Insper

Machine Learning

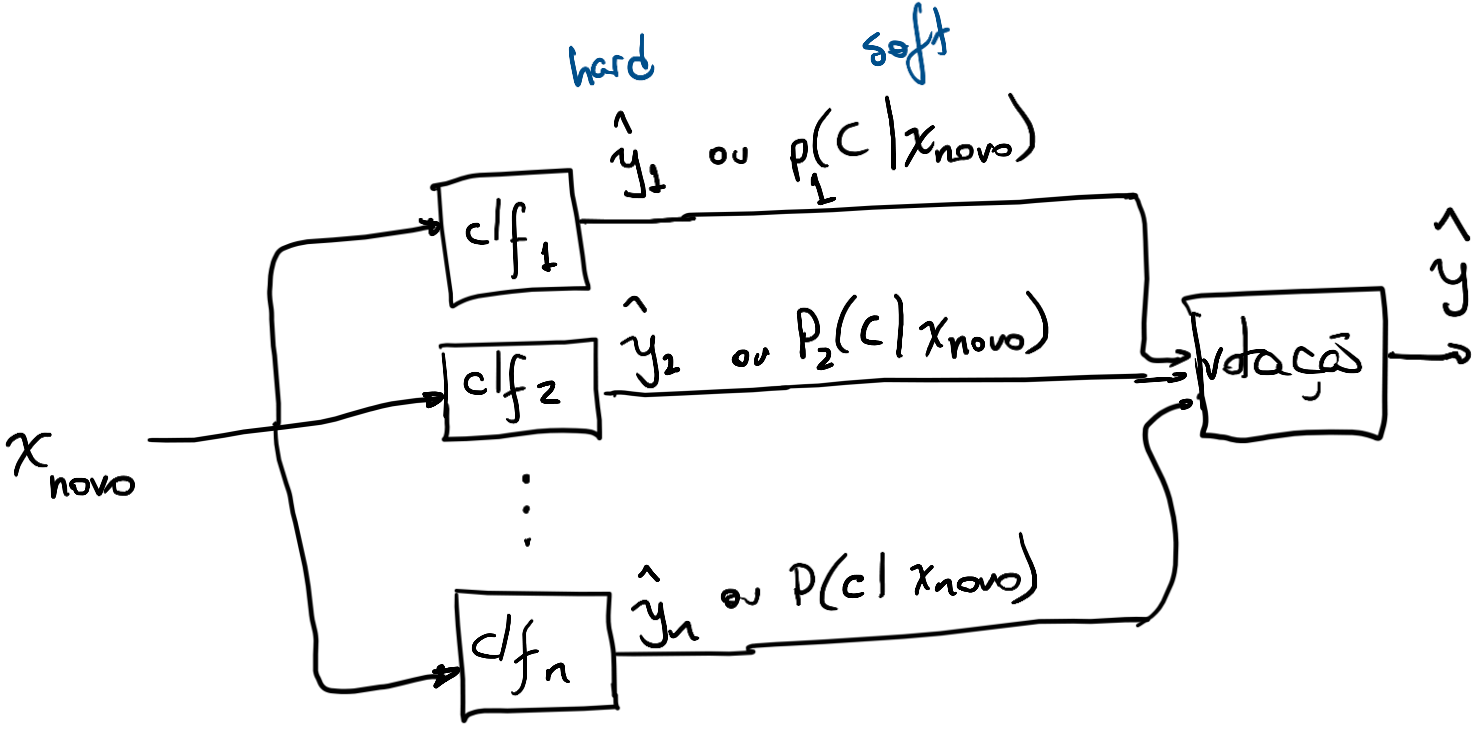
Ensemble Models

Fábio Ayres <fabioja@insper.edu.br>

- Wisdom of the crowds
  - multiplos modelos "fracos" em modelo "forte"

- Requer:
  - prob. de acerto > 50% p/ cada modelo fraco
  - num. grande de modelos p/ cancelar o erro
  - independencia
    - efeito manada não existe
      - $P(A \mid B) = P(A)$
    - Algoritmos diferentes, mesmos dados
    - Mesmo algoritmo, dados diferentes.

Voting Classifier



Bootstrap

- distribuição desconhecida F (população)
- Coleto m amostras i.i.d.

