

# Tópicos Avançados em IA

## Aula #11

### **Agregação de Algoritmos de Aprendizado de Máquina (AM)**

**Professor: Eduardo R. Hruschka**  
**Estagiário PAE: Luiz F. S. Coletta**  
**([luizfsc@icmc.usp.br](mailto:luizfsc@icmc.usp.br))**

# Sumário

1. **Motivação**
2. **Bagging**
3. **Random Forest**
4. **Boosting**
5. **C<sup>3</sup>E**

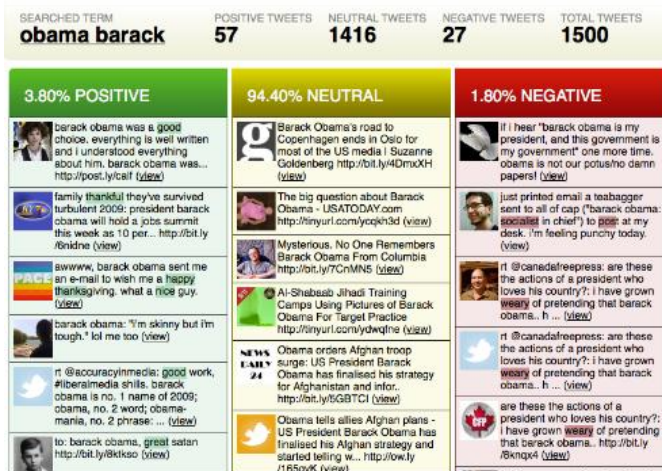
# Motivação

- ▶ **Métodos de Aprendizado de Máquina (AM)**
  - **Classificadores**
    - Árvore de decisão, rede neural, SVM, etc.
  - **Agrupadores**
    - EM, K-Means, Bisecting K-Means, etc.

# Motivação

## ► Aplicações

- Análise de sentimentos;
- Churn prediction (prever cancelamento de serviços);
- Detecção de comportamento em vídeos
  - Anormalidades, monitoramento de idosos, pacientes, etc.
- Identificação de padrões em imagens;
- Sensoriamento remoto;
- Reconhecimento de pessoas (face, íris, etc.).



# Motivação

## ▶ **Uso de algoritmos individualmente**

- Cada modelo assume um conjunto de suposições e, portanto, está sujeito a um “bias”
  - Diferentes algoritmos podem convergir para diferentes soluções podendo falhar sobre determinadas circunstâncias;
  - Dados de treinamento podem ser escassos ou oriundos de uma amostragem deficiente.

# Agregação de Algoritmos de AM

- ▶ **Forma de melhorar a qualidade dos resultados**
  - **Agregação (ensemble) de classificadores**
    - Maior precisão e/ou robustez [[Oza & Tumer, 2008](#)].
  - **Agregação de agrupadores (cluster ensembles)**
    - Conciliar diversas partições com o objetivo de obter uma partição de consenso [[Ghosh & Acharya, 2011](#)].

Oza, Nikunj C., and Kagan Tumer. "Classifier ensembles: Select real-world applications." *Information Fusion* 9.1 (2008): 4-20.

Ghosh, Joydeep, and Ayan Acharya. "Cluster ensembles." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 1.4 (2011): 305-315.

# Agregação de Algoritmos de AM

## ▶ Exemplo didático para classificação

- Um objeto no conjunto de treinamento pode ser visto como uma tupla  $\langle \mathbf{x}_i, C_i \rangle$ ,  $C_i \in \{C_1, C_2, \dots, C_c\}$ ;
- O problema de inferir a classe de novos objetos desconhecidos é chamado de classificação;
  - Árvores de decisão, redes neurais, SVM, K-NN, Naive Bayes, etc...

# Agregação de Algoritmos de AM

- ▶ Dados três classificadores  $h_1$ ,  $h_2$  e  $h_3$  e um novo objeto  $\mathbf{x}_i$  a ser classificado:
  - Se  $h_1$ ,  $h_2$  e  $h_3$  são idênticos, então quando  $h_1(\mathbf{x}_i)$  erra,  $h_2(\mathbf{x}_i)$  e  $h_3(\mathbf{x}_i)$  também erram...
  - Se são diferentes, quando  $h_1(\mathbf{x}_i)$  está errado,  $h_2(\mathbf{x}_i)$  e  $h_3(\mathbf{x}_i)$  podem estar certos!
    - Usando-se voto majoritário  $\mathbf{x}_i$  tem a chance de ser corretamente classificado!



