Supervised Classification

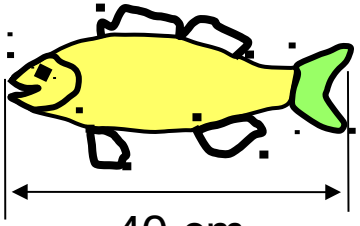
Sumary

- Introduction
- K-NN
- Neural Networks
- Support Vector Machines (SVM)

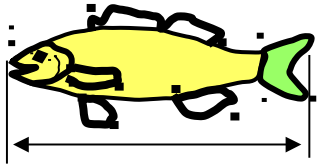
Un Ejemplo

Fish Identification: Only two measurements are available

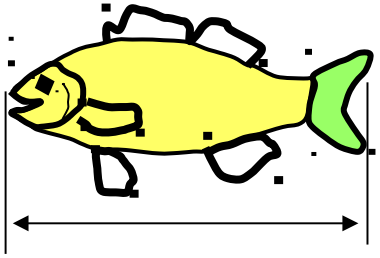
Class A



1 kg

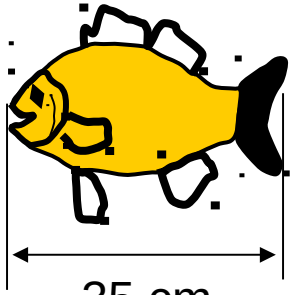


750 gr

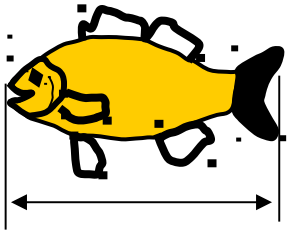


1,25 Kg

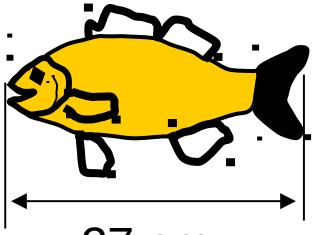
Class B



1,2 kg



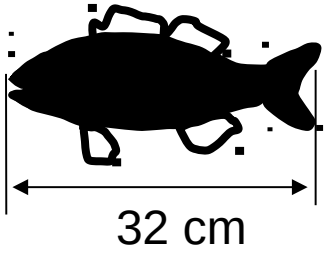
850 gr



1 Kg

Example

Which one ?