$(x_1, x_2, \cdots, x_m)$ ← conjunto de observações

- Quero estimar $\overline{F} = E_F[X]$

$$\overline{F} = \frac{1}{n}\sum_{i=1}^{m} x_i \quad \text{(media amostral)}$$

Considere o seguinte:

- cada x_i foi amostrado de F
- Se eu repetir a amostragem, obtenho outros x_i

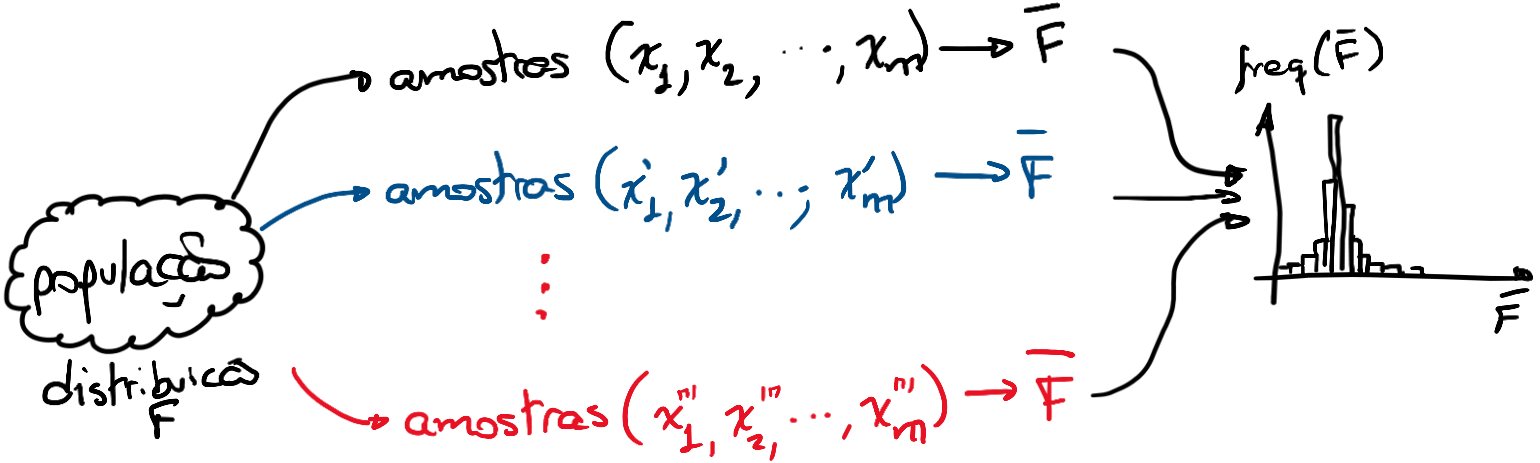
⇒ obtenho outro $\overline{F}$

⇒ $\overline{F}$ é variável aleatória ⇒ tem uma distribuição!

Histograma de $\bar{F} = \frac{1}{m}\sum_{i=1}^{m} X_i$ onde X_i v.a. com distribuição F

---

em outras palavras:



Exemplo: Será que o pãozinho da padaria pesa em média mais que 50 g?

⇒ teste de hipótese!

