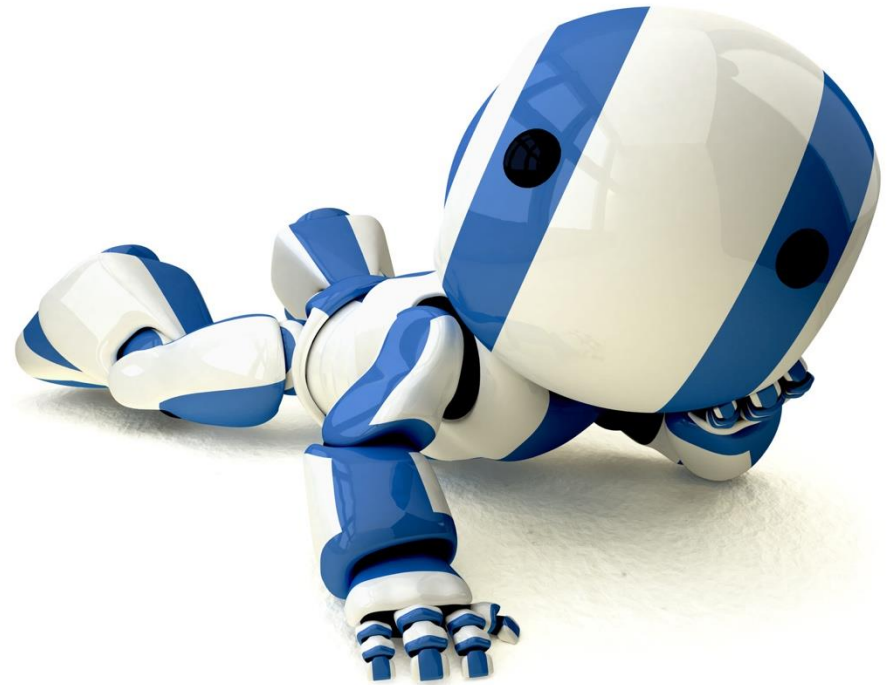


# Aprendizado de Máquina

Paradigma baseado em Comitês:  
*Ensemble Learning*

Prof. Me. Otávio Parraga



# MALTA

Machine Learning Theory  
and Applications Lab

# Aula de Hoje

- *Trade-off Bias-Variância*
- *Ensemble Learning*
  - Bagging (redução de variância)
    - Random Forests
  - Boosting (redução de *bias*)
    - AdaBoost

# Aula de Hoje

- *Trade-off Bias-Variância*
- *Ensemble Learning*
  - Bagging (redução de variância)
    - Random Forests
  - Boosting (redução de *bias*)
    - AdaBoost