Peso: 900 gr

Example

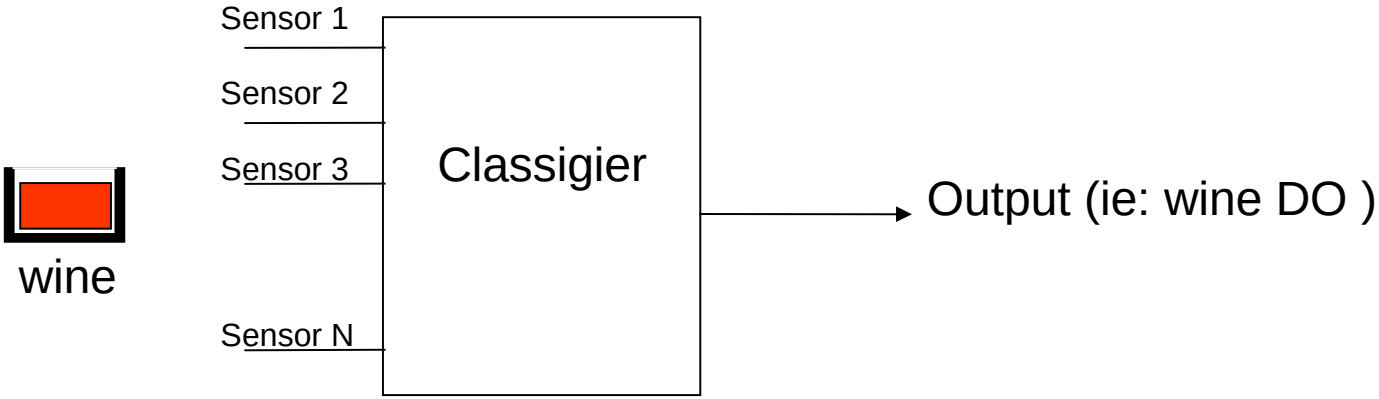
Pattern = Fish

Features/Characteristics = Measurements

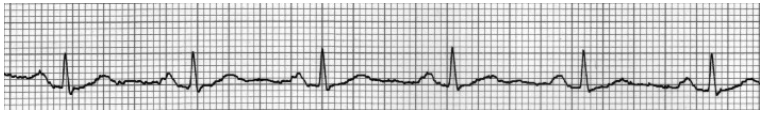
Training Set = Pattern with their labels

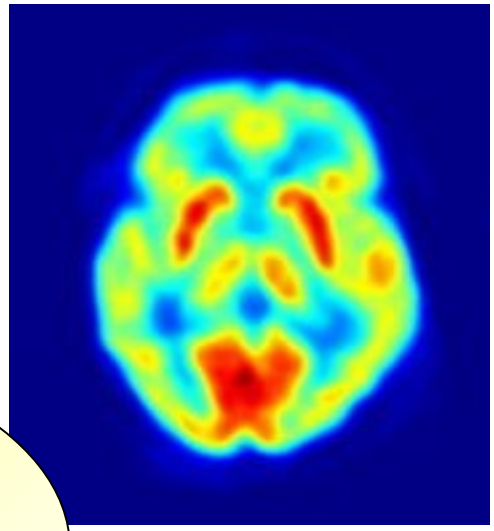
Decision Rule = Rule to decide how to classify a new sample.

Otro Ejemplo

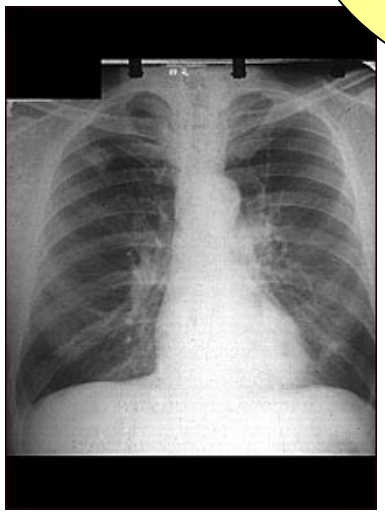


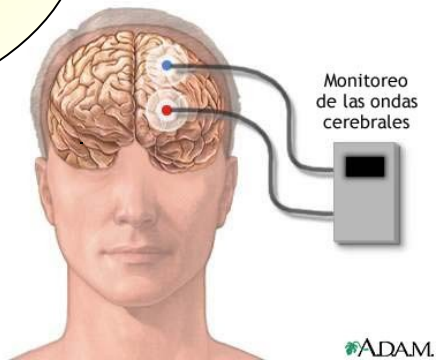
Campos de Aplicación del Reconocimiento de Patrones