# Agregação de Algoritmos de AM

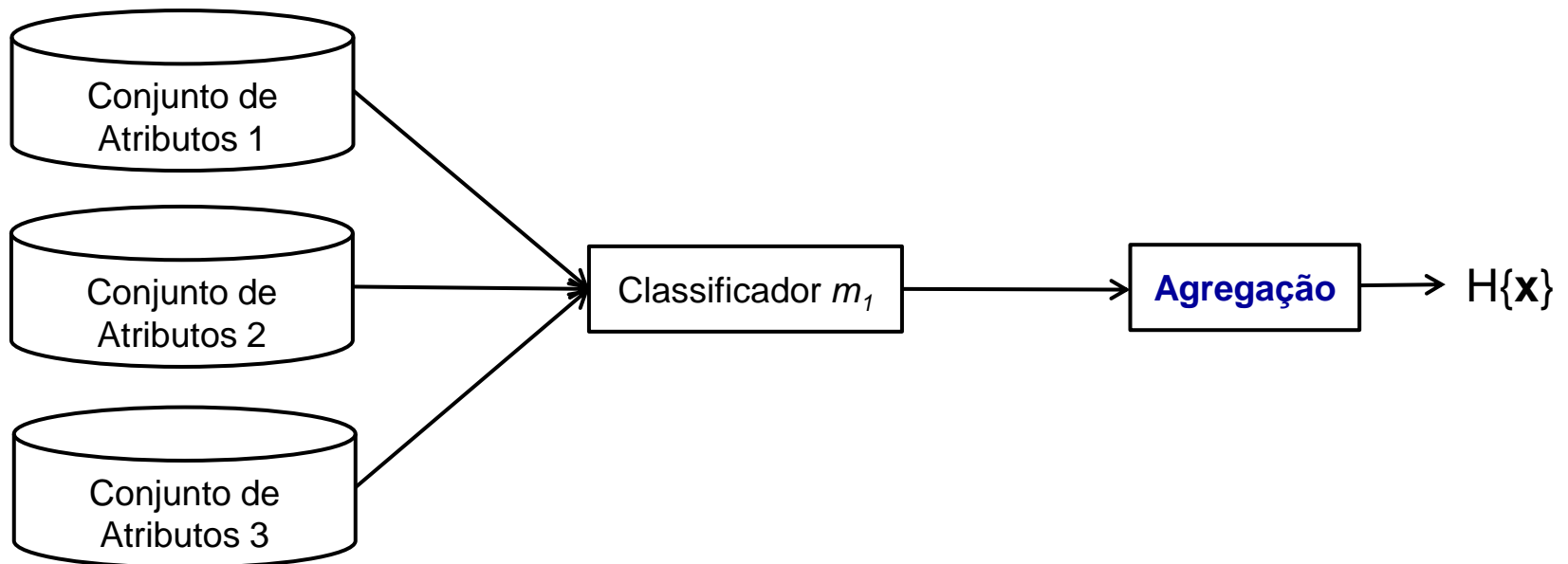
- ▶ Dados três classificadores  $h_1$ ,  $h_2$  e  $h_3$  e um novo objeto  $\mathbf{x}_i$  a ser classificado:
  - Se  $h_1$ ,  $h_2$  e  $h_3$  são idênticos, então quando  $h_1(\mathbf{x}_i)$  erra,  $h_2(\mathbf{x}_i)$  e  $h_3(\mathbf{x}_i)$  também erram...
  - Se são diferentes, quando  $h_1(\mathbf{x}_i)$  está errado,  $h_2(\mathbf{x}_i)$  e  $h_3(\mathbf{x}_i)$  podem estar certos!
    - Usando-se voto majoritário  $\mathbf{x}_i$  tem a chance de ser corretamente classificado!

**Classificadores base diversificados!!!**

# Agregação de Algoritmos de AM

## ▶ Arquitetura

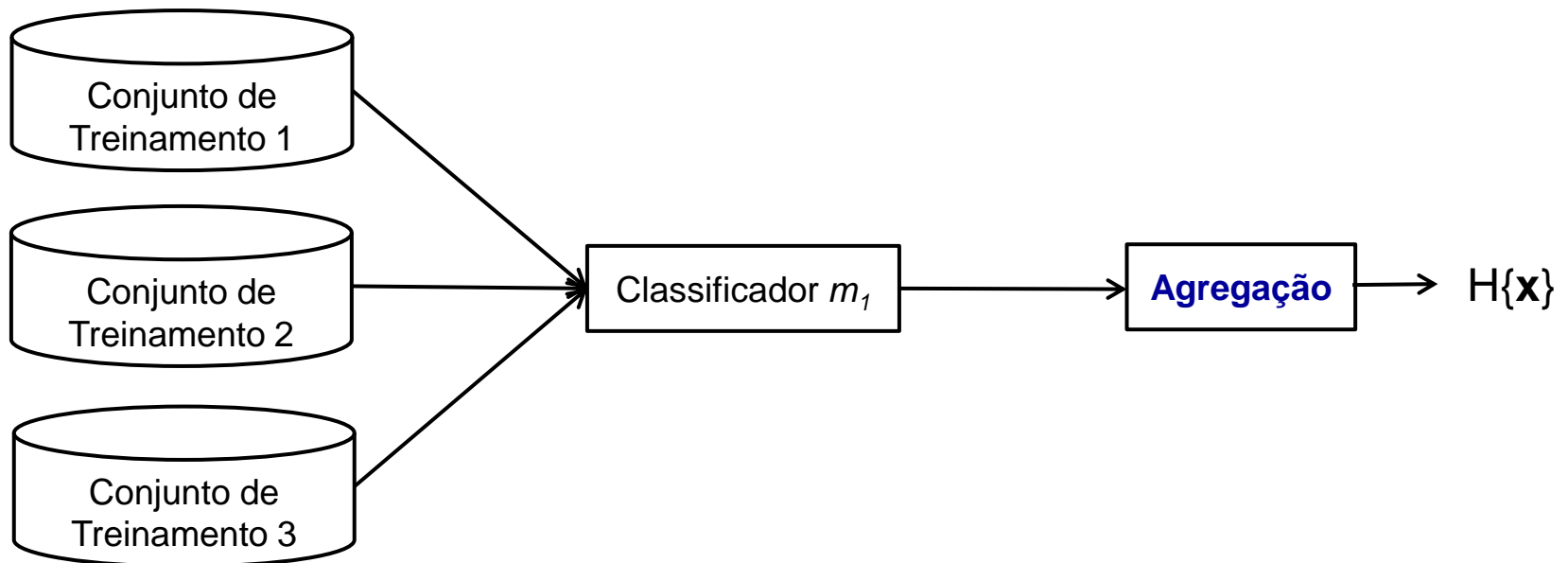
- Diferentes espaços de atributos, mesmo classificador