# Decomposição Bias-Variância

VARIÂNCIA

baixa

alta

baixo

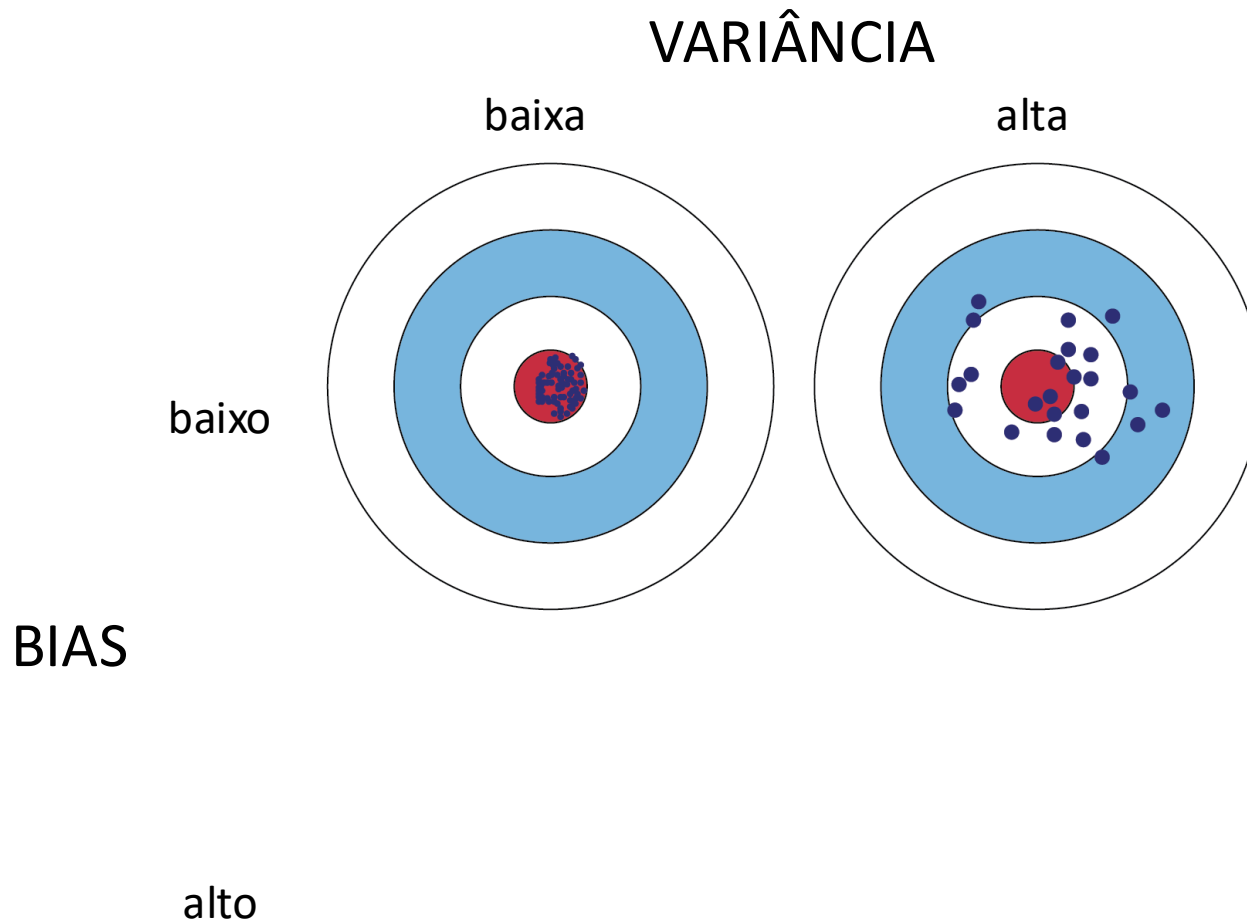
BIAS

alto

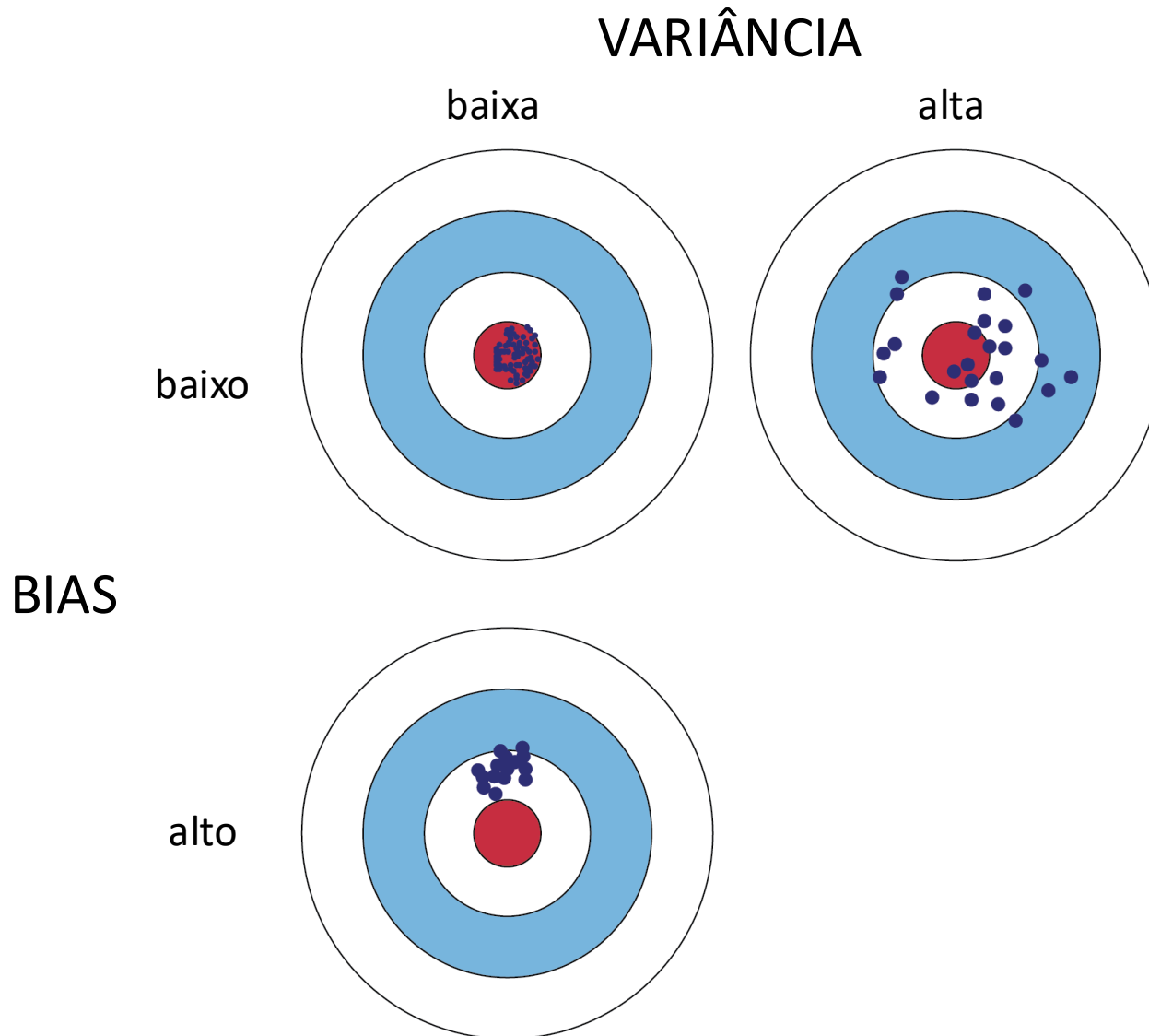
# Decomposição Bias-Variância



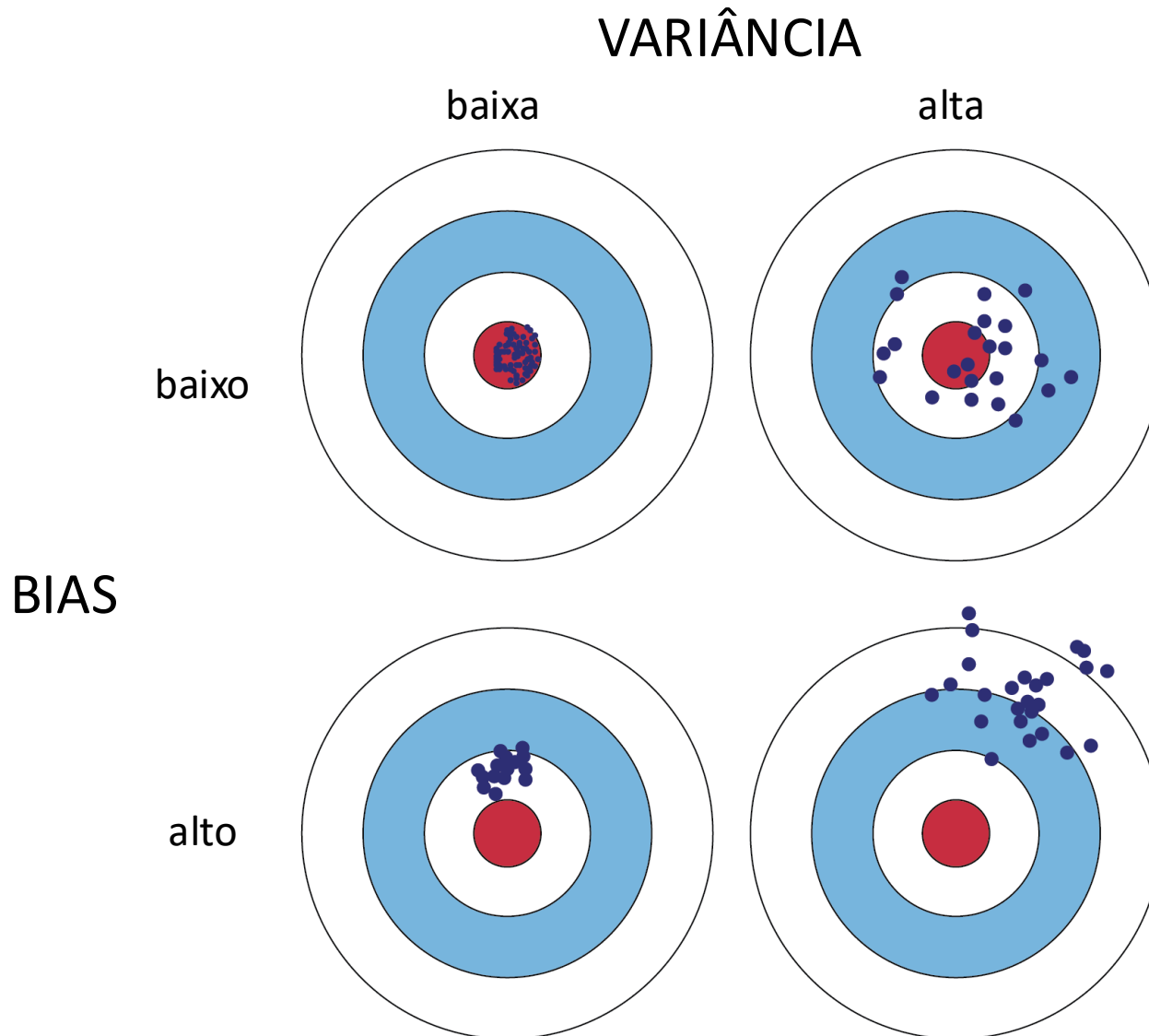
# Decomposição Bias-Variância



# Decomposição Bias-Variância



# Decomposição Bias-Variância





# Aula de Hoje

- *Trade-off Bias-Variância*
- *Ensemble Learning*
  - Bagging (redução de variância)
    - Random Forests
  - Boosting (redução de *bias*)
    - AdaBoost

# *Ensemble Learning*

- Ideia
  - A opinião de vários especialistas é **melhor** que a de apenas um.
  - “Na média, o **consenso da maioria** acerta mais que uma **opinião individual**”

# *Ensemble Learning*

- Ideia
  - Em vez de treinarmos um **único classificador**, treinamos **vários** (comitê) e agregamos os resultados para gerar a predição final
  - Muitos jeitos diferentes de fazer isso. Veremos os dois principais:
    - Bagging (técnica de redução de variância)
    - Boosting (técnica de redução de bias)

# *Ensemble Learning*

- Em Teoria:
  - Cada modelo do comitê possui a **mesma taxa de erro**
  - Todos os erros são **independentes**
  - Dessa forma: erro do Ensemble é menor que o erro dos modelos individuais

# *Ensemble Learning*

- Um comitê de modelos iguais não melhora os resultados
- Necessidade de adicionar **diversidade**
  - Manipulação do *Dataset*
  - Manipulação das *Features*
  - Manipulação do *Método de Aprendizagem*
  - ...

# *Ensemble Learning*

- A combinação dos resultados dos modelos pode ocorrer por
  - Votação
    - Uniforme ou com pesos
  - Por serialização
    - Soma, média, produto, máxima,...

# Aula de Hoje

- *Trade-off Bias-Variância*
- *Ensemble Learning*
  - Bagging (redução de variância)
    - Random Forests
  - Boosting (redução de *bias*)
    - AdaBoost

# Bagging

- Concebido por Leo Breiman em 1996
- A ideia é **marginalizar a variância** de vários classificadores gerados a partir de bootstrapping através da agregação de seus desempenhos
- Método que **reduz variância**



# Bagging

- Utilizamos um algoritmo de baixo bias e com alta variância e tentamos reduzir a variância

---

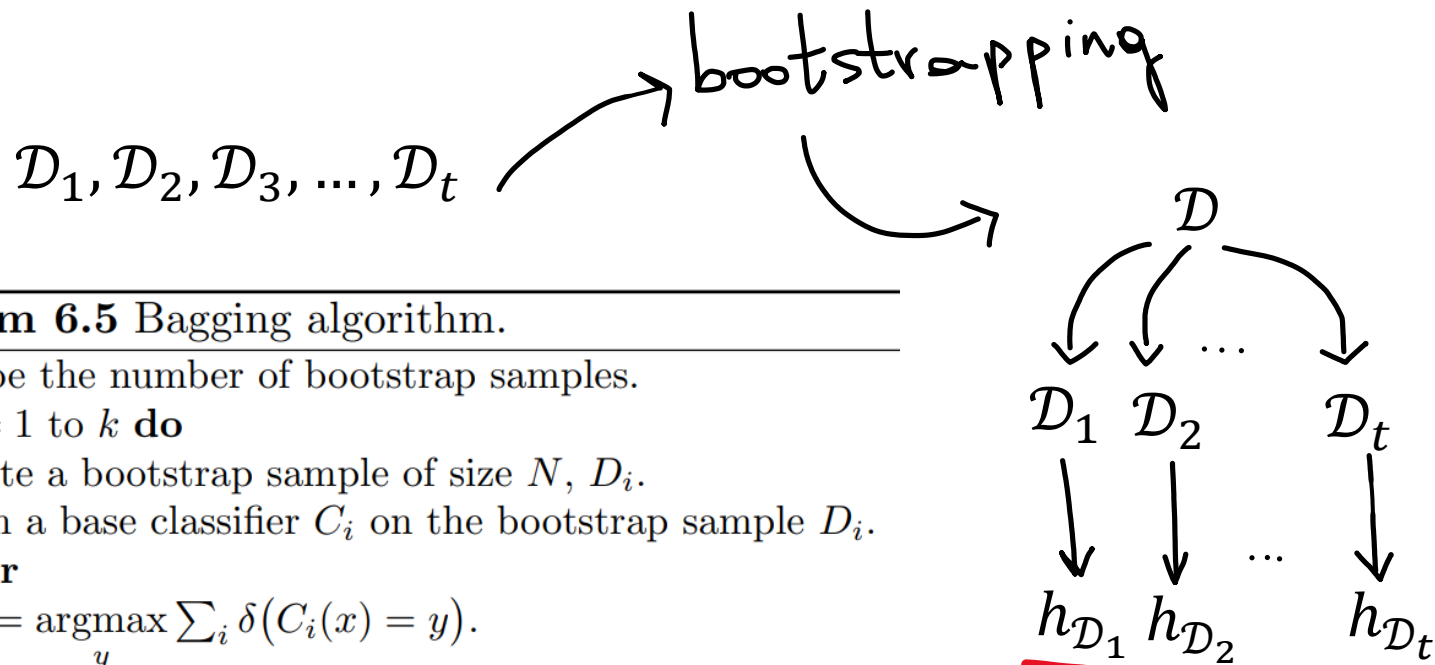
**Algorithm 6.5** Bagging algorithm.

---

- 1: Let  $k$  be the number of bootstrap samples.
  - 2: **for**  $i = 1$  to  $k$  **do**
  - 3:   Create a bootstrap sample of size  $N$ ,  $D_i$ .
  - 4:   Train a base classifier  $C_i$  on the bootstrap sample  $D_i$ .
  - 5: **end for**
  - 6:  $C^*(x) = \operatorname{argmax}_y \sum_i \delta(C_i(x) = y)$ .  
     $\{\delta(\cdot) = 1 \text{ if its argument is true and } 0 \text{ otherwise.}\}$
-

# Bagging

Vantagem imediata além da redução de variância:  
geração de **soluções probabilísticas** (basta vermos a fração  
de modelos que estão prevendo cada uma das classes!)



---

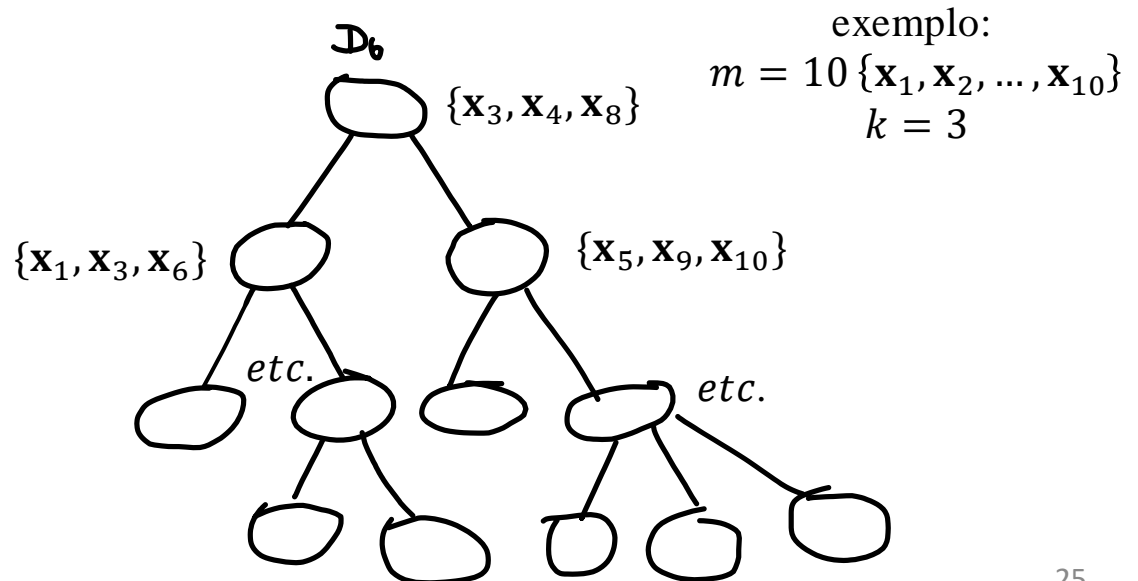
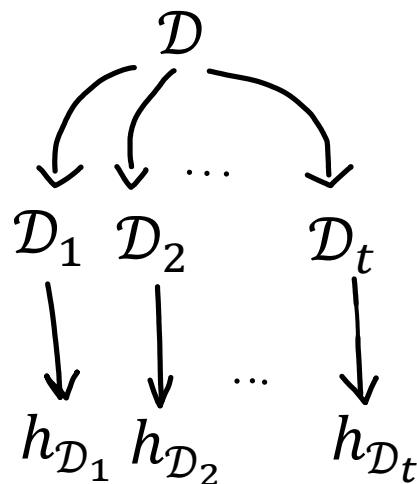
**Algorithm 6.5** Bagging algorithm.