Medicine

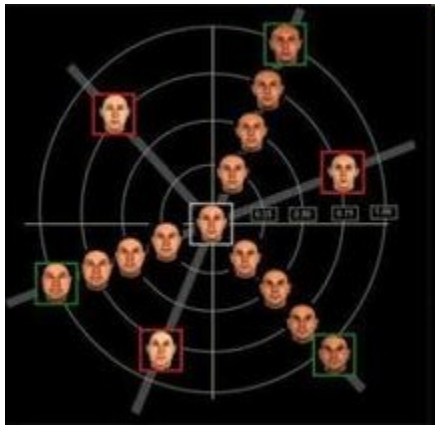




Campos de Aplicación del Reconocimiento de Patrones

Chemistry

Applications



Face Recognition

Surveillance

License Plate recognition

Trajectory recognition

Campos de Aplicación del Reconocimiento de Patrones

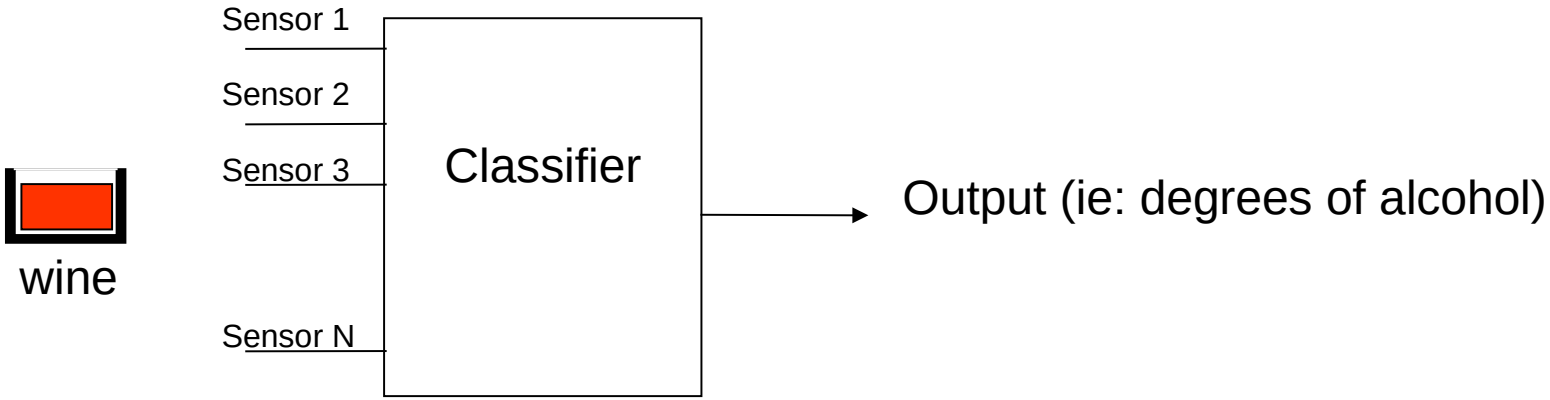
Energy

Binary classification: Two Classes

Clasification Vs Regression

Regression: For a given training set, find the transformation to get a real number $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5, ..., \mathbf{x}_N\}, x \hat{I}$ ¡ ^{d}

Example:

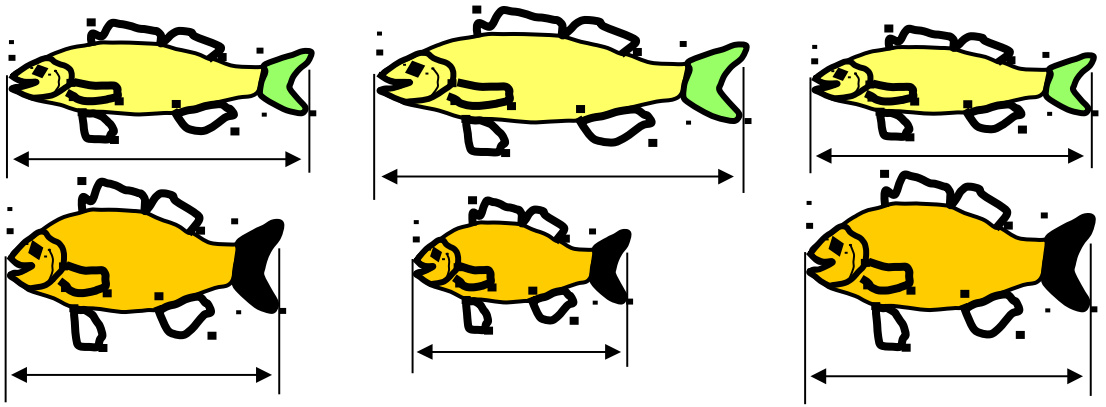


Supervised Learning vs non supervised

Surpervised learning:

The labels (classes) of the training samples are known.