1) $H_0: \mu \leq 50$
$H_A: \mu > 50$

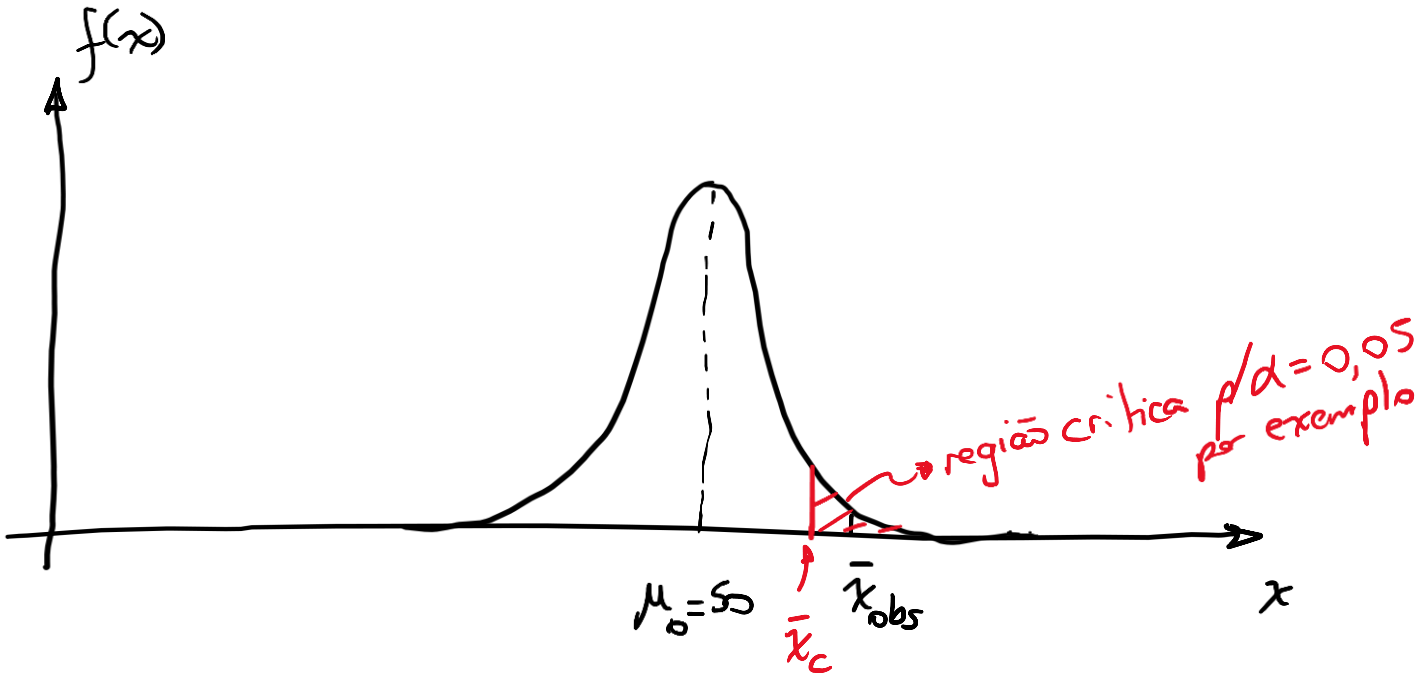
↖ média populacional (desconhecida)

requer populações com distr. normal!

2) Estatística de teste: estatística t-student

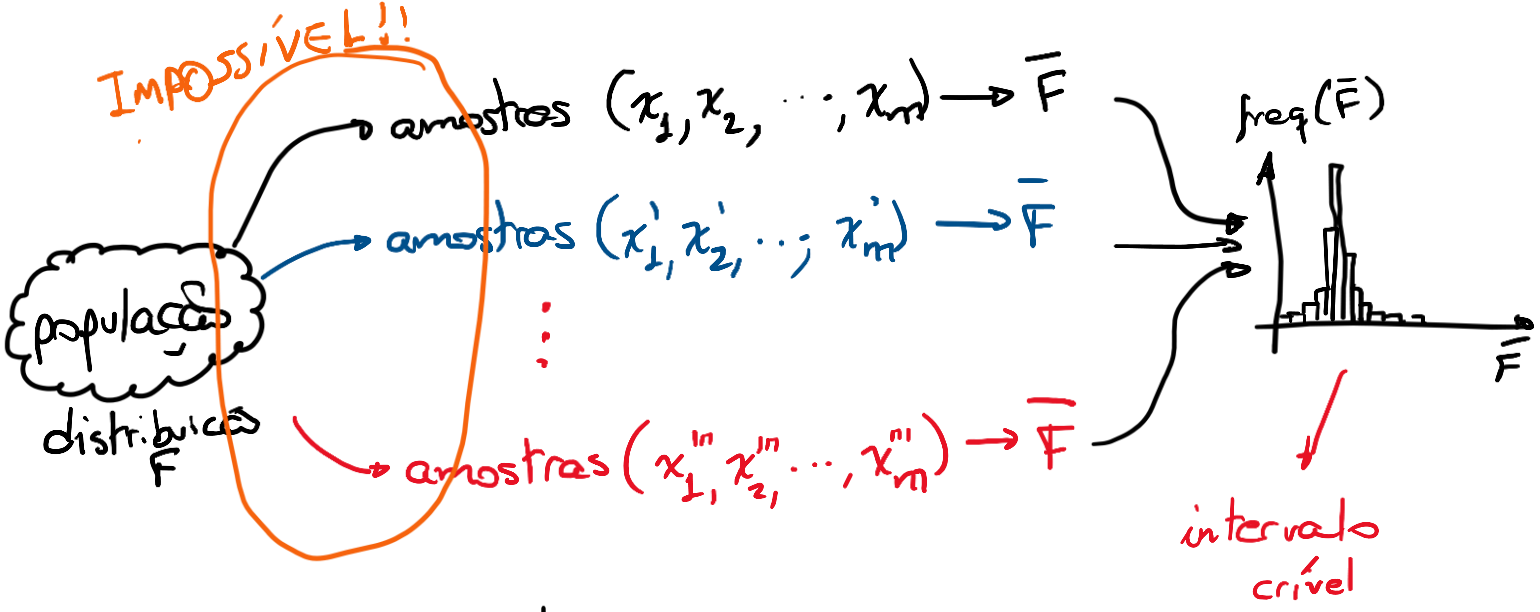
3) Critério de rejeição de H_0: → Região crítica
→ valor-p
↳ escolher um α