---

- 1: Let  $k$  be the number of bootstrap samples.
  - 2: **for**  $i = 1$  to  $k$  **do**
  - 3:   Create a bootstrap sample of size  $N$ ,  $\mathcal{D}_i$ .
  - 4:   Train a base classifier  $C_i$  on the bootstrap sample  $\mathcal{D}_i$ .
  - 5: **end for**
  - 6:  $C^*(x) = \operatorname{argmax}_y \sum_i \delta(C_i(x) = y)$ .  
     $\{\delta(\cdot) = 1 \text{ if its argument is true and } 0 \text{ otherwise.}\}$
-

# Random Forests

- Também inventado por Leo Breiman (2001)
- **TL;DR**: bagging com árvores de decisão
- Twist: **aleatorizar** (por isso *random*) subsets de atributos pra cada divisão candidata



# Random Forests

- Apenas dois **hiperparâmetros**
  - **Número de árvores** (datasets)  $t$
  - **Número de atributos**  $k$  nos *subsets*
- Na prática, **não precisamos** otimizar esses valores
  - Valor de  $t$  limitado apenas pelo *budget* disponível
  - Utilizar  $k = \lceil \sqrt{m} \rceil$

# Random Forests

- Random Forests tem uma vantagem importante em termos de **protocolo de avaliação de resultados**
  - Na verdade é uma vantagem geral de bagging
- *Out-of-bag error*

$$h(\mathbf{x}^{(i)}) = \frac{1}{t} \sum_{j=1}^t h_{\mathcal{D}_j}(\mathbf{x}^{(i)})$$

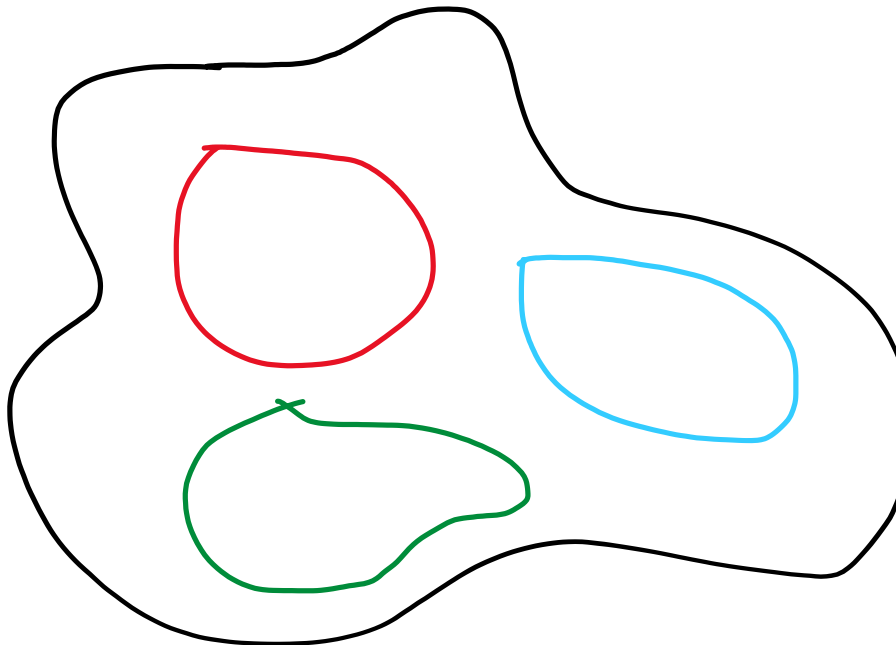
$$\ell(\mathbf{x}^{(i)}, y^{(i)}) = \frac{1}{t} \sum_{j=1}^t \ell(h_{\mathcal{D}_j}(\mathbf{x}^{(i)}), y^{(i)})$$

$$\varepsilon_{oob} = \frac{1}{N} \sum_{i=1}^N \frac{1}{Z_i} \sum_{\substack{j \\ (\mathbf{x}^{(i)}, y^{(i)}) \notin \mathcal{D}_j}} \ell(h_{\mathcal{D}_j}(\mathbf{x}^{(i)}), y^{(i)})$$

$$Z_i = \sum_{\substack{j \\ (\mathbf{x}^{(i)}, y^{(i)}) \notin \mathcal{D}_j}} 1$$

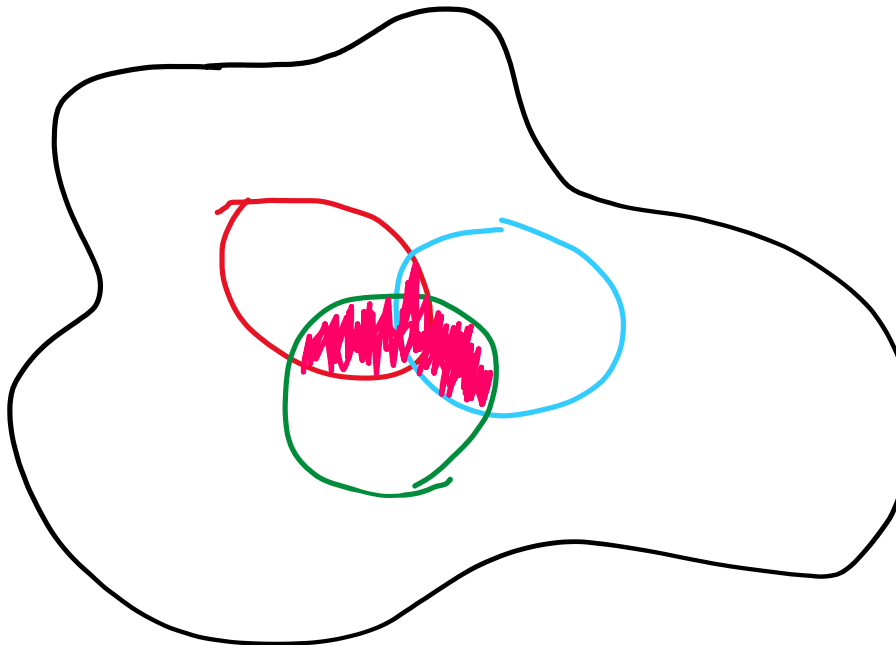
# Bagging

- Uma limitação importante de bagging diz respeito aos erros cometidos por cada aprendiz no espaço de objetos



# Bagging

- Uma limitação importante de bagging diz respeito aos erros cometidos por cada aprendiz no espaço de objetos



# Aula de Hoje

- *Trade-off Bias-Variância*
- *Ensemble Learning*
  - Bagging (redução de variância)
    - Random Forests
  - Boosting (redução de *bias*)
    - AdaBoost



# Boosting

- $\hat{f}(\mathbf{x}) = \text{sinal}(\underbrace{\hat{f}^1(\mathbf{x})}_{\text{red}} + \underbrace{\hat{f}^2(\mathbf{x})}_{\text{blue}} + \underbrace{\hat{f}^3(\mathbf{x})}_{\text{green}} + \dots), \quad \hat{f}^i(\mathbf{x}) = \{-1, +1\}$

Dados Originais



$\hat{f}^1$

Dados com número exagerado  
de erros cometidos por  $\hat{f}^1$



$\hat{f}^2$

Dados com número exagerado  
de objetos onde  $\hat{f}^1 \neq \hat{f}^2$

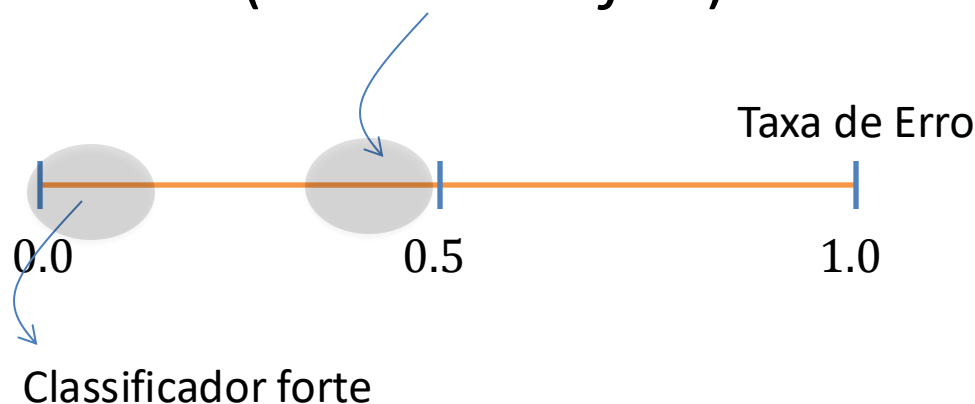


$\hat{f}^3$


...

# Adaptive Boosting (AdaBoost)

- Desenvolvido por Freund e Schapire em 1997
- Classificação binária,  $Y = \{-1, +1\}$
- A ideia é impulsionar (*boost*) classificadores fracos, combinando ao final suas previsões
- Algoritmo que reduz *bias*
- Classificador fraco (*weak classifier*): **Decision Stumps**



# Adaptive Boosting (AdaBoost)

- $\hat{f}(\mathbf{x}) = \text{sinal}(\underbrace{\hat{f}^1(\mathbf{x})}_{\text{red}} + \underbrace{\hat{f}^2(\mathbf{x})}_{\text{blue}} + \underbrace{\hat{f}^3(\mathbf{x})}_{\text{green}})$
- $\hat{f}(\mathbf{x}) = \text{sinal}(\alpha^1 \hat{f}^1(\mathbf{x}) + \alpha^2 \hat{f}^2(\mathbf{x}) + \alpha^3 \hat{f}^3(\mathbf{x}) + \dots)$ 


1)  $w_t^{(i)} = \frac{1}{N} \forall i \in X$

2) Treinar  $\hat{f}^t$  para minimizar  $\epsilon_t$

3) Computar  $\alpha^t$  de acordo com  $\epsilon_t$

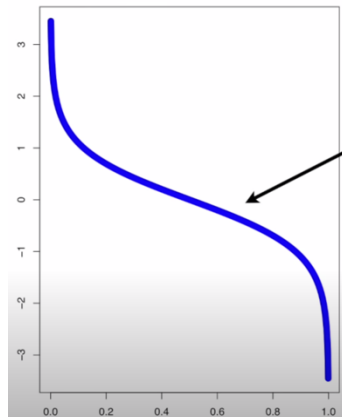
4) Computar  $w_{t+1}^{(i)}$  de acordo com  $\alpha^t$

# Adaptive Boosting (AdaBoost)

- $\hat{f}(\mathbf{x}) = \text{sinal}(\underbrace{\hat{f}^1(\mathbf{x})}_{\text{red}} + \underbrace{\hat{f}^2(\mathbf{x})}_{\text{blue}} + \underbrace{\hat{f}^3(\mathbf{x})}_{\text{green}})$
- $\hat{f}(\mathbf{x}) = \text{sinal}(\alpha^1 \hat{f}^1(\mathbf{x}) + \alpha^2 \hat{f}^2(\mathbf{x}) + \alpha^3 \hat{f}^3(\mathbf{x}) + \dots)$   



