = negativo

P = positivo




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

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

TFP = Taxa de falsos positivos.

TPV = Taxa de verdadeiros positivos.