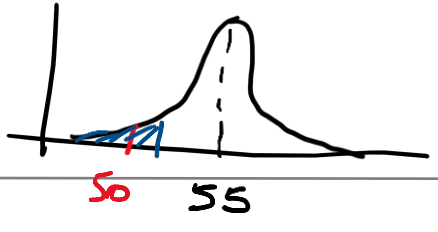
4) Calcula $\bar{X}$ → verifica se H_0 foi rejeitada


f(x)
região crítica p/ α = 0,05 por exemplo
μ0 = 55
x̄c
x̄obs
x

E se não quero supor $x_i \sim \text{Normal}(\mu, \sigma^2)$?
Ou nenhuma outra?

⇒ métodos não-paramétricos !

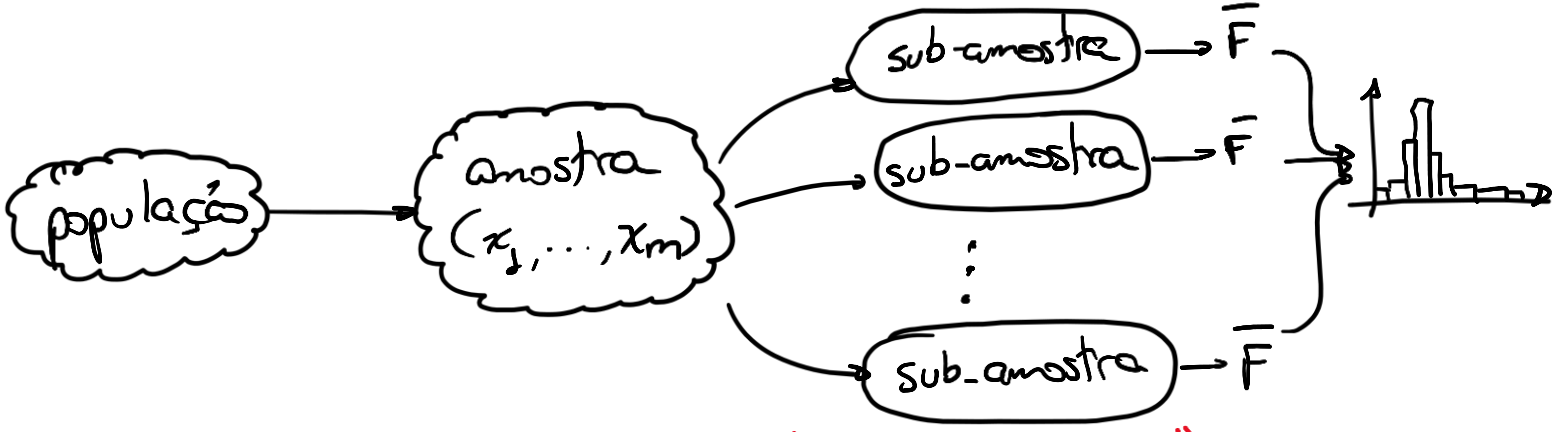




E se eu tenho apenas 1 conjunto de observações?

BOOTSTRAP!

- Gerar conjuntos de observações a partir de um único conjunto através de amostragem com reposição