3) Computar  $\alpha^t$

$$\alpha^t = \frac{1}{2} \ln \frac{1 - \epsilon_t}{\epsilon_t}$$



valor de  $\alpha$  para  
diferentes valores  
de erros

# Adaptive Boosting (AdaBoost)

- $\hat{f}(\mathbf{x}) = \text{sinal}(\underbrace{\hat{f}^1(\mathbf{x})}_{\text{red}} + \underbrace{\hat{f}^2(\mathbf{x})}_{\text{blue}} + \underbrace{\hat{f}^3(\mathbf{x})}_{\text{green}})$
- $\hat{f}(\mathbf{x}) = \text{sinal}(\alpha^1 \hat{f}^1(\mathbf{x}) + \alpha^2 \hat{f}^2(\mathbf{x}) + \alpha^3 \hat{f}^3(\mathbf{x}) + \dots)$   


4) Computar  $w_{t+1}^{(i)}$

$$w_{t+1}^{(i)} = \frac{w_t^{(i)}}{Z} e^{-\alpha^t \hat{f}^t(\mathbf{x}^{(i)}) f(\mathbf{x}^{(i)})}$$

+ 1 SE  $\hat{f}$  ACERTOU  
- 1 CASO CONTRÁRIO

Lembrando que ambos

$$\hat{f} \text{ e } f \in \{-1, +1\}$$

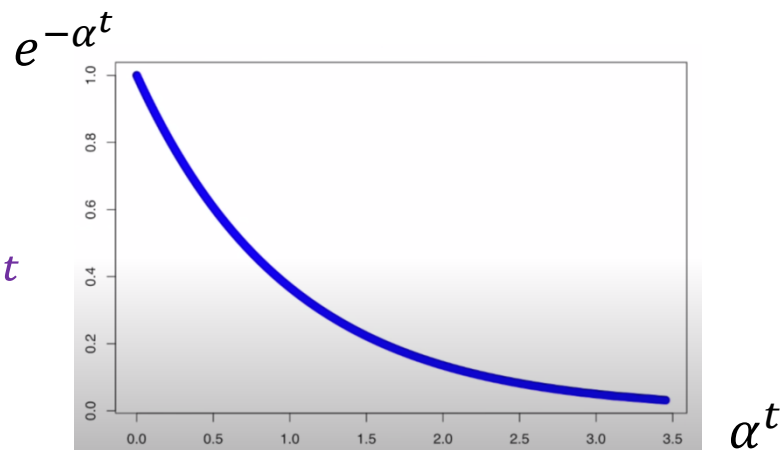
# Adaptive Boosting (AdaBoost)

- $\hat{f}(\mathbf{x}) = \text{sinal}(\underbrace{\hat{f}^1(\mathbf{x})}_{\text{red}} + \underbrace{\hat{f}^2(\mathbf{x})}_{\text{blue}} + \underbrace{\hat{f}^3(\mathbf{x})}_{\text{green}})$
- $\hat{f}(\mathbf{x}) = \text{sinal}(\alpha^1 \hat{f}^1(\mathbf{x}) + \alpha^2 \hat{f}^2(\mathbf{x}) + \alpha^3 \hat{f}^3(\mathbf{x}) + \dots)$


4) Computar  $w_{t+1}^{(i)}$

se o objeto foi corretamente classificado por  $\hat{f}^t$ , podemos reduzir seu peso de acordo com o fator de redução  $e^{-\alpha^t}$

$$w_{t+1}^{(i)} = \frac{w_t^{(i)}}{Z} e^{-\alpha^t}$$



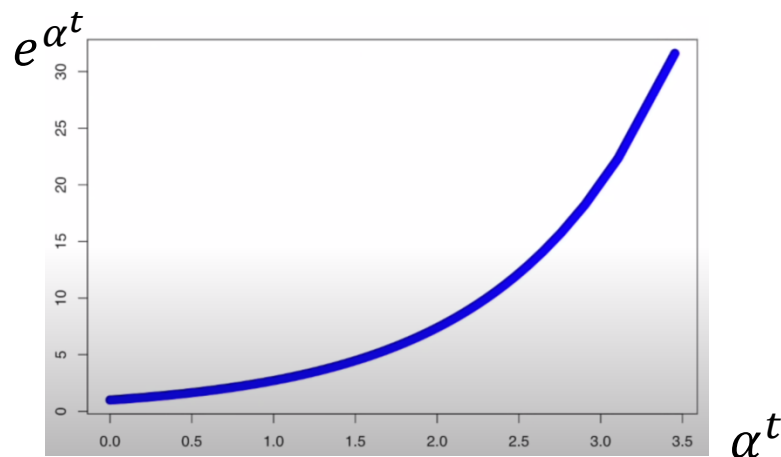
# Adaptive Boosting (AdaBoost)

- $\hat{f}(\mathbf{x}) = \text{sinal}(\underbrace{\hat{f}^1(\mathbf{x})}_{\text{red}} + \underbrace{\hat{f}^2(\mathbf{x})}_{\text{blue}} + \underbrace{\hat{f}^3(\mathbf{x})}_{\text{green}})$
- $\hat{f}(\mathbf{x}) = \text{sinal}(\alpha^1 \hat{f}^1(\mathbf{x}) + \alpha^2 \hat{f}^2(\mathbf{x}) + \alpha^3 \hat{f}^3(\mathbf{x}) + \dots)$   


4) Computar  $w_{t+1}^{(i)}$

se o objeto foi incorretamente classificado por  $\hat{f}^t$ , podemos aumentar seu peso de acordo com o fator de aumento  $e^{\alpha^t}$

$$w_{t+1}^{(i)} = \frac{w_t^{(i)}}{Z} e^{\alpha^t}$$



# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca
SIM	SIM	100	SIM
NÃO	SIM	90	SIM
SIM	NÃO	105	SIM
SIM	SIM	80	SIM
NÃO	SIM	70	NÃO
NÃO	SIM	50	NÃO
SIM	NÃO	79	NÃO
SIM	SIM	83	NÃO

$$1) w_t^{(i)} = \frac{1}{N} \quad \forall i \in X$$

2) Treinar  $\hat{f}^t$  para minimizar  $\epsilon_t$

3) Computar  $\alpha^t$  de acordo com  $\epsilon_t$

4) Computar  $w_{t+1}^{(i)}$  de acordo com  $\alpha^t$



# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
SIM	SIM	100	SIM	1/8
NÃO	SIM	90	SIM	1/8
SIM	NÃO	105	SIM	1/8
SIM	SIM	80	SIM	1/8
NÃO	SIM	70	NÃO	1/8
NÃO	SIM	50	NÃO	1/8
SIM	NÃO	79	NÃO	1/8
SIM	SIM	83	NÃO	1/8

$$1) w_t^{(i)} = \frac{1}{N} \quad \forall i \in X$$

# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
SIM	SIM	100	SIM	1/8
NÃO	SIM	90	SIM	1/8
SIM	NÃO	105	SIM	1/8
SIM	SIM	80	SIM	1/8
NÃO	SIM	70	NÃO	1/8
NÃO	SIM	50	NÃO	1/8
SIM	NÃO	79	NÃO	1/8
SIM	SIM	83	NÃO	1/8

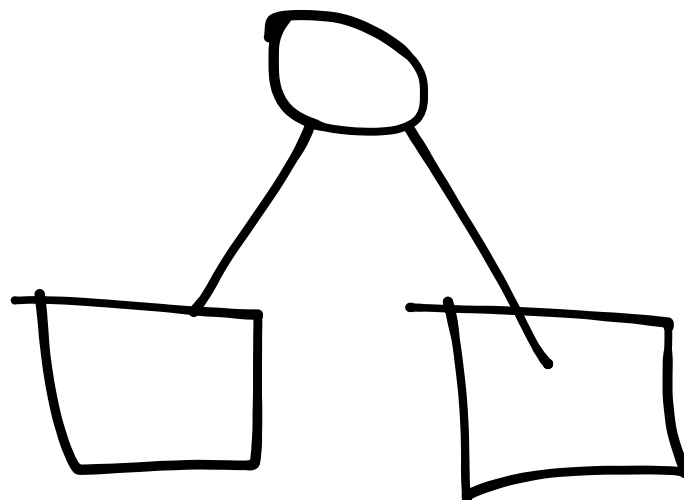
2) Treinar  $\hat{f}^t$  para minimizar  $\epsilon_t$

# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
SIM	SIM	100	SIM	1/8
NÃO	SIM	90	SIM	1/8
SIM	NÃO	105	SIM	1/8
SIM	SIM	80	SIM	1/8
NÃO	SIM	70	NÃO	1/8
NÃO	SIM	50	NÃO	1/8
SIM	NÃO	79	NÃO	1/8
SIM	SIM	83	NÃO	1/8

2) Treinar  $\hat{f}^t$  para minimizar  $\epsilon_t$

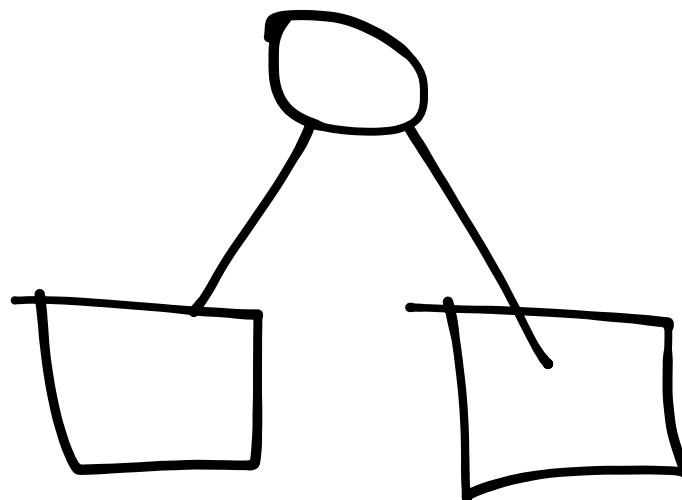


# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
SIM	SIM	100	SIM	1/8
NÃO	SIM	90	SIM	1/8
SIM	NÃO	105	SIM	1/8
SIM	SIM	80	SIM	1/8
NÃO	SIM	70	NÃO	1/8
NÃO	SIM	50	NÃO	1/8
SIM	NÃO	79	NÃO	1/8
SIM	SIM	83	NÃO	1/8