# Agregação de Algoritmos de AM

## ► Arquitetura

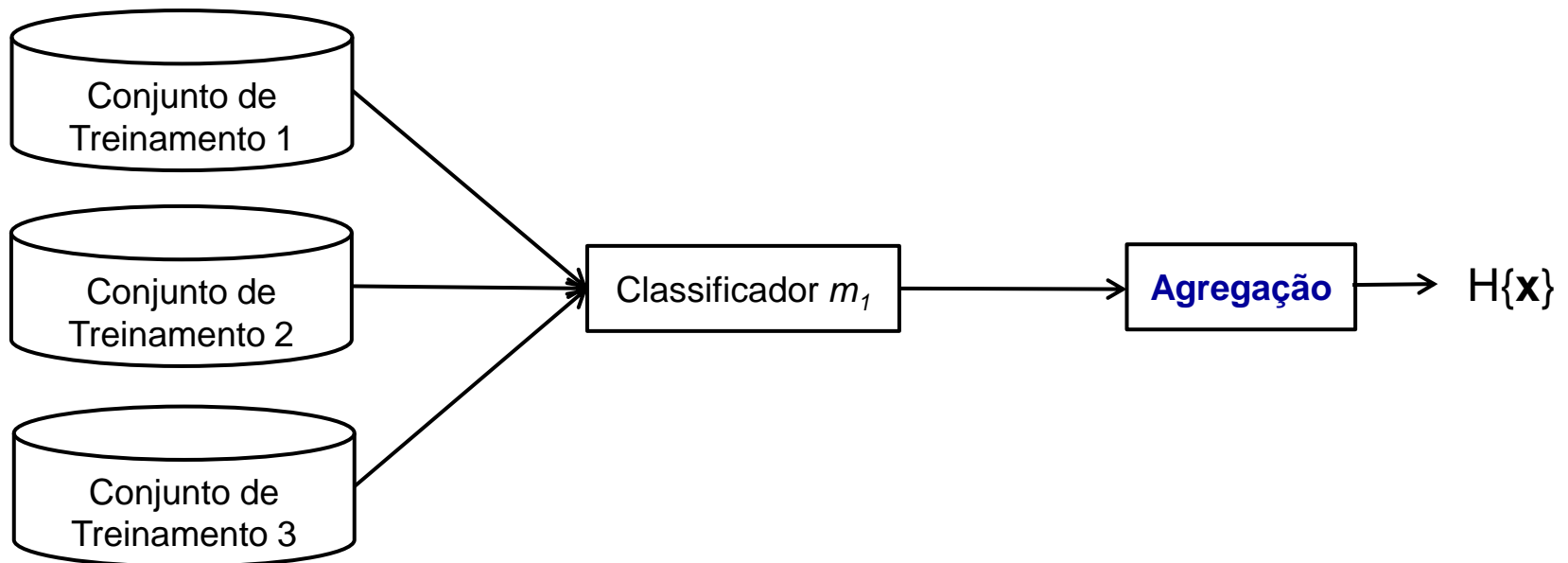
- Diferentes conjuntos de treinamento, mesmo classificador



# Agregação de Algoritmos de AM

## ► Arquitetura

- Diferentes conjuntos de treinamento, mesmo classificador



Bagging, Random Forests, Boosting...

# Bagging [Breiman, 1996]

- ▶ Cada classificador é treinado usando-se uma distribuição específica dos dados
  - Amostragem baseada em Bootstrap:
    - Mesmo tamanho do conjunto original;
    - Com reposição a partir de distribuição uniforme:
      - Podem haver instâncias repetidas, outras podem não aparecer;
      - 63% dos dados originais aproximadamente!

**Breiman, Leo. "Bagging predictors." *Machine learning* 24.2 (1996): 123-140.**

# Bagging [Breiman, 1996]

- ▶ Uso de classificadores instáveis
  - Pequenas mudanças no conjunto de treinamento levam a grandes mudanças no resultado de classificação
    - Instáveis: Redes Neurais, Árvores de decisão;
    - Estável: K-NN.
- ▶ Naturalmente paralelizável;
- ▶ Agregação por voto majoritário;
- ▶ Robusto a ruídos nos dados.

# Random Forests [Breiman, 2001]

## ▶ Extensão de Bagging

- Amostragem baseada em bootstrap
- Cada Árvore de Decisão:
  - É construída a partir de amostragem das instâncias do conjunto de treinamento e também de um subconjunto de atributos!
- Tem tido grande destaque na área de ensembles!

# Boosting [Schapire, 1990]

- ▶ Classificadores fracos são capazes de se tornarem classificadores fortes?
- ▶ Induz sequencialmente um conjunto de classificadores
  - O classificador corrente depende dos anteriores tendo maior foco no erro destes últimos
    - Distribuição do conjunto de treinamento é modificada
      - Instâncias incorretamente previstas anteriormente são escolhidas com maior frequência / ponderadas com maior peso.