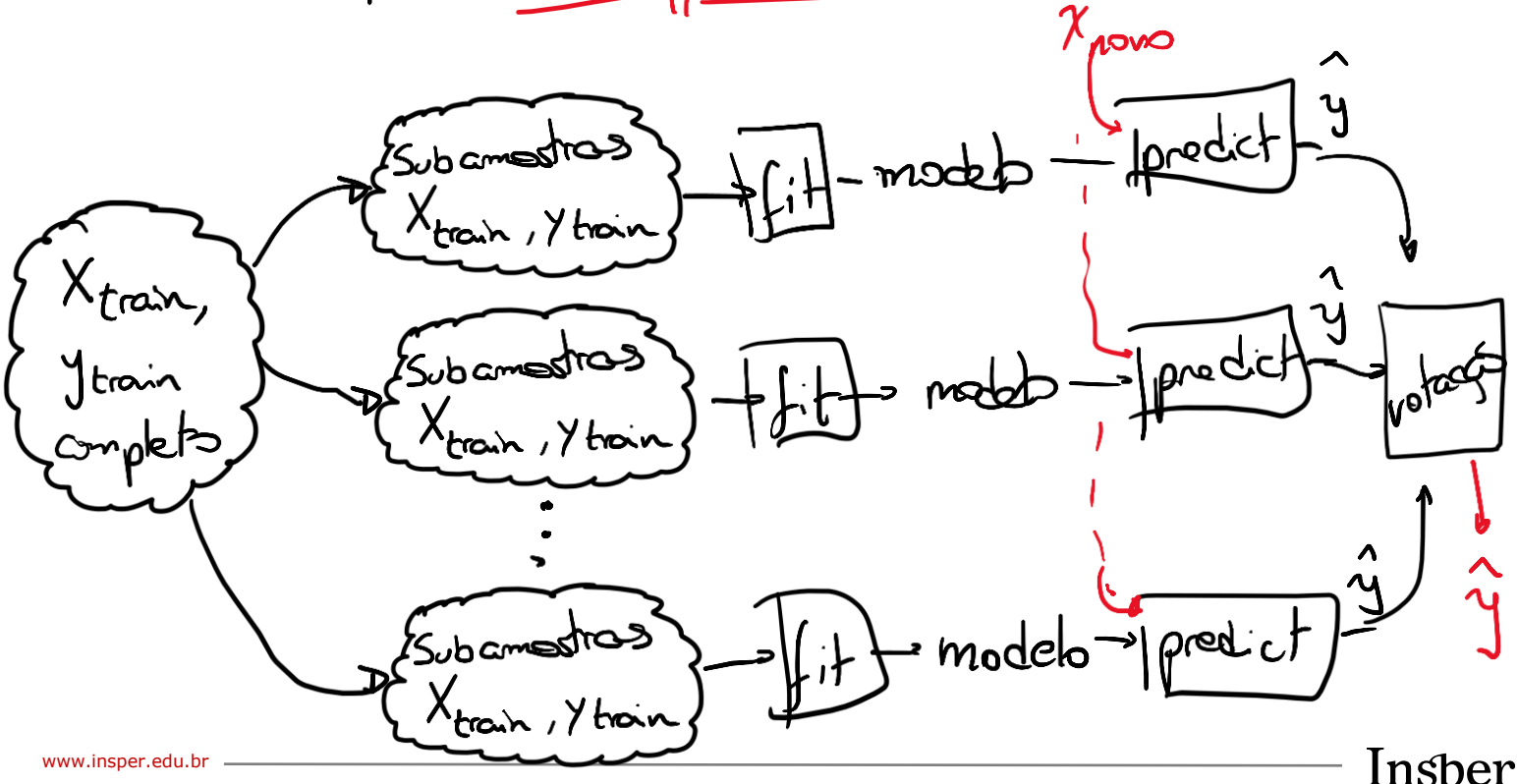


"to pull oneself up by one's own bootstraps"

se erguer sozinho a partir do nada

Bradley Efron (1979)