2) Treinar  $\hat{f}^t$  para minimizar  $\epsilon_t$

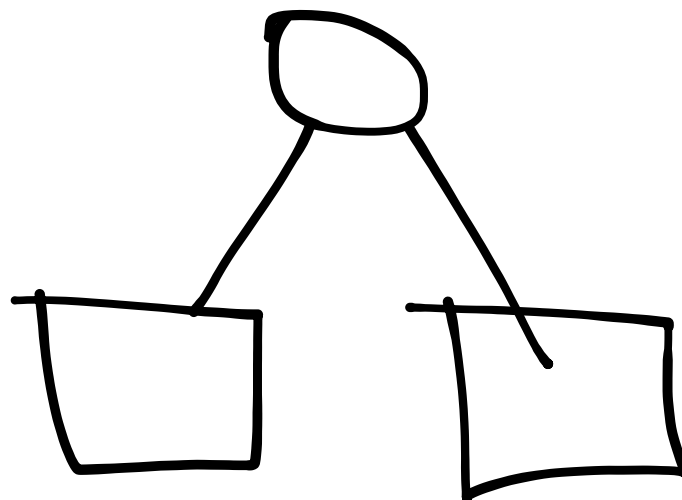


# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
SIM	SIM	100	SIM	1/8
NÃO	SIM	90	SIM	1/8
SIM	NÃO	105	SIM	1/8
SIM	SIM	80	SIM	1/8
NÃO	SIM	70	NÃO	1/8
NÃO	SIM	50	NÃO	1/8
SIM	NÃO	79	NÃO	1/8
SIM	SIM	83	NÃO	1/8

2) Treinar  $\hat{f}^t$  para minimizar  $\epsilon_t$

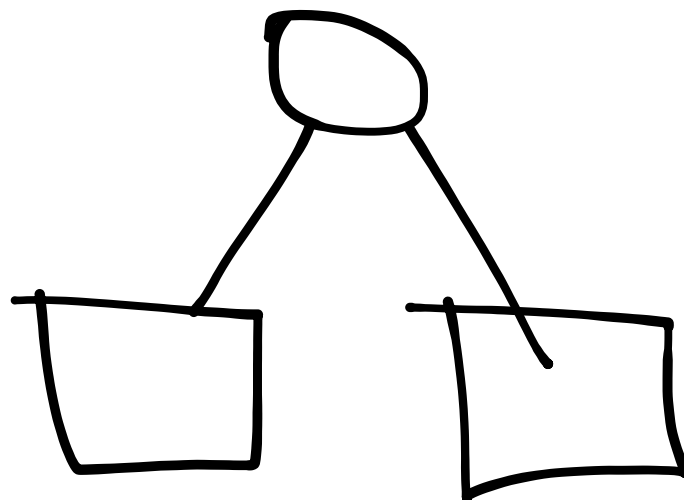


# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
SIM	SIM	100	SIM	1/8
NÃO	SIM	90	SIM	1/8
SIM	NÃO	105	SIM	1/8
SIM	SIM	80	SIM	1/8
NÃO	SIM	70	NÃO	1/8
NÃO	SIM	50	NÃO	1/8
SIM	NÃO	79	NÃO	1/8
SIM	SIM	83	NÃO	1/8

2) Treinar  $\hat{f}^t$  para minimizar  $\epsilon_t$

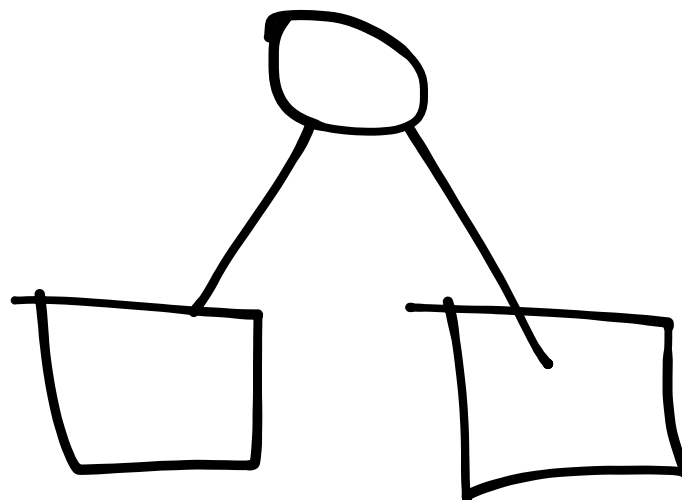


# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
SIM	SIM	100	SIM	1/8
NÃO	SIM	90	SIM	1/8
SIM	NÃO	105	SIM	1/8
SIM	SIM	80	SIM	1/8
NÃO	SIM	70	NÃO	1/8
NÃO	SIM	50	NÃO	1/8
SIM	NÃO	79	NÃO	1/8
SIM	SIM	83	NÃO	1/8

2) Treinar  $\hat{f}^t$  para minimizar  $\epsilon_t$



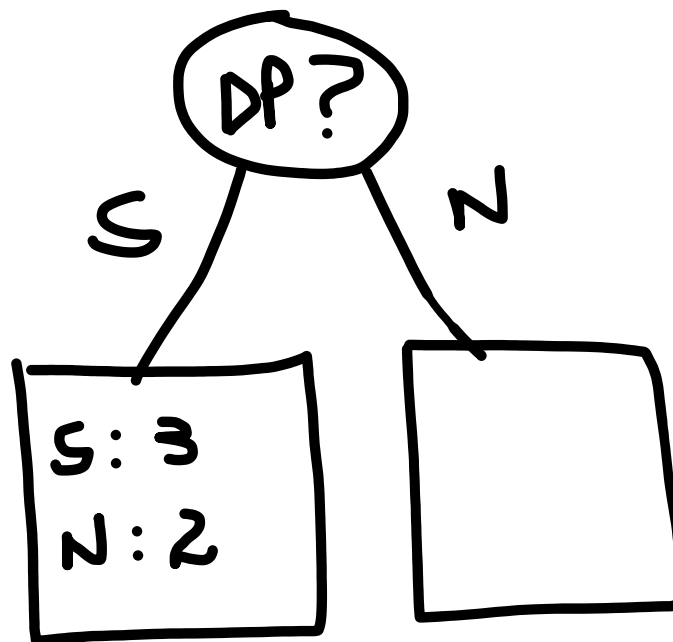
# Adaptive Boosting (AdaBoost)

- Um exemplo didático



Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
SIM	SIM	100	SIM	1/8
NÃO	SIM	90	SIM	1/8
SIM	NÃO	105	SIM	1/8
SIM	SIM	80	SIM	1/8
NÃO	SIM	70	NÃO	1/8
NÃO	SIM	50	NÃO	1/8
SIM	NÃO	79	NÃO	1/8
SIM	SIM	83	NÃO	1/8

2) Treinar  $\hat{f}^t$  para minimizar  $\epsilon_t$





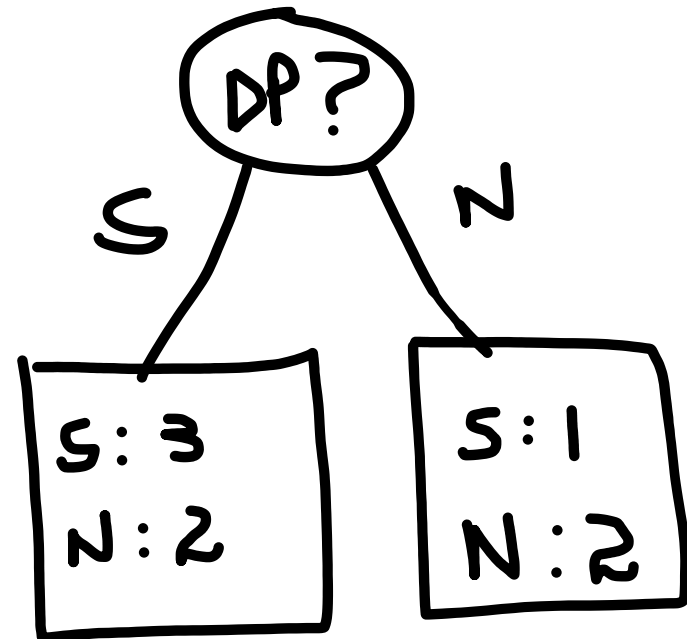
# Adaptive Boosting (AdaBoost)

- Um exemplo didático



Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
SIM	SIM	100	SIM	1/8
NÃO	SIM	90	SIM	1/8
SIM	NÃO	105	SIM	1/8
SIM	SIM	80	SIM	1/8
NÃO	SIM	70	NÃO	1/8
NÃO	SIM	50	NÃO	1/8
SIM	NÃO	79	NÃO	1/8
SIM	SIM	83	NÃO	1/8

2) Treinar  $\hat{f}^t$  para minimizar  $\epsilon_t$



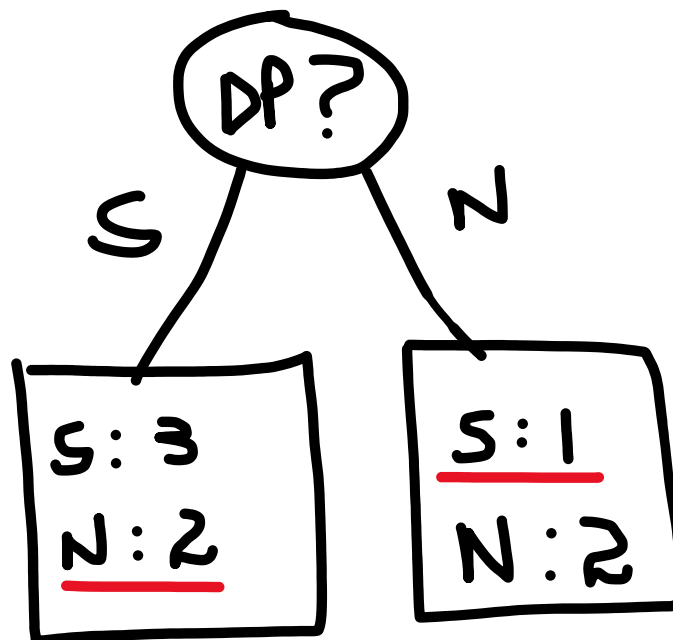
# Adaptive Boosting (AdaBoost)

- Um exemplo didático



Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
SIM	SIM	100	SIM	1/8
NÃO	SIM	90	SIM	1/8
SIM	NÃO	105	SIM	1/8
SIM	SIM	80	SIM	1/8
NÃO	SIM	70	NÃO	1/8
NÃO	SIM	50	NÃO	1/8
SIM	NÃO	79	NÃO	1/8
SIM	SIM	83	NÃO	1/8

2) Treinar  $\hat{f}^t$  para minimizar  $\epsilon_t$

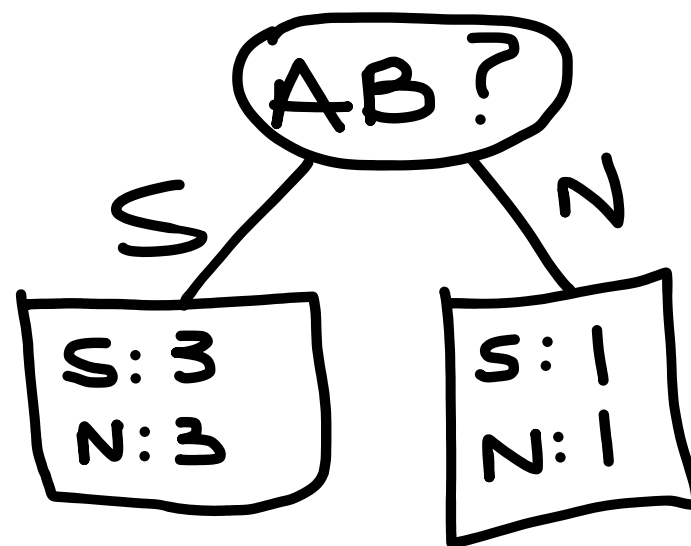


# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
SIM	SIM	100	SIM	1/8
NÃO	SIM	90	SIM	1/8
SIM	NÃO	105	SIM	1/8
SIM	SIM	80	SIM	1/8
NÃO	SIM	70	NÃO	1/8
NÃO	SIM	50	NÃO	1/8
SIM	NÃO	79	NÃO	1/8
SIM	SIM	83	NÃO	1/8

