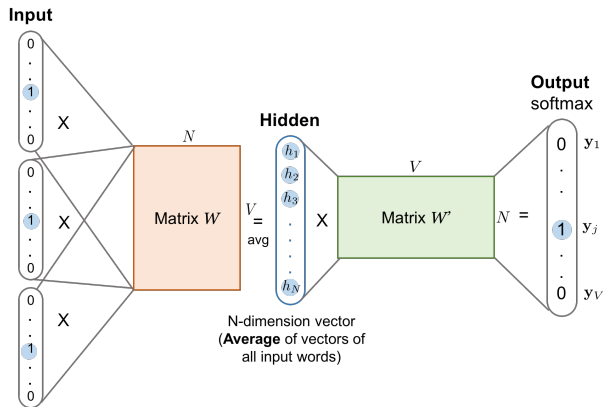


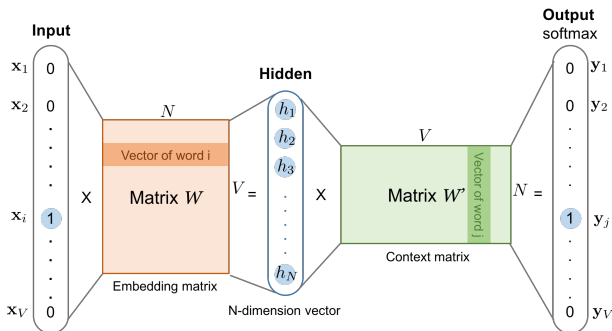
ROTEIRO

AUXILIAR

Embedding - CBoW

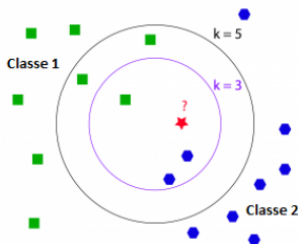


Embedding - Skip-gram



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

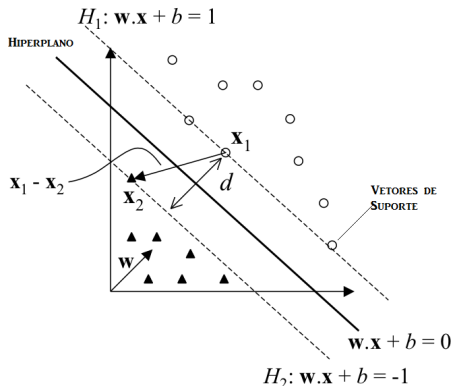
Classifica novas amostras de acordo com as K (5) amostras do conjunto de treinamento mais próximas a essas novas amostras. O KNN usa uma medida de distância (Euclidiana) para definir a semelhança (proximidade) de uma amostra com outra.¹



¹DUDA, 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 ². *Kernel*: RBF.



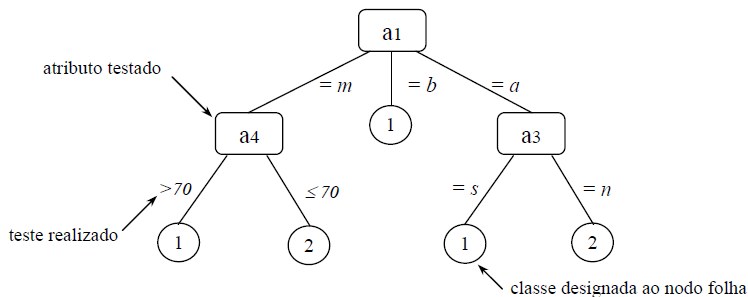
²DUTRA, L. P. Detecção das doenças olho de boi e mancha de sarna em maçãs utilizando máquina de vetores de suporte. 2017.

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.³

Algoritmo de construção utilizado:

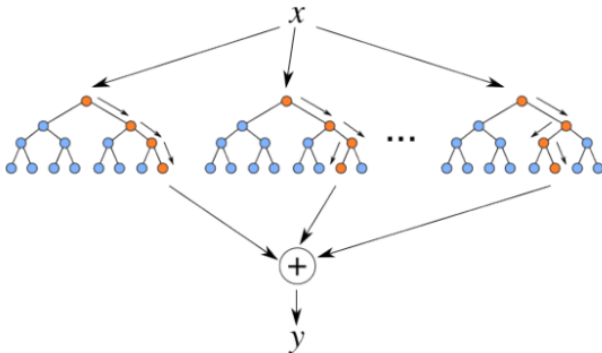
- ▶ CART (*Classification and Regression Trees*)



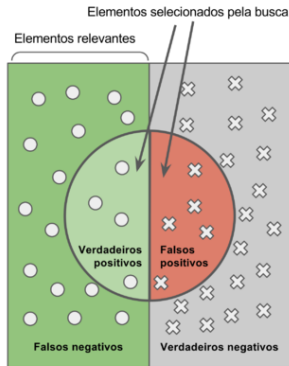
³CARACIOLO, M. P. Introdução a AD para classificação e MD. 2009.

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?"

TREINAMENTOS E TESTES

Combinação	Z5News (M)	Z5News (O)	Z5BrasilNews (M)	Z5BrasilNews (O)	Z12News (M)	Z12News (O)
RF+BoW	0,7818	0,7273	0,7651	0,6745	0,5671	0,4699
SVM(RBF)+FT (E2V-IDF)	0,7846	0,8309	0,7501	0,6888	0,5828	0,6595
SVM(RBF)+W2V (E2V-IDF)	0,7838	0,8047	0,7111	0,6104	0,5954	0,6614

M - Treinado e testado na mesma coleção

O - Treinado em uma coleção e testado em outras coleções

Observação

A Z5NewsBrasil foi testada apenas com outra fonte; As demais foram testadas com outras fonte, mas tem também notícias da própria fonte (ainda não vistas).