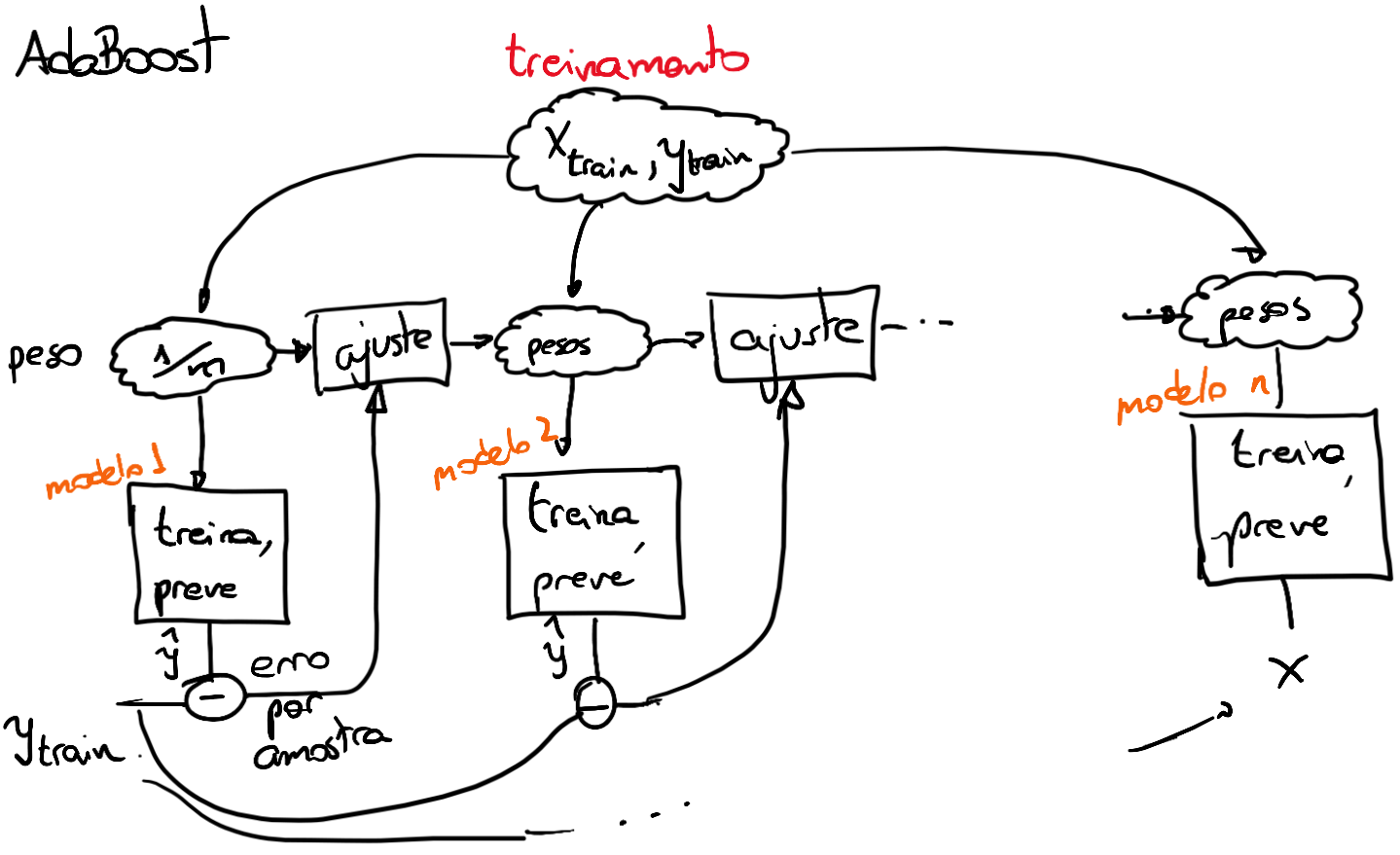
Bootstrap: permite obter amostras de uma estatística de interesse a partir de um conjunto de observações, sem supor nada sobre a distribuição populacional



Bagging : Bootstrap aggregating (com repetições)