2) Treinar  $\hat{f}^t$  para minimizar  $\epsilon_t$

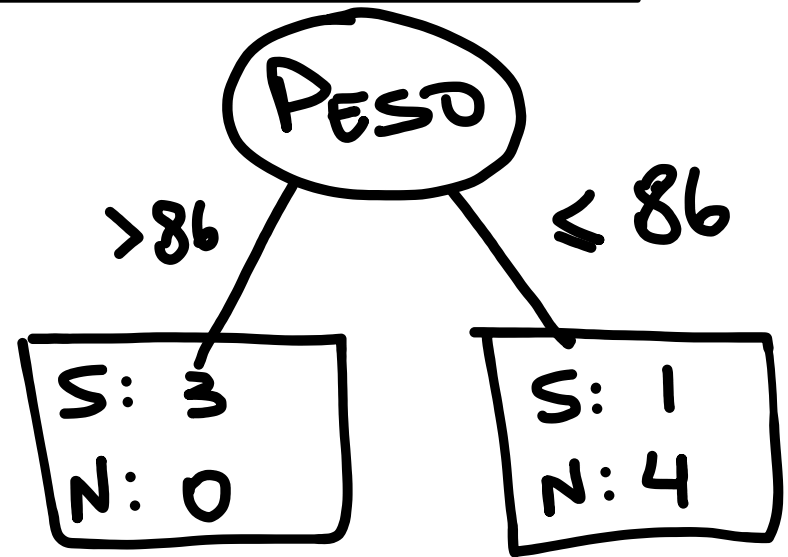


# Adaptive Boosting (AdaBoost)

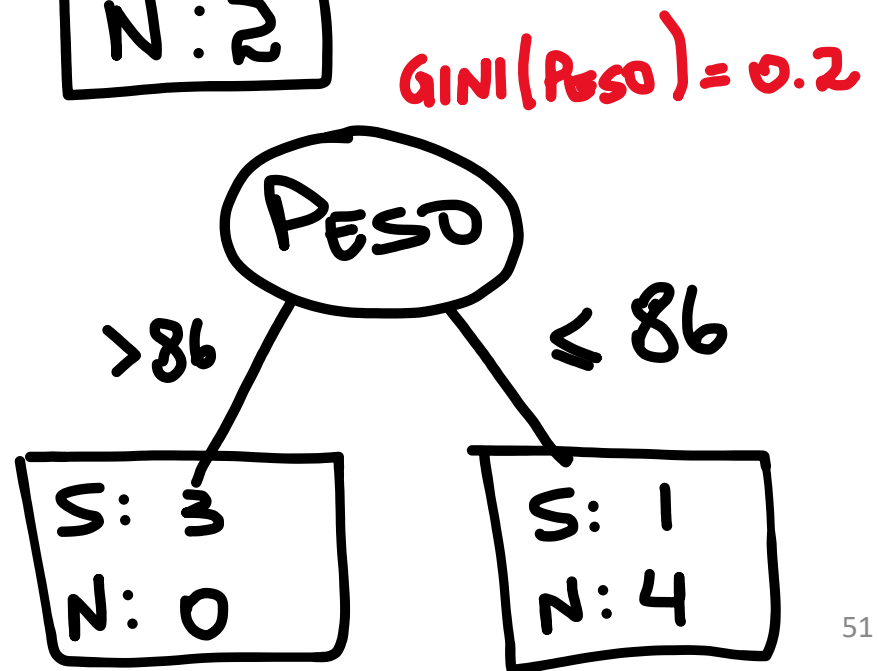
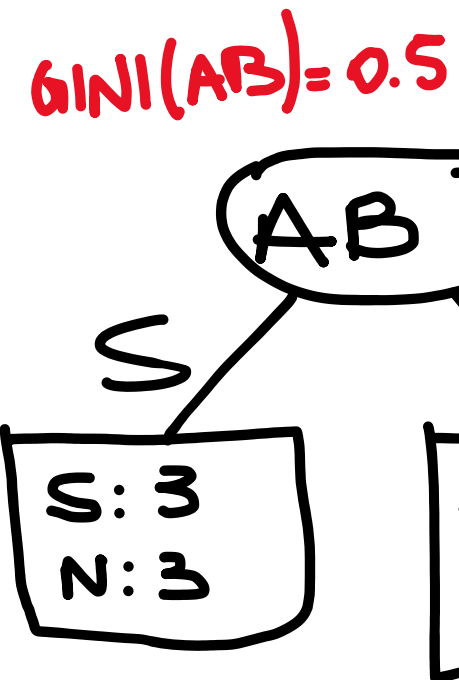
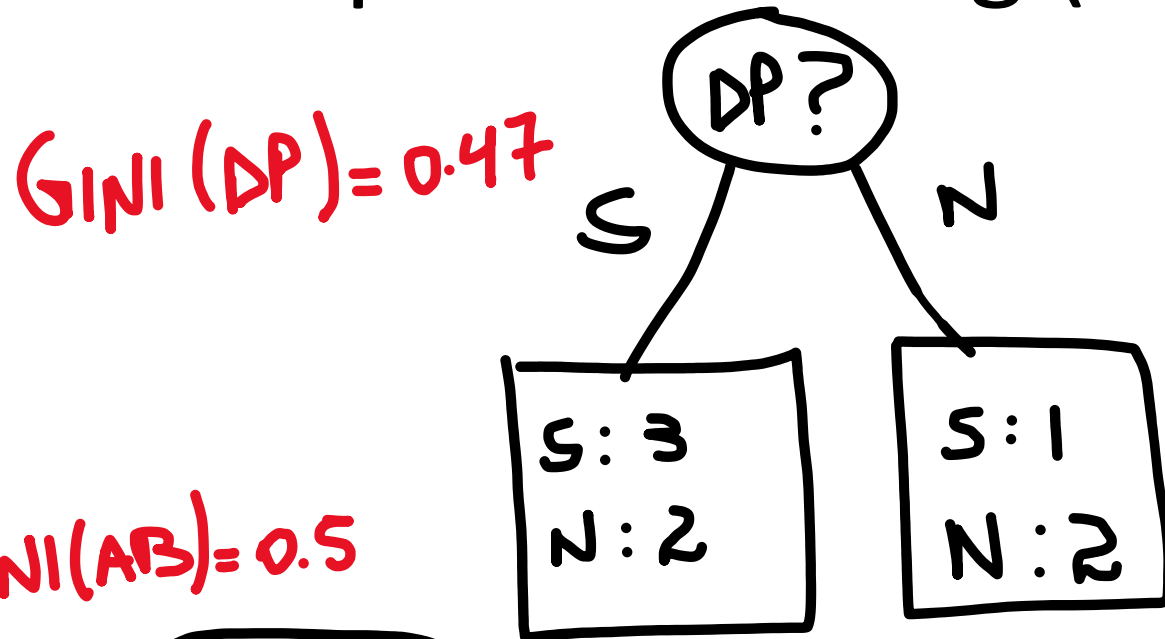
- Um exemplo didático

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
SIM	SIM	100	SIM	1/8
NÃO	SIM	90	SIM	1/8
SIM	NÃO	105	SIM	1/8
SIM	SIM	80	SIM	1/8
NÃO	SIM	70	NÃO	1/8
NÃO	SIM	50	NÃO	1/8
SIM	NÃO	79	NÃO	1/8
SIM	SIM	83	NÃO	1/8

2) Treinar  $\hat{f}^t$  para minimizar  $\epsilon_t$



# Adaptive Boosting (AdaBoost)

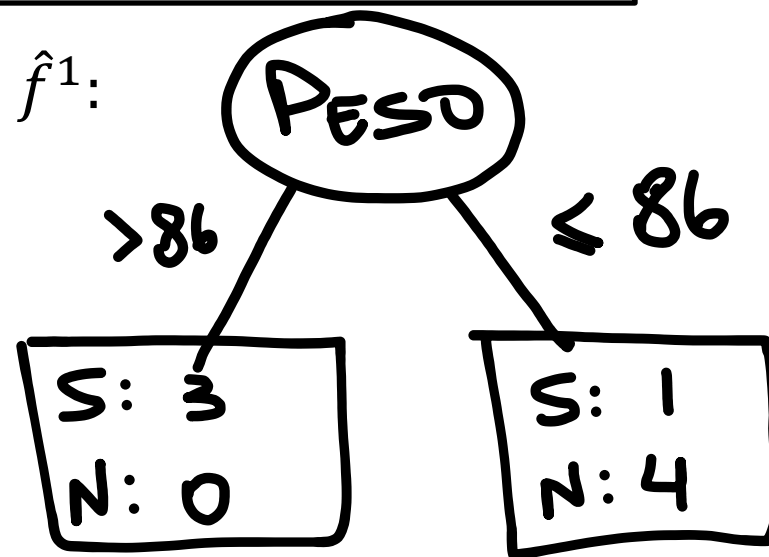


# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
SIM	SIM	100	SIM	1/8
NÃO	SIM	90	SIM	1/8
SIM	NÃO	105	SIM	1/8
SIM	SIM	80	SIM	1/8
NÃO	SIM	70	NÃO	1/8
NÃO	SIM	50	NÃO	1/8
SIM	NÃO	79	NÃO	1/8
SIM	SIM	83	NÃO	1/8

2) Treinar  $\hat{f}^t$  para minimizar  $\epsilon_t$

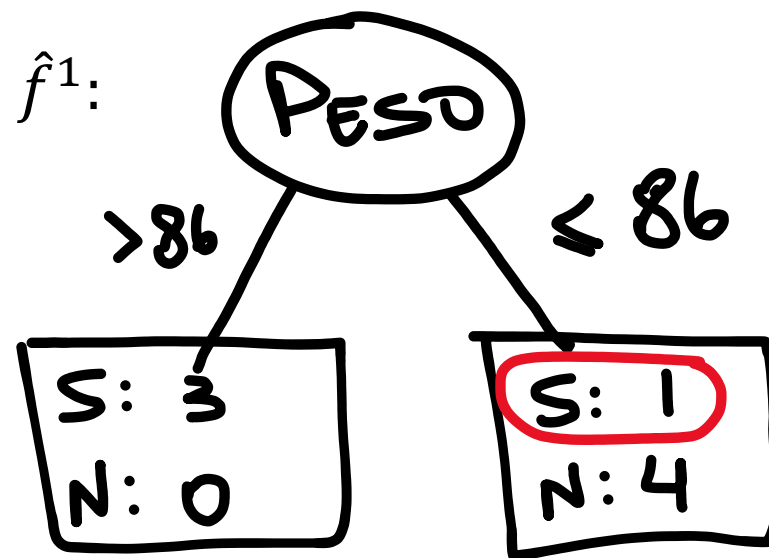


# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
SIM	SIM	100	SIM	1/8
NÃO	SIM	90	SIM	1/8
SIM	NÃO	105	SIM	1/8
SIM	SIM	80	SIM	1/8
NÃO	SIM	70	NÃO	1/8
NÃO	SIM	50	NÃO	1/8
SIM	NÃO	79	NÃO	1/8
SIM	SIM	83	NÃO	1/8