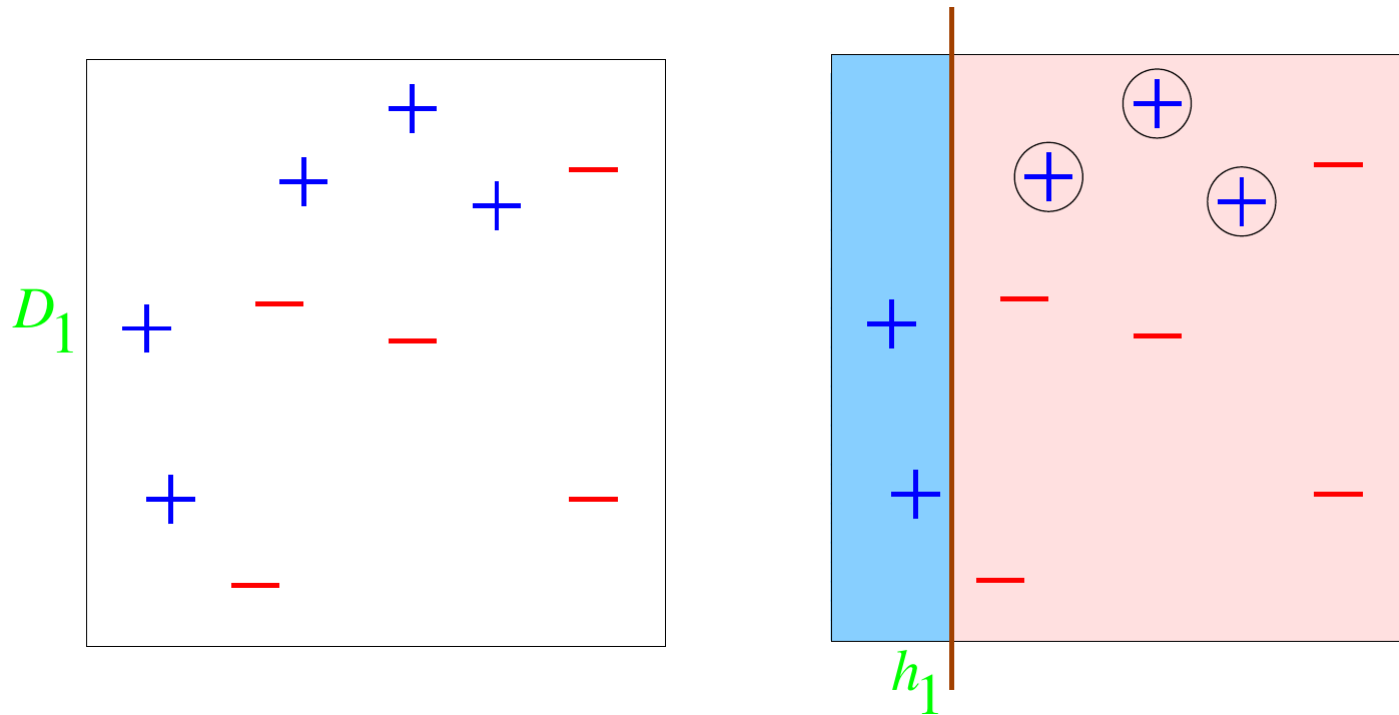
# Boosting [Schapire, 1990]

- ▶ Classificadores fracos são capazes de se tornarem classificadores fortes?
- ▶ Induz sequencialmente um conjunto de classificadores
  - O classificador corrente depende dos anteriores tendo maior foco no erro destes últimos
    - Distribuição do conjunto de treinamento é modificada
      - Instâncias incorretamente preditas anteriormente são escolhidas com maior frequência / ponderadas com maior peso.

Computação distribuída?