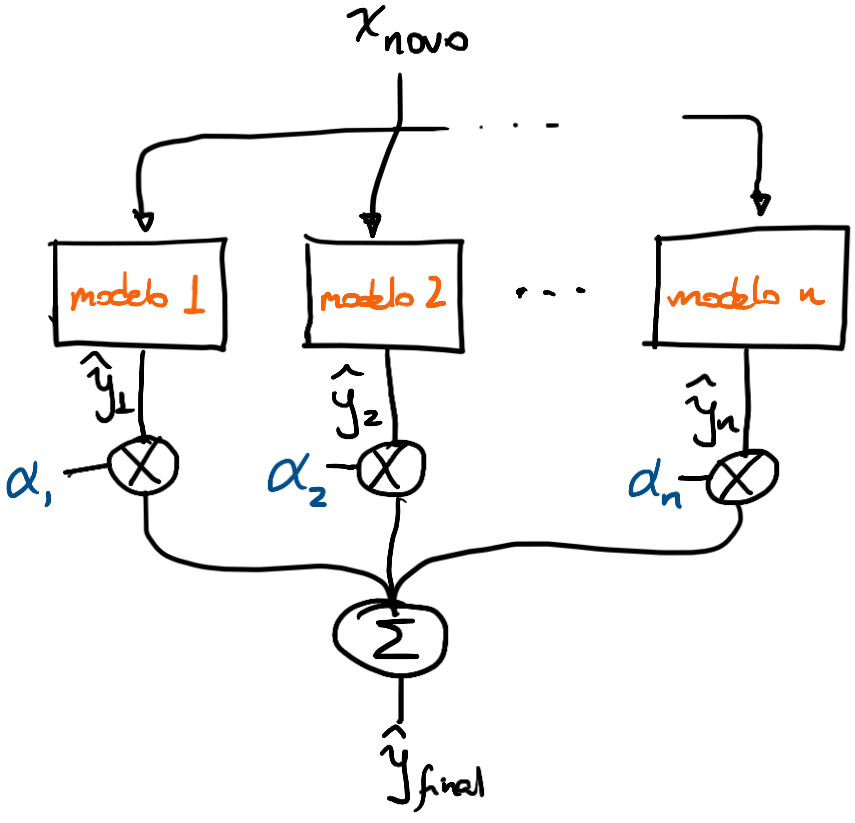
Pasting : o mesmo, mas sem repetições

Boosting: aprendizado sequencial de uma coleção de modelos



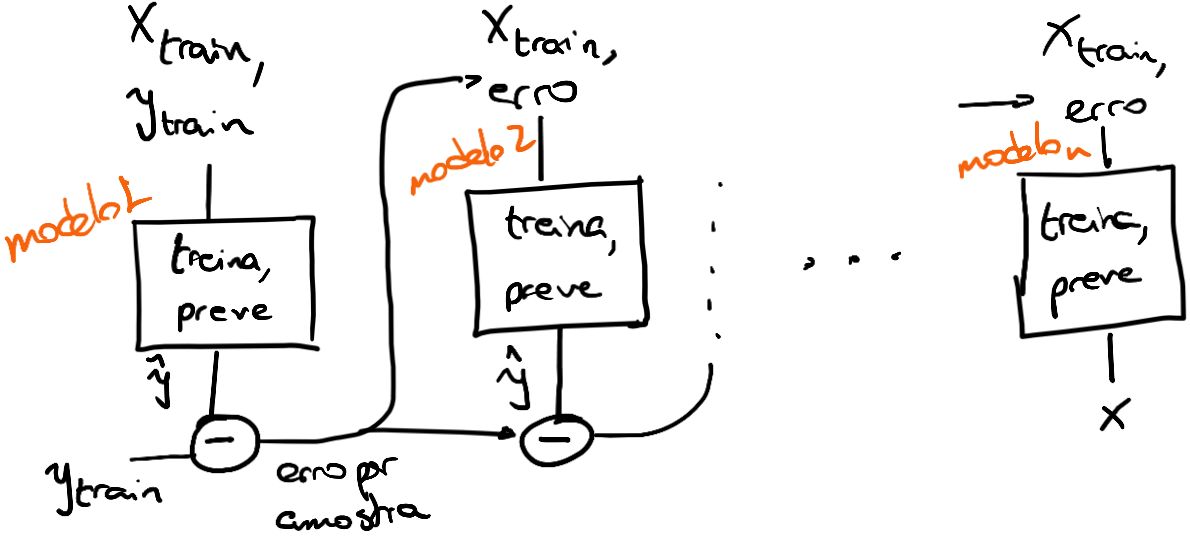
AdaBoost

predições

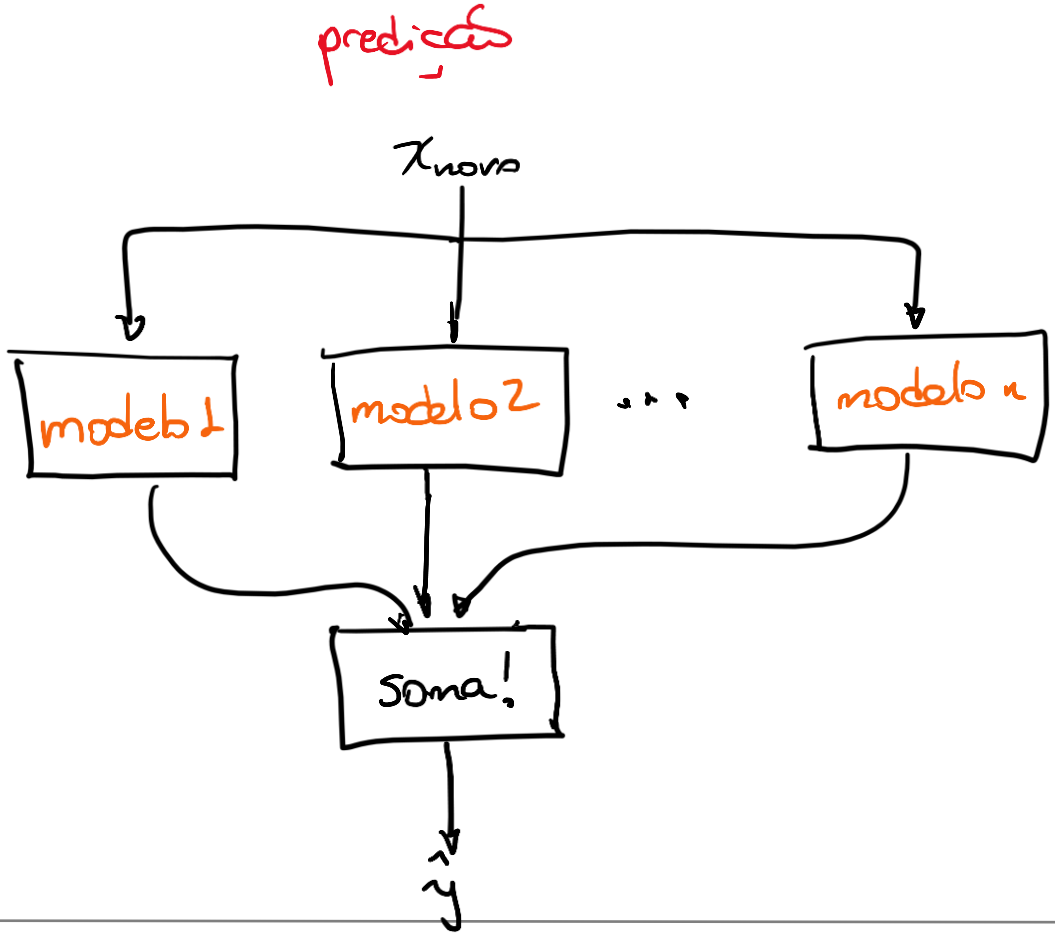