3) Computar  $\alpha^1$  de acordo com  $\epsilon_1$

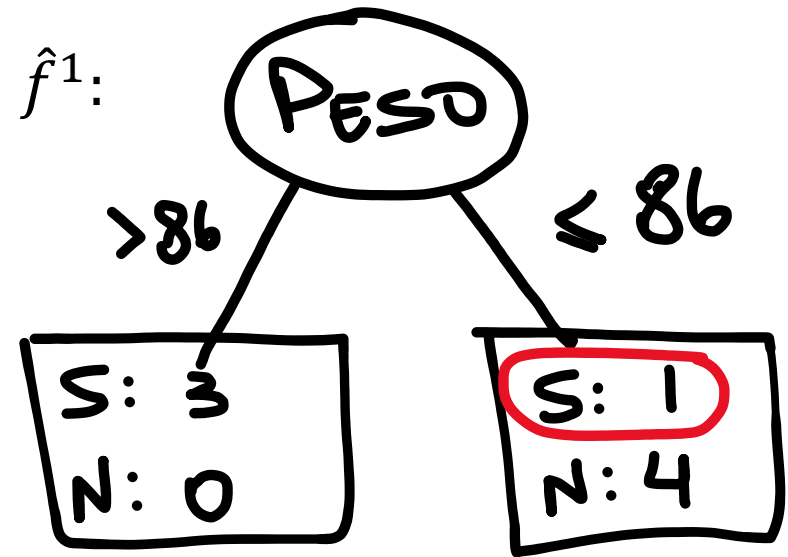


# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
SIM	SIM	100	SIM	1/8
NÃO	SIM	90	SIM	1/8
SIM	NÃO	105	SIM	1/8
SIM	SIM	80	SIM	1/8
NÃO	SIM	70	NÃO	1/8
NÃO	SIM	50	NÃO	1/8
SIM	NÃO	79	NÃO	1/8
SIM	SIM	83	NÃO	1/8

3) Computar  $\alpha^1$  de acordo com  $\epsilon_1$



$$\epsilon_1 = \sum_{\text{ERROS}} w_i^{(1)} = 1/8$$

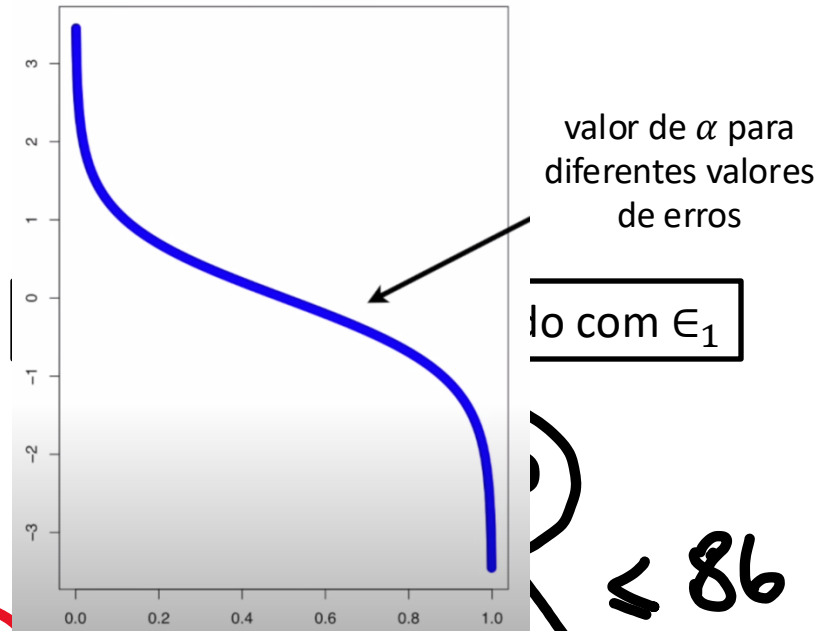


# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
SIM	SIM	100	SIM	1/8
NÃO	SIM	90	SIM	1/8
SIM	NÃO	105	SIM	1/8
SIM	SIM	80	SIM	1/8
NÃO	SIM	70	NÃO	1/8
NÃO	SIM	50	NÃO	1/8
SIM	NÃO	79	NÃO	1/8
SIM	SIM	83	NÃO	1/8

$$\alpha^1 = \frac{1}{2} \ln \frac{1 - \epsilon_1}{\epsilon_1} = \frac{1}{2} \ln \left( \frac{7/8}{1/8} \right) = 0.97$$



S: 3  
N: 0

S: 1  
N: 4

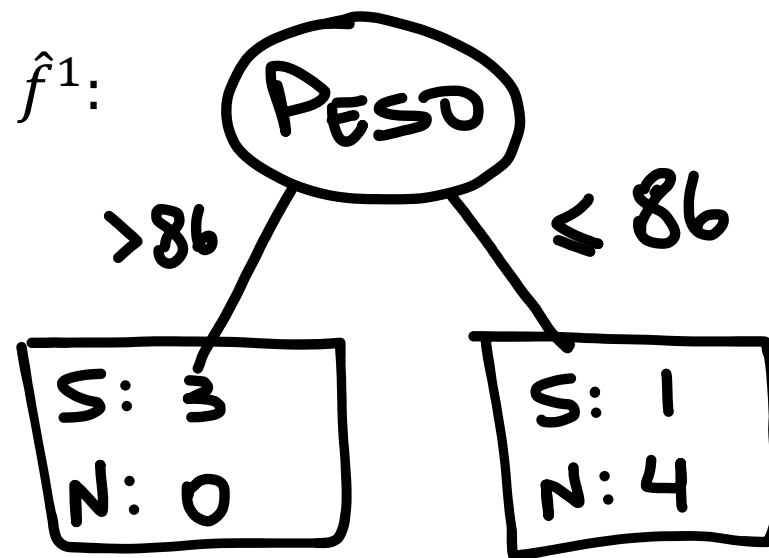
$$\epsilon_1 = \sum \text{ERROS } w_i^{(1)} = 1/8$$

# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
SIM	SIM	100	SIM	1/8
NÃO	SIM	90	SIM	1/8
SIM	NÃO	105	SIM	1/8
SIM	SIM	80	SIM	1/8
NÃO	SIM	70	NÃO	1/8
NÃO	SIM	50	NÃO	1/8
SIM	NÃO	79	NÃO	1/8
SIM	SIM	83	NÃO	1/8

3) Computar  $\alpha^1$  de acordo com  $\epsilon_1$



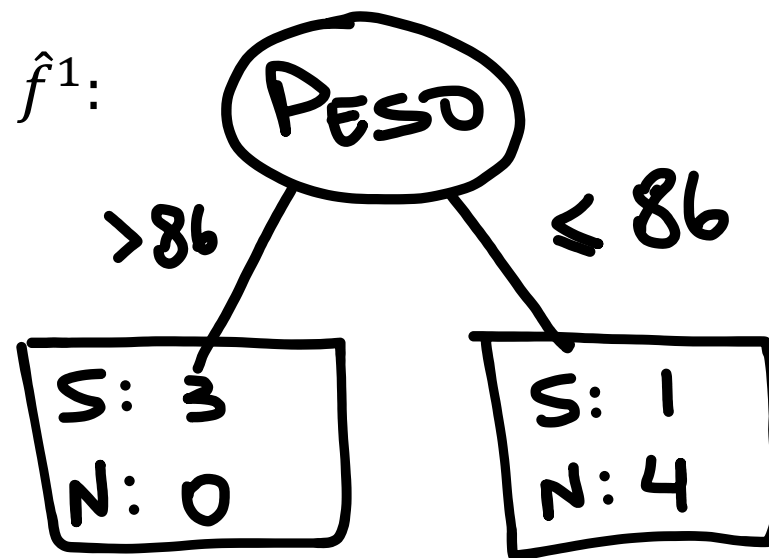
$$\hat{f}(\mathbf{x}) = \text{sinal}(0.97 \times \hat{f}^1(\mathbf{x}) + \alpha^2 \hat{f}^2(\mathbf{x}) + \alpha^3 \hat{f}^3(\mathbf{x}) + \dots)$$

# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
SIM	SIM	100	SIM	1/8
NÃO	SIM	90	SIM	1/8
SIM	NÃO	105	SIM	1/8
SIM	SIM	80	SIM	1/8
NÃO	SIM	70	NÃO	1/8
NÃO	SIM	50	NÃO	1/8
SIM	NÃO	79	NÃO	1/8
SIM	SIM	83	NÃO	1/8

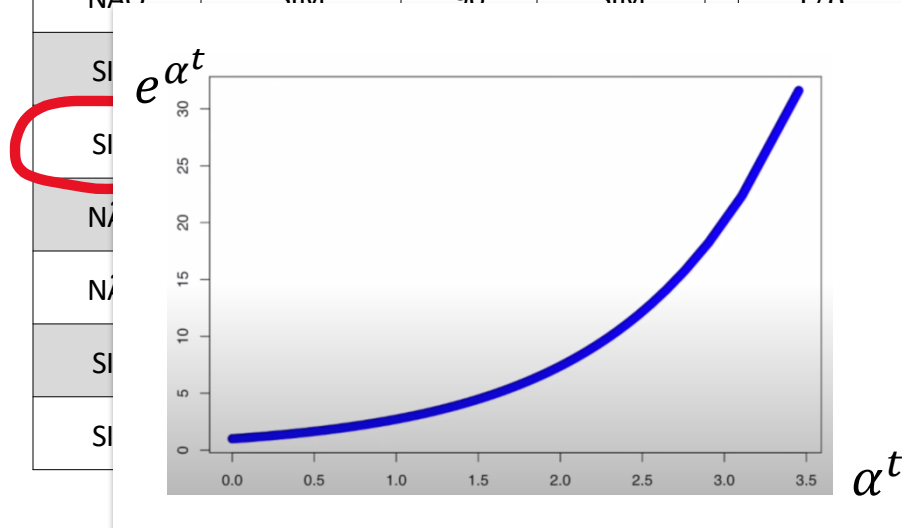
4) Computar  $w_{t+1}^{(i)}$  de acordo com  $\alpha^t$



# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
SIM	SIM	100	SIM	1/8
NÃO	SIM	90	SIM	1/8
SIM	NÃO	10	NÃO	1/8
SIM	NÃO	10	SIM	1/8
NÃO	NÃO	10	NÃO	1/8
NÃO	NÃO	10	SIM	1/8