# Boosting

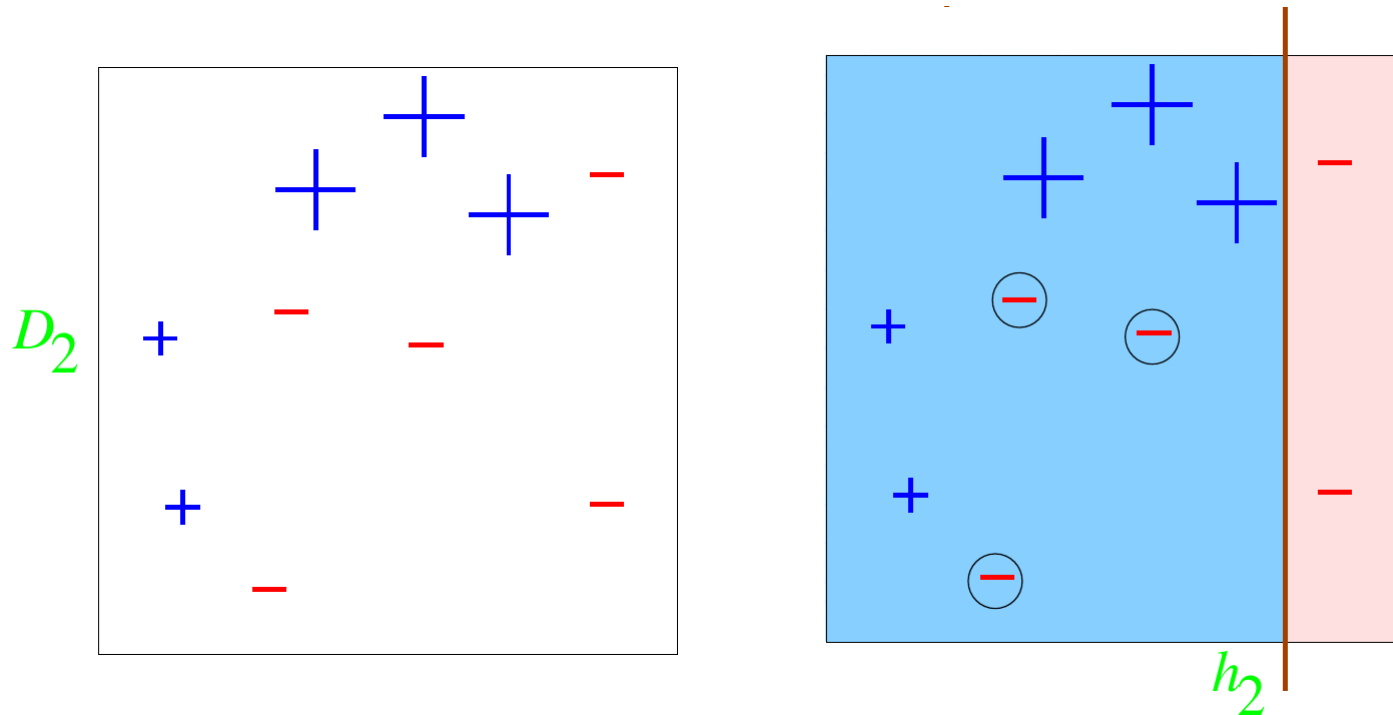
## ▶ Exemplo didático – passo 1



De “*A Tutorial on Boosting*” de Y. Freund & R. Schapire

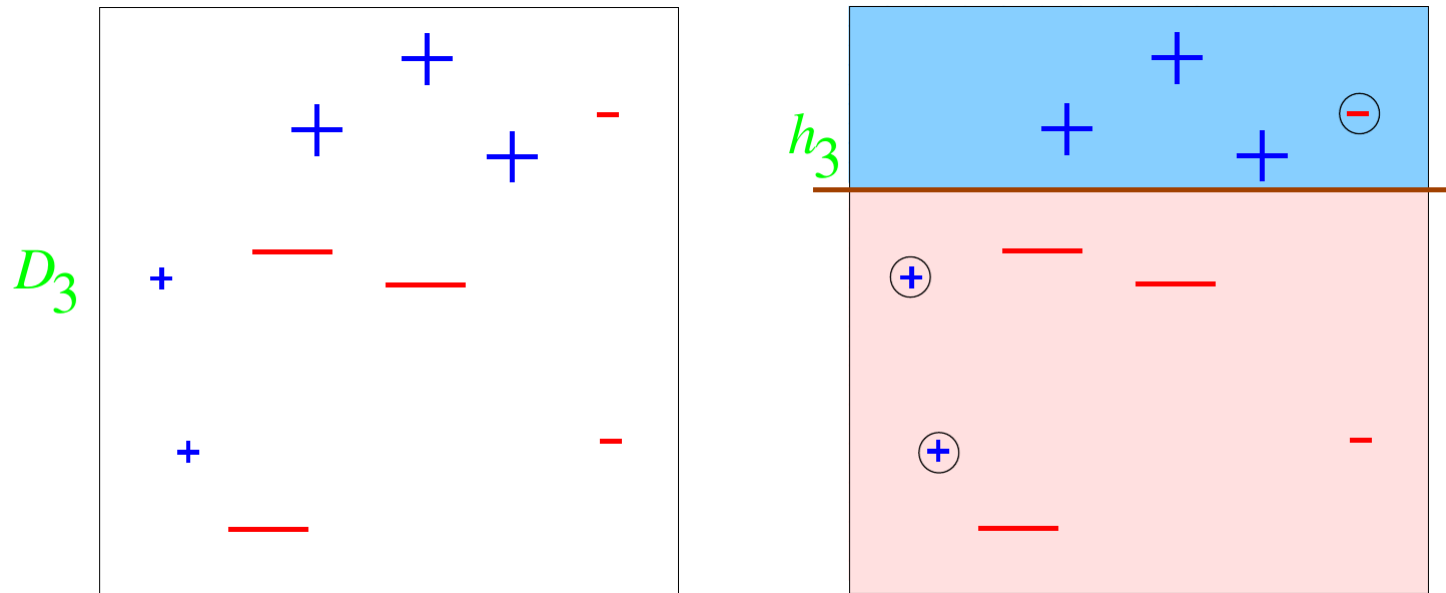
# Boosting

## ▶ Exemplo didático – passo 2



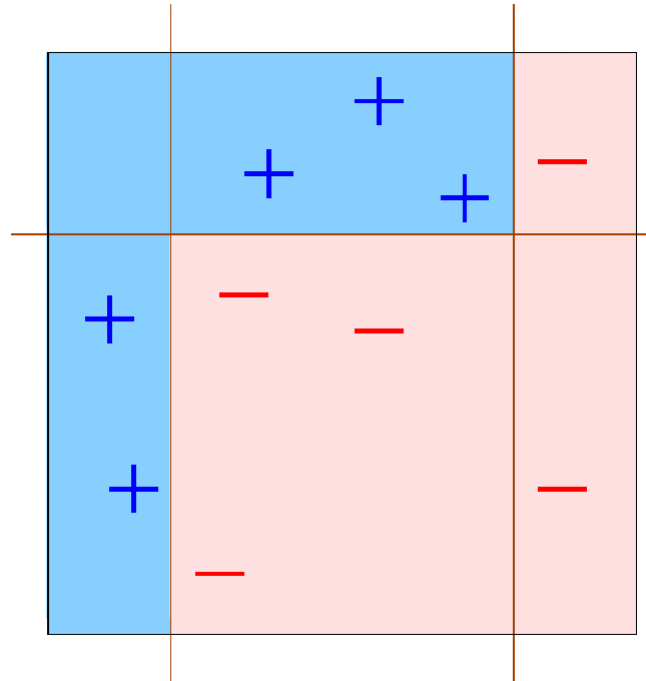
# Boosting

## ▶ Exemplo didático – passo 3



# Boosting

## ▶ Exemplo didático – hipótese final

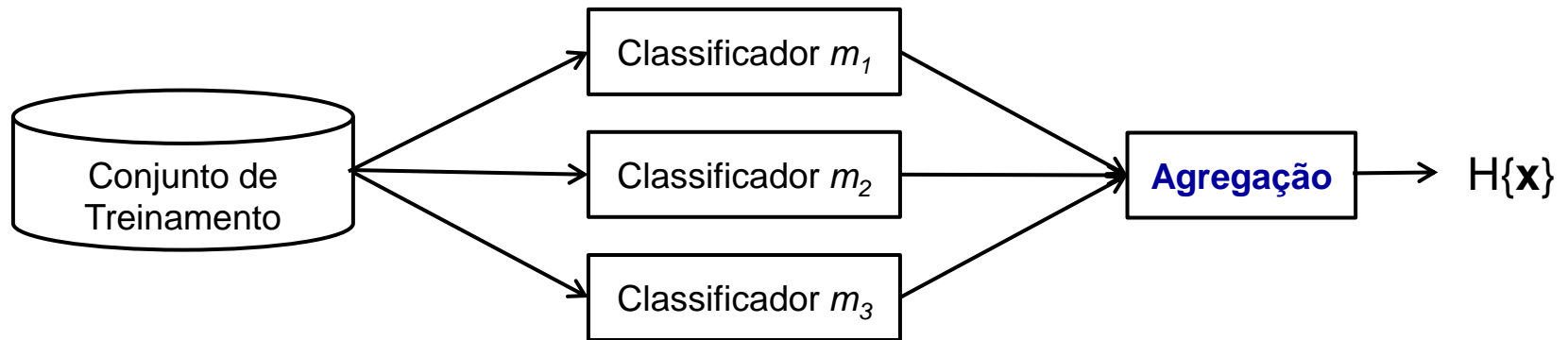


- Análises empíricas sugerem sensibilidade a ruídos!

# Agregação de Algoritmos de AM

## ▶ Arquitetura

- Mesmo conjunto de treinamento, diferentes classificadores



