

O que é melhor?

1 salário
95% certo

10k góias
51% certo

Qual o desempenho?

salário: 95% certo

góias: ?

$$\text{Prob}[X=10000] = p^{10000}$$

$$\text{Prob}[X=9999] = 10000 p^{9999} \cdot (1-p)$$

$$\text{Prob}[X=9998] = \binom{10000}{9998} p^{9998} \cdot (1-p)^2$$

$$\text{Prob}[X=k] = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

$$n = 10000$$

$$p = 0,51$$

X: n.º de acertos

$$X \sim \text{Bin}(n=10000, p=0,51)$$

Queremos saber:

$$\text{Prob}[X > 5000] = ?$$

$$= 1 - \underbrace{\text{Prob}[X \leq 5000]} = 97,675$$

cumulativa da Bin ($n=10000, p=0,51$)

Conclusão:

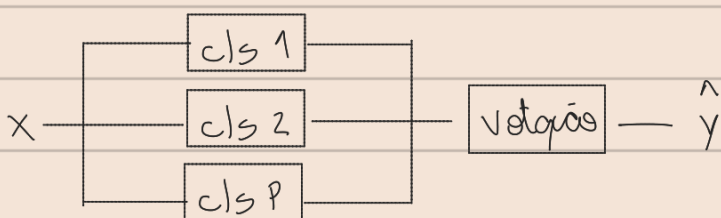
A democracia funciona!

DESDE QUE

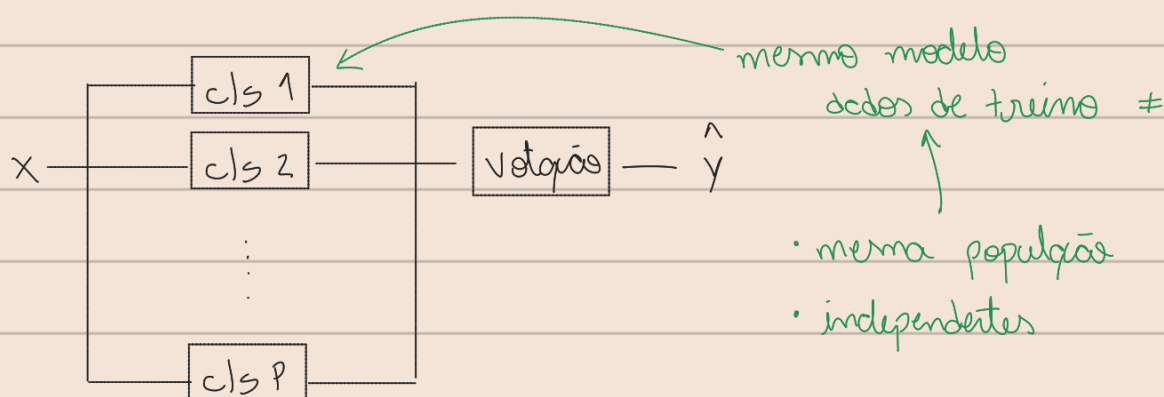
- Muitos indiv
- Indivíduos minimamente preparados
- Indivíduos independentes

Modulos de ensemble

Votoção



Outra alternativa: mesmo modelo, dados diferentes



Treinos:

cls 1 com $X_{train}^{(1)}, y_{train}^{(1)}$
cls 1 com $X_{train}^{(2)}, y_{train}^{(2)}$
⋮
cls 1 com $X_{train}^{(k)}, y_{train}^{(k)}$

obtidos independentemente da mesma população

Bootstrap (Efron 1979)

"To lift yourself by one's own bootstrap"

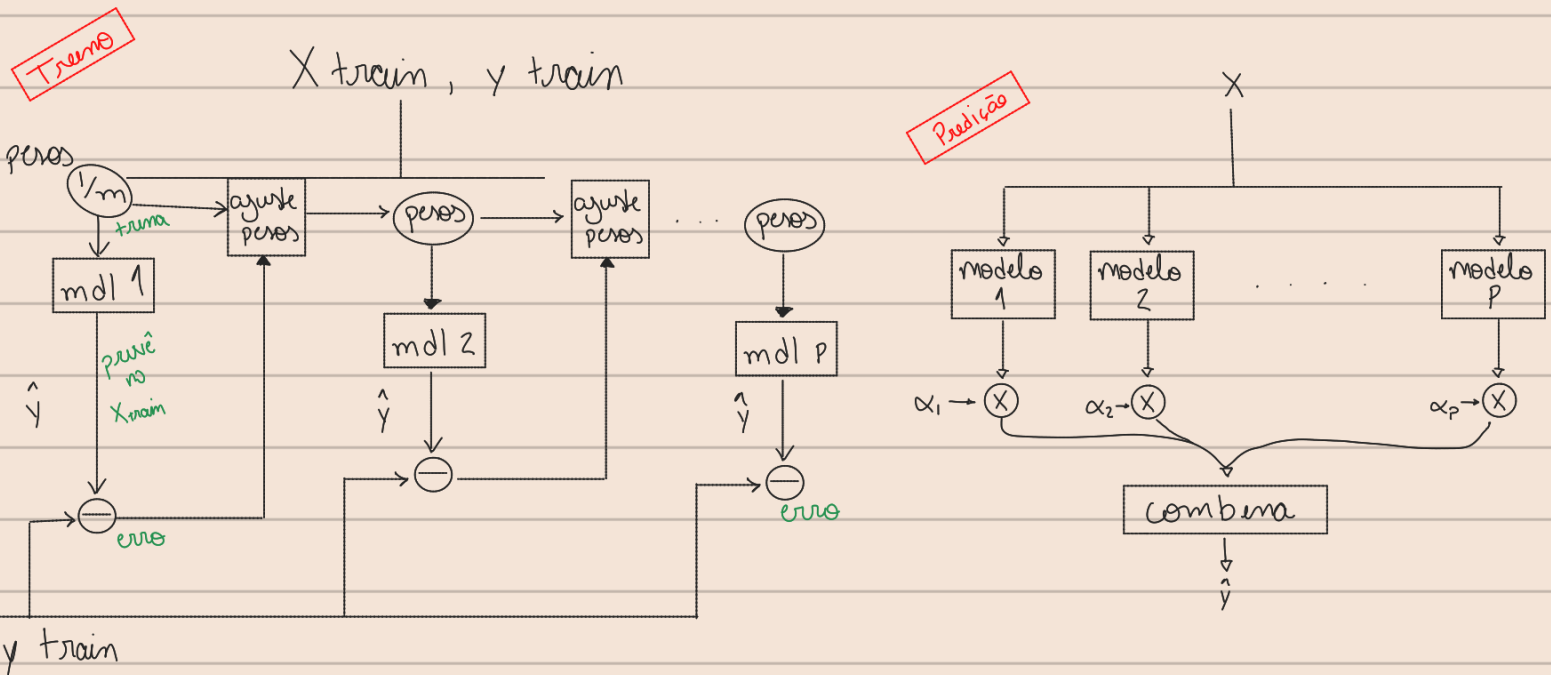
- amostrar COM REPOSIÇÃO m exemplos de (X_{train}, y_{train}) original (de n exemplos)
- repetir P vezes

Bagging → Bootstrap Aggregating

Decision Tree $\xrightarrow{\text{Bagging}}$ Random Forest

Boosting

Adaboosting: Adaptive Boosting



Gradient Boosting