4) Computar  $w_{t+1}^{(i)}$  de acordo com  $\alpha^t$

se o objeto foi incorretamente classificado por  $\hat{f}^t$ , podemos aumentar seu peso de acordo com o fator de aumento  $e^{\alpha^t}$

$$w_{t+1}^{(i)} = \frac{w_t^{(i)}}{Z} e^{\alpha^t}$$

$$= \frac{1/8}{Z} e^{0.97}$$

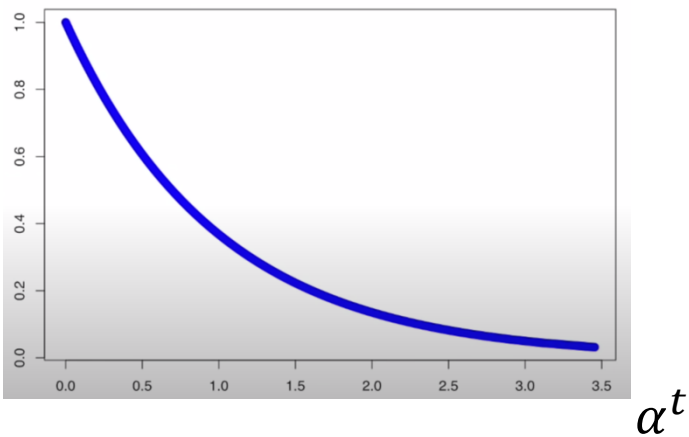
$$= \frac{1/8}{Z} \times 2.64 = \frac{0.33}{Z}$$

# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
SIM	SIM	100	SIM	1/8
NÃO	SIM	90	SIM	1/8

$e^{-\alpha^t}$



4) Computar  $w_{t+1}^{(i)}$  de acordo com  $\alpha^t$

se o objeto foi corretamente classificado por  $\hat{f}^t$ , podemos reduzir seu peso de acordo com o fator de redução  $e^{-\alpha^t}$

$$w_{t+1}^{(i)} = \frac{w_t^{(i)}}{Z} e^{-\alpha^t}$$

$$= \frac{1/8}{Z} e^{-0.97}$$

$$= \frac{1/8}{Z} \times 0.38 = \frac{0.05}{Z}$$

# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca
SIM	SIM	100	SIM
NÃO	SIM	90	SIM
SIM	NÃO	105	SIM
SIM	SIM	80	SIM
NÃO	SIM	70	NÃO
NÃO	SIM	50	NÃO
SIM	NÃO	79	NÃO
SIM	SIM	83	NÃO

Peso da Instância	Novos Pesos
1/8	0.05/Z
1/8	0.05/Z
1/8	0.05/Z
1/8	0.33/Z
1/8	0.05/Z
1/8	0.05/Z
1/8	0.05/Z
1/8	0.05/Z

4) Computar  $w_{t+1}^{(i)}$  de acordo com  $\alpha^t$

→  $Z = 0.68$

# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca
SIM	SIM	100	SIM
NÃO	SIM	90	SIM
SIM	NÃO	105	SIM
SIM	SIM	80	SIM
NÃO	SIM	70	NÃO
NÃO	SIM	50	NÃO
SIM	NÃO	79	NÃO
SIM	SIM	83	NÃO

Peso da Instância
1/8
1/8
1/8
1/8
1/8
1/8
1/8
1/8

4) Computar  $w_{t+1}^{(i)}$  de acordo com  $\alpha^t$

Novos Pesos
0.07
0.07
0.07
0.49
0.07
0.07
0.07
0.07

# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
SIM	SIM	100	SIM	0.07
NÃO	SIM	90	SIM	0.07
SIM	NÃO	105	SIM	0.07
SIM	SIM	80	SIM	0.49
NÃO	SIM	70	NÃO	0.07
NÃO	SIM	50	NÃO	0.07
SIM	NÃO	79	NÃO	0.07
SIM	SIM	83	NÃO	0.07

Ajustar o *dataset* via amostragem c/ reposição

Sortear um número no intervalo  $[0,1]$

$[0.0, 0.07)$

$[0.07, 0.14)$

$[0.14, 0.21)$

$[0.21, 0.70)$

...

...

...

...

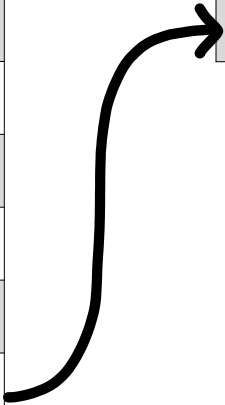


# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
SIM	SIM	100	SIM	0.07
NÃO	SIM	90	SIM	0.07
SIM	NÃO	105	SIM	0.07
SIM	SIM	80	SIM	0.49
NÃO	SIM	70	NÃO	0.07
NÃO	SIM	50	NÃO	0.07
SIM	NÃO	79	NÃO	0.07
SIM	SIM	83	NÃO	0.07

Ajustar o *dataset* via amostragem c/ reposição



Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca
NÃO	SIM	50	NÃO

# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Ajustar o *dataset* via amostragem c/ reposição

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
SIM	SIM	100	SIM	0.07
NÃO	SIM	90	SIM	0.07
SIM	NÃO	105	SIM	0.07
SIM	SIM	80	SIM	0.49
NÃO	SIM	70	NÃO	0.07
NÃO	SIM	50	NÃO	0.07
SIM	NÃO	79	NÃO	0.07
SIM	SIM	83	NÃO	0.07

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca
NÃO	SIM	50	NÃO
SIM	NÃO	105	SIM

# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Ajustar o *dataset* via amostragem c/ reposição

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
SIM	SIM	100	SIM	0.07
NÃO	SIM	90	SIM	0.07
SIM	NÃO	105	SIM	0.07
SIM	SIM	80	SIM	0.49
NÃO	SIM	70	NÃO	0.07
NÃO	SIM	50	NÃO	0.07
SIM	NÃO	79	NÃO	0.07
SIM	SIM	83	NÃO	0.07



Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca
NÃO	SIM	50	NÃO
SIM	NÃO	105	SIM
SIM	SIM	80	SIM

# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Ajustar o *dataset* via amostragem c/ reposição

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
SIM	SIM	100	SIM	0.07
NÃO	SIM	90	SIM	0.07
SIM	NÃO	105	SIM	0.07
SIM	SIM	80	SIM	0.49
NÃO	SIM	70	NÃO	0.07
NÃO	SIM	50	NÃO	0.07
SIM	NÃO	79	NÃO	0.07
SIM	SIM	83	NÃO	0.07

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca
NÃO	SIM	50	NÃO
SIM	NÃO	105	SIM
SIM	SIM	80	SIM
SIM	NÃO	79	NÃO

# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Ajustar o *dataset* via amostragem c/ reposição

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
SIM	SIM	100	SIM	0.07
NÃO	SIM	90	SIM	0.07
SIM	NÃO	105	SIM	0.07
SIM	SIM	80	SIM	0.49
NÃO	SIM	70	NÃO	0.07
NÃO	SIM	50	NÃO	0.07
SIM	NÃO	79	NÃO	0.07
SIM	SIM	83	NÃO	0.07



Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca
NÃO	SIM	50	NÃO
SIM	NÃO	105	SIM
SIM	SIM	80	SIM
SIM	NÃO	79	NÃO
SIM	SIM	80	SIM

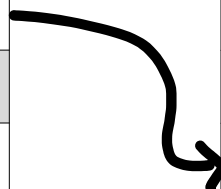
# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Ajustar o *dataset* via amostragem c/ reposição

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
SIM	SIM	100	SIM	0.07
NÃO	SIM	90	SIM	0.07
SIM	NÃO	105	SIM	0.07
SIM	SIM	80	SIM	0.49
NÃO	SIM	70	NÃO	0.07
NÃO	SIM	50	NÃO	0.07
SIM	NÃO	79	NÃO	0.07
SIM	SIM	83	NÃO	0.07

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca
NÃO	SIM	50	NÃO
SIM	NÃO	105	SIM
SIM	SIM	80	SIM
SIM	NÃO	79	NÃO
SIM	SIM	80	SIM
SIM	SIM	80	SIM



# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Ajustar o *dataset* via amostragem c/ reposição

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
SIM	SIM	100	SIM	0.07
NÃO	SIM	90	SIM	0.07
SIM	NÃO	105	SIM	0.07
SIM	SIM	80	SIM	0.49
NÃO	SIM	70	NÃO	0.07
NÃO	SIM	50	NÃO	0.07
SIM	NÃO	79	NÃO	0.07
SIM	SIM	83	NÃO	0.07

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca
NÃO	SIM	50	NÃO
SIM	NÃO	105	SIM
SIM	SIM	80	SIM
SIM	NÃO	79	NÃO
SIM	SIM	80	SIM
SIM	SIM	80	SIM
NÃO	SIM	90	SIM

# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Ajustar o *dataset* via amostragem c/ reposição

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
SIM	SIM	100	SIM	0.07
NÃO	SIM	90	SIM	0.07
SIM	NÃO	105	SIM	0.07
SIM	SIM	80	SIM	0.49
NÃO	SIM	70	NÃO	0.07
NÃO	SIM	50	NÃO	0.07
SIM	NÃO	79	NÃO	0.07
SIM	SIM	83	NÃO	0.07

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca
NÃO	SIM	50	NÃO
SIM	NÃO	105	SIM
SIM	SIM	80	SIM
SIM	NÃO	79	NÃO
SIM	SIM	80	SIM
SIM	SIM	80	SIM
NÃO	SIM	90	SIM
SIM	SIM	100	SIM



# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Ajustar o *dataset* via amostragem c/ reposição

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
SIM	SIM	100	SIM	0.07
NÃO	SIM	90	SIM	0.07
SIM	NÃO	105	SIM	0.07
SIM	SIM	80	SIM	0.49
NÃO	SIM	70	NÃO	0.07
NÃO	SIM	50	NÃO	0.07
SIM	NÃO	79	NÃO	0.07
SIM	SIM	83	NÃO	0.07

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca
NÃO	SIM	50	NÃO
SIM	NÃO	105	SIM
SIM	SIM	80	SIM
SIM	NÃO	79	NÃO
SIM	SIM	80	SIM
SIM	SIM	80	SIM
NÃO	SIM	90	SIM
SIM	SIM	100	SIM

# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca
NÃO	SIM	50	NÃO
SIM	NÃO	105	SIM
SIM	SIM	80	SIM
SIM	NÃO	79	NÃO
SIM	SIM	80	SIM
SIM	SIM	80	SIM
NÃO	SIM	90	SIM
SIM	SIM	100	SIM

# Adaptive Boosting (AdaBoost)

- Um exemplo didático

Dor no Peito	Artérias Bloqueadas	Peso	Doença cardíaca	Peso da Instância
NÃO	SIM	50	NÃO	1/8
SIM	NÃO	105	SIM	1/8
SIM	SIM	80	SIM	1/8
SIM	NÃO	79	NÃO	1/8
SIM	SIM	80	SIM	1/8
SIM	SIM	80	SIM	1/8
NÃO	SIM	90	SIM	1/8
SIM	SIM	100	SIM	1/8

$$1) w_2^{(i)} = \frac{1}{N} \quad \forall i \in X$$

Repetir todo o processo novamente!

# Adaptive Boosting (AdaBoost)

- Existe muita **controvérsia** sobre o nível de overfitting atingido por AdaBoost
  - Há argumentos teóricos (e práticos) para os que defendem que existe overfitting, e os que defendem que não existe
- Como medida preventiva, sugere-se fazer **early-stopping**, controlando o nível de erro do comitê em um conjunto de dados de validação

# Sugestão de Leitura

- Capítulo 17 (17.6 e 17.7) (Alpaydin, 2010)
- Capítulo 8 (Faceli et al. 2011)
- Seção 5.6 (Tan et al., 2006)

# Agradecimentos

- Baseado nos slides gentilmente cedidos pelo Prof. Dr. Rodrigo Coelho Barros (PUCRS)