# Agregação de Classificadores e Agrupadores

- ▶ Modelos não supervisionados podem fornecer uma série de restrições complementares para classificar novos dados
  - **Premissa subjacente:** objetos similares no conjunto-alvo provavelmente compartilham do mesmo rótulo de classe.

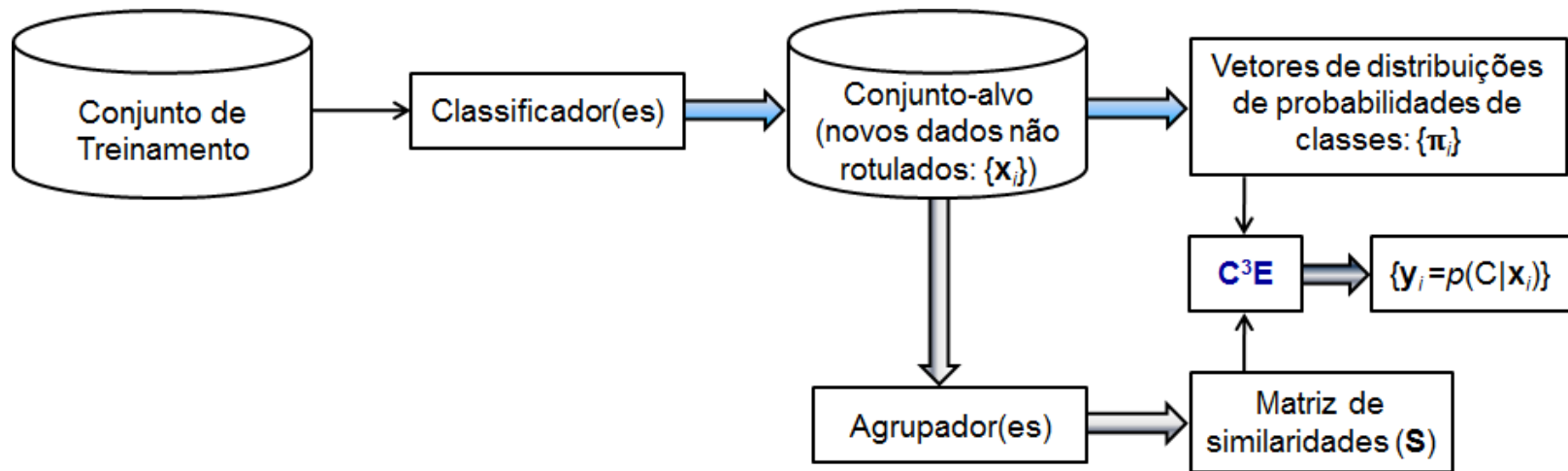
# Agregação de Classificadores e Agrupadores

## ▶ **Benefícios**

- Aumentar a capacidade de generalização e a robustez (especialmente em aprendizado semi-supervisionado);
- Auxiliar na identificação de mudanças de conceito e novas classes não previstas.