Gradient Boosting:

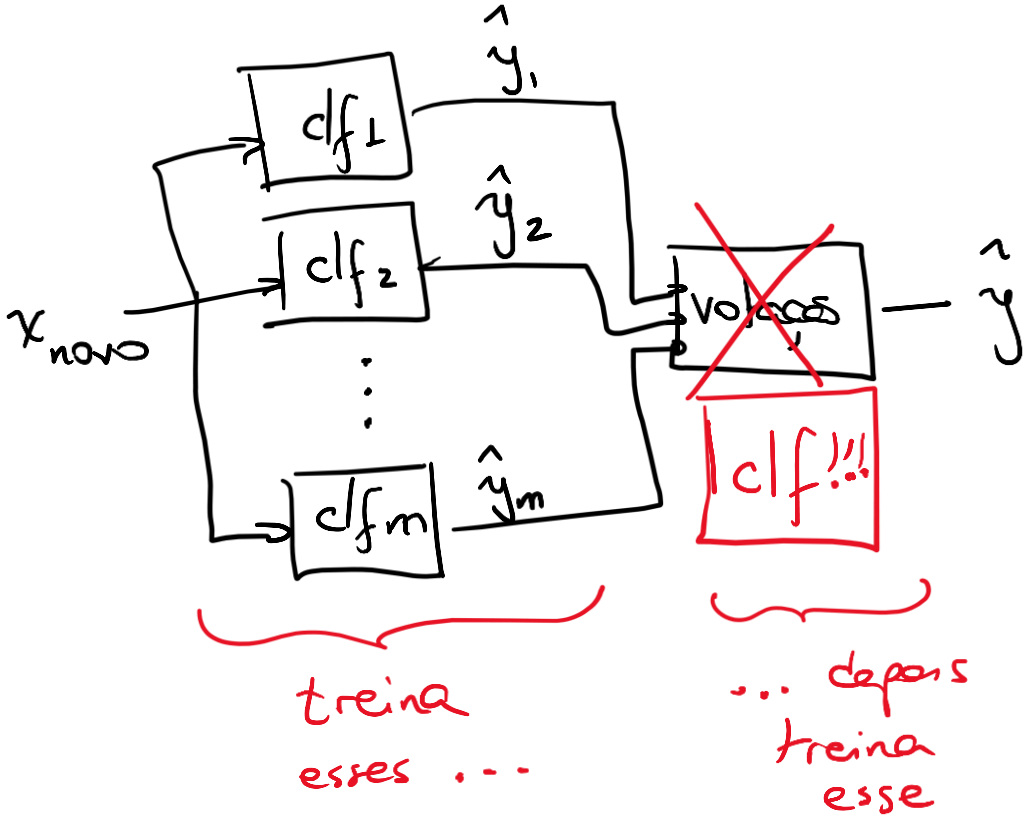
treinamento



Gradient Boosting



Stacking


$\hat{y}_1$
clf 1
$\hat{y}_2$
clf 2
x_{novo}
votação
$\hat{y}$
clf m
$\hat{y}_m$
clf !!!
treina esses ...
... depois treina esse

Insper