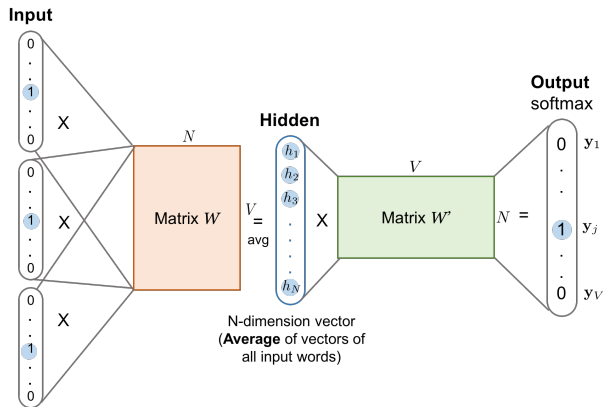


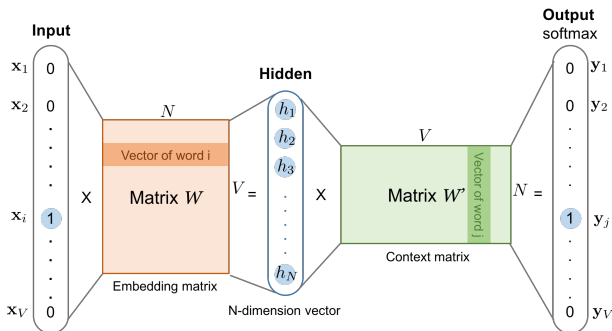
# ROTEIRO

AUXILIAR

# Embedding - CBoW

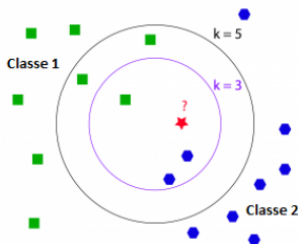


# Embedding - Skip-gram



## CLASSIFICADORES (*KNN (K-Nearest Neighbor)*)

Classifica novas amostras de acordo com as  $K$  (5) amostras do conjunto de treinamento mais próximas a essas novas amostras. O KNN usa uma medida de distância (Euclidiana) para definir a semelhança (proximidade) de uma amostra com outra.<sup>1</sup>

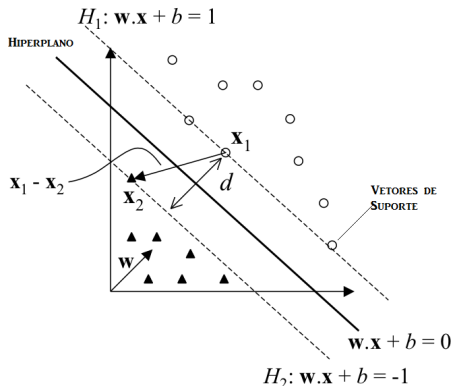


---

<sup>1</sup>DUDA, R. O.; HART, P. E.; STORK, D. G. Pattern classification and scene analysis 2nd ed. ed: Wiley Interscience, 1995.

# CLASSIFICADORES (*SVM (Support Vector Machines)*)

Se baseia na margem de separação das classes, onde o objetivo do treinamento é encontrar um hiperplano separador ótimo, aquele em que a distância de separação entre as classes é máxima - hiperplano de margem máxima <sup>2</sup>. *Kernel*: RBF.



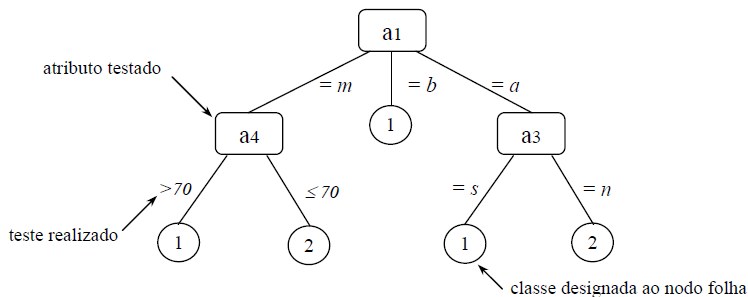
<sup>2</sup>DUTRA, L. P. Detecção das doenças olho de boi e mancha de sarna em maçãs utilizando máquina de vetores de suporte. 2017.

## CLASSIFICADORES (*DT (Decision Tree)*)

É constituído essencialmente uma série de decisões *if-else*. Os dados vão sendo particionados em subconjuntos e alguma medida de pureza (gini) dos subconjuntos vai sendo avaliada para decidir quando parar.<sup>3</sup>

Algoritmo de construção utilizado:

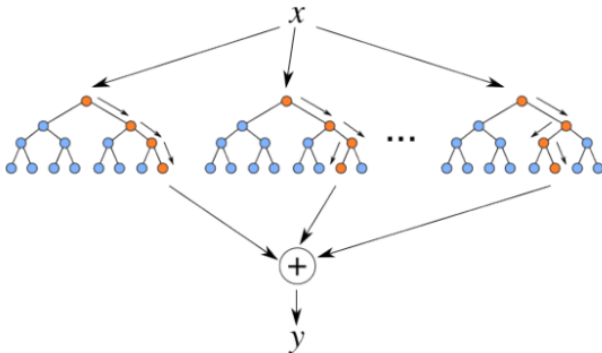
- ▶ CART (*Classification and Regression Trees*)



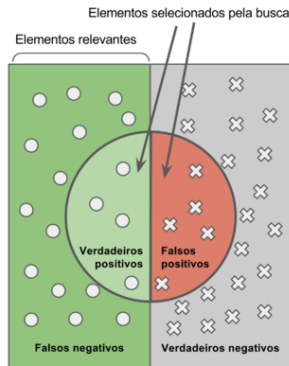
<sup>3</sup>CARACIOLO, M. P. Introdução a AD para classificação e MD. 2009.

## CLASSIFICADORES (*RF (Random Forest)*)

Tem o objetivo de efetuar a criação de várias árvores de decisão usando um subconjunto de atributos selecionados aleatoriamente a partir do conjunto original, contendo todos os atributos e que estes possuem um tipo de amostragem chamado de *bootstrap*, a qual é do tipo com reposição, possibilitando assim melhor análise dos dados.



# Precision e Recall



$$\text{Precisão} = \frac{\text{Verdadeiros positivos}}{\text{Verdadeiros positivos} + \text{Falsos positivos}}$$

"Quanto elementos selecionados são relevantes?"

$$\text{Revocação} = \frac{\text{Verdadeiros positivos}}{\text{Verdadeiros positivos} + \text{Falsos negativos}}$$

"Quanto elementos relevantes foram selecionados?"