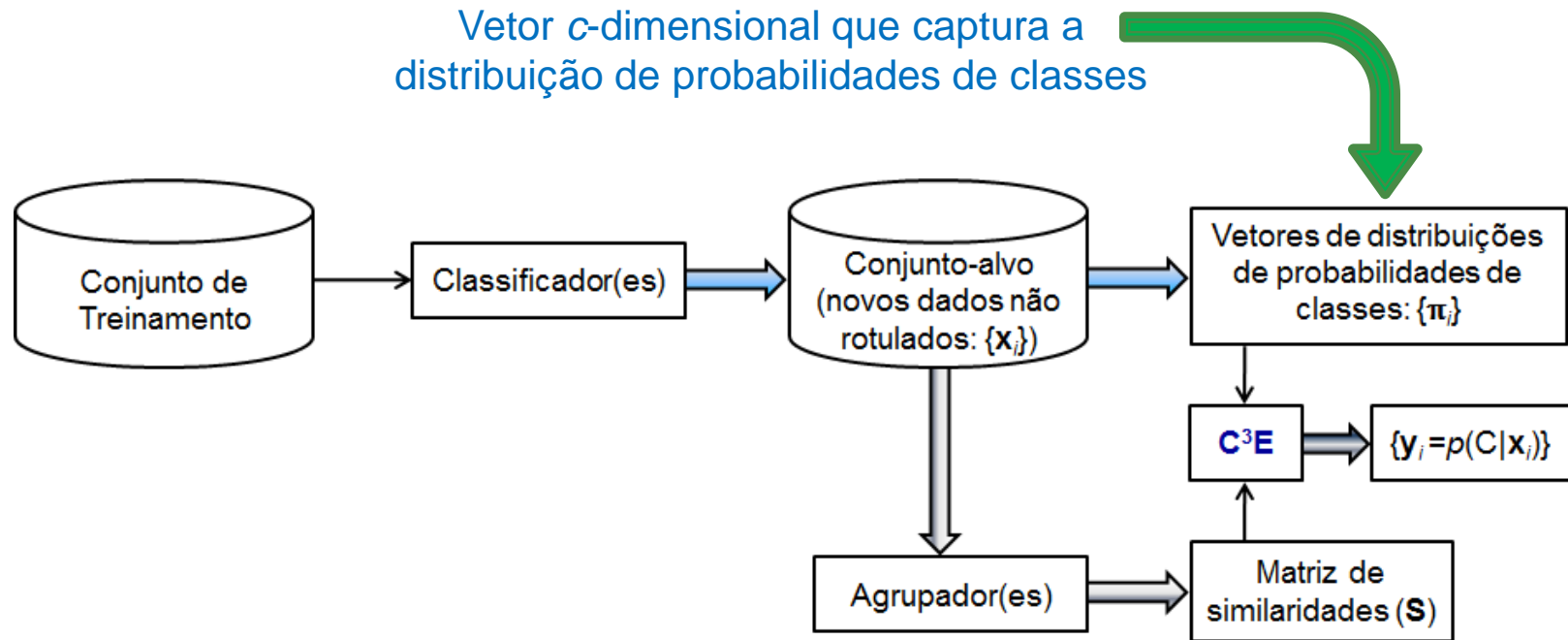


# Aggregação de Classificadores e Agrupadores



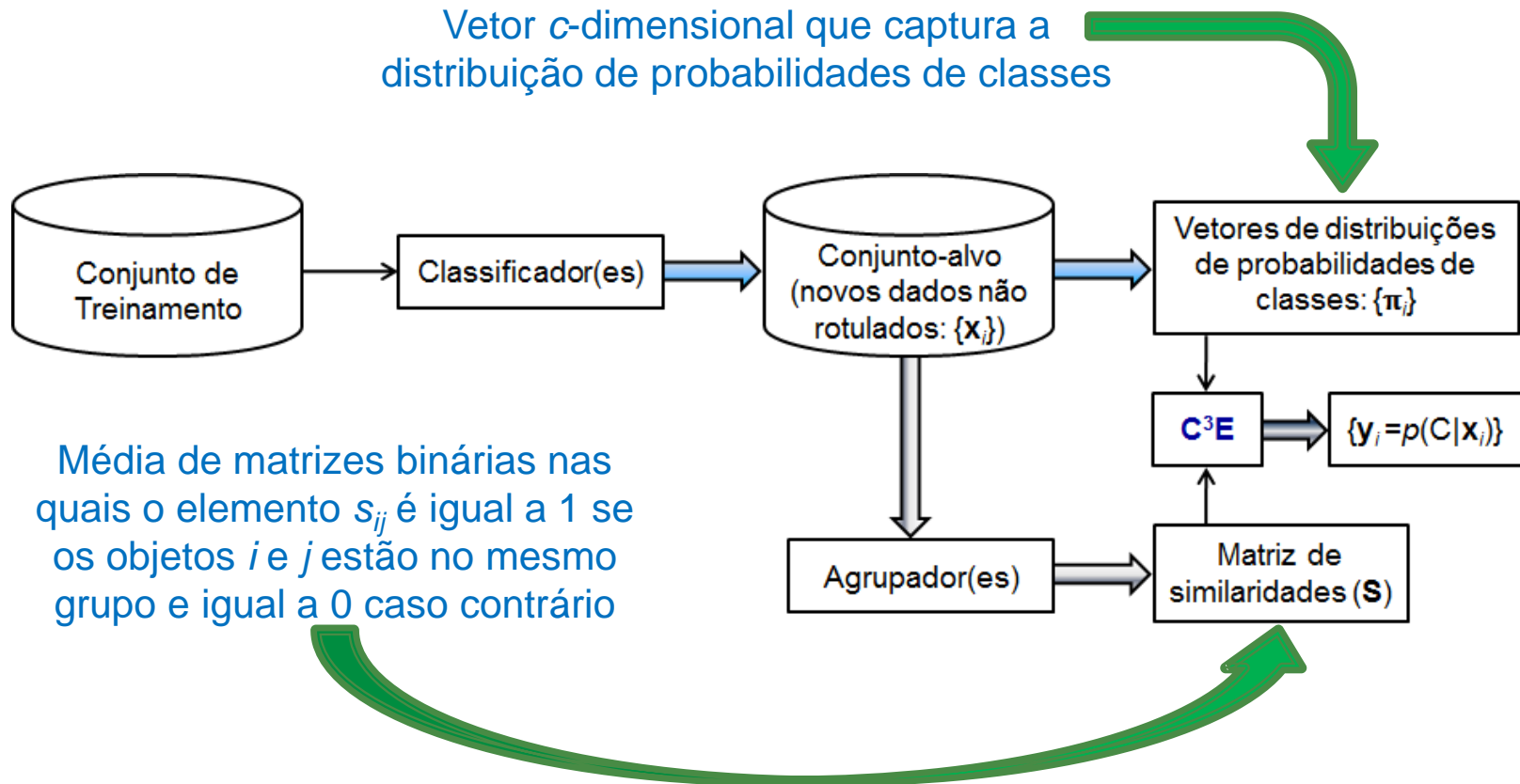
**Acharya, A., Hruschka, E. R., Ghosh, J., and Acharyya, S. (2011). C3e: A framework for combining ensembles of classifiers and clusterers. In Multiple Classifier Systems, pages 269-278. LNCS Vol. 6713, Springer.**

# Agregação de Classificadores e Agrupadores



Acharya, A., Hruschka, E. R., Ghosh, J., and Acharyya, S. (2011). C3e: A framework for combining ensembles of classifiers and clusterers. In *Multiple Classifier Systems*, pages 269-278. LNCS Vol. 6713, Springer.

# Aggregação de Classificadores e Agrupadores



Acharya, A., Hruschka, E. R., Ghosh, J., and Acharyya, S. (2011). C3e: A framework for combining ensembles of classifiers and clusterers. In *Multiple Classifier Systems*, pages 269-278. LNCS Vol. 6713, Springer.

# Agregação de Classificadores e Agrupadores

## ► Problema de otimização (minimizar função $J$ )

$$J_{PQ} = \frac{1}{2} \sum_{i \in \mathcal{X}} \|\mathbf{y}_i - \boldsymbol{\pi}_i\|^2 + \alpha \frac{1}{2} \sum_{(i,j) \in \mathcal{X}} s_{ij} \|\mathbf{y}_i - \mathbf{y}_j\|^2$$

- **1º termo:** captura as dissimilaridades entre as distribuições de probabilidades de classes iniciais ( $\boldsymbol{\pi}_i$ ) e finais ( $\mathbf{y}_i$ );
- **2º termo:** captura a dissimilaridade de todos os possíveis pares ( $\mathbf{y}_i, \mathbf{y}_j$ ) ponderados por  $s_{ij} \in \mathbf{S}$ ;
- $\alpha$ : controla a importância relativa entre agrupadores e classificadores;
- **Processo iterativo de computação da seguinte equação:**

$$\mathbf{y}_i = \frac{\boldsymbol{\pi}_i + \alpha' \sum_{j \neq i} s_{ij} \mathbf{y}_j}{1 + \alpha' \sum_{j \neq i} s_{ij}},$$

# Agregação de Classificadores e Agrupadores

## ► Resultados de classificação

Taxa de classificação correta média (%) – desvio entre parênteses		
Problema	NB + AD	C <sup>3</sup> E
Wisconsin (câncer de mama)	95,60 (2,8)	96,48 (1,2)
Yeast (localização de proteínas)	95,61 (3,2)	97,56 (1,7)
Spam (detecção de spams)	81,20 (1,2)	92,81 (1,1)
Wine Red (qualidade de vinhos)	55,91 (4,5)	57,48 (9,1)
Blood (doação de sangue)	75,80 (1,9)	76,60 (1,7)
Seeds (classificação de trigo)	88,57 (7,4)	91,90 (3,2)

**COLETTA, L. F. S. ; HRUSCHKA, E. R. ; ACHARYA, A. ; GHOSH, J. . Towards the Use of Metaheuristics for Optimizing the Combination of Classifier and Cluster Ensembles. Proceedings of the 1st BRICS Countries Congress on Computational Intelligence (BRICS-CCI), 2013. p. 1-6.**

# Agregação de Classificadores e Agrupadores

## ► Resultados de classificação