Conjunto de entrenamiento:

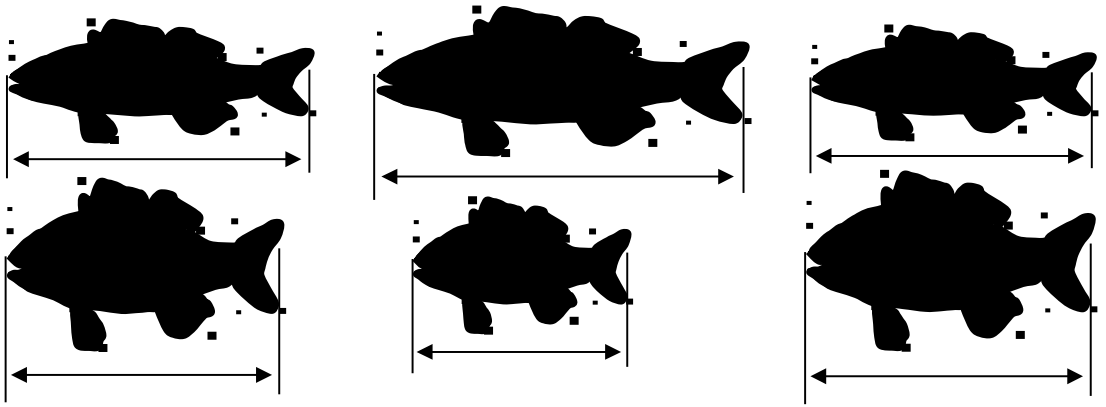


Supervised Learning vs non supervised

Non supervised learning:

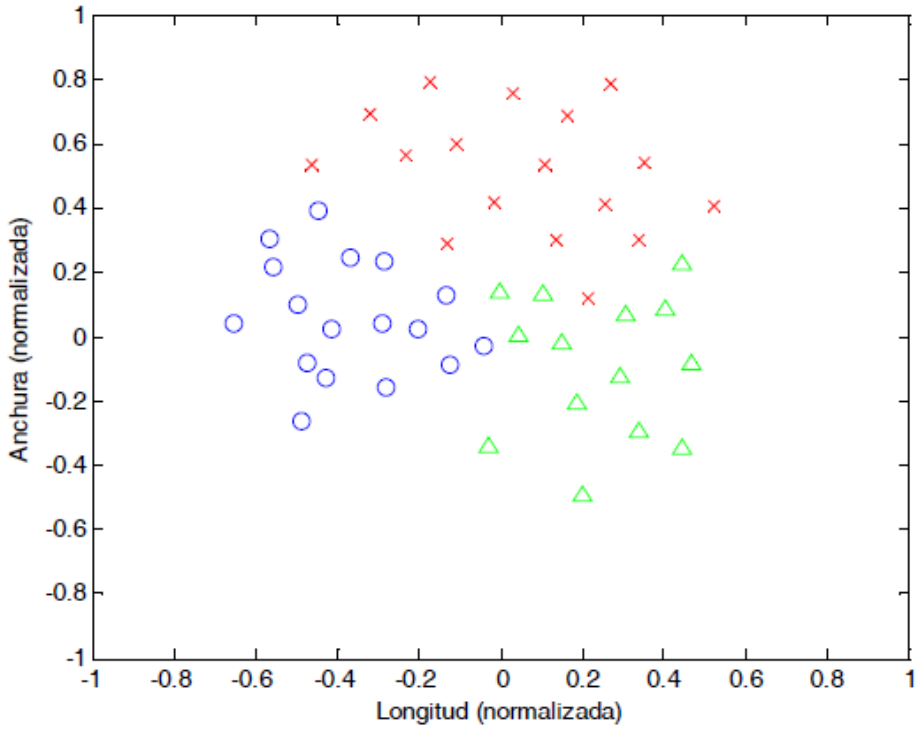
There is no information about the classes of the training data

Conjunto de entrenamiento:



K- Nearest Neighbours

Given a problem with 3 classes, with dimension 2: Length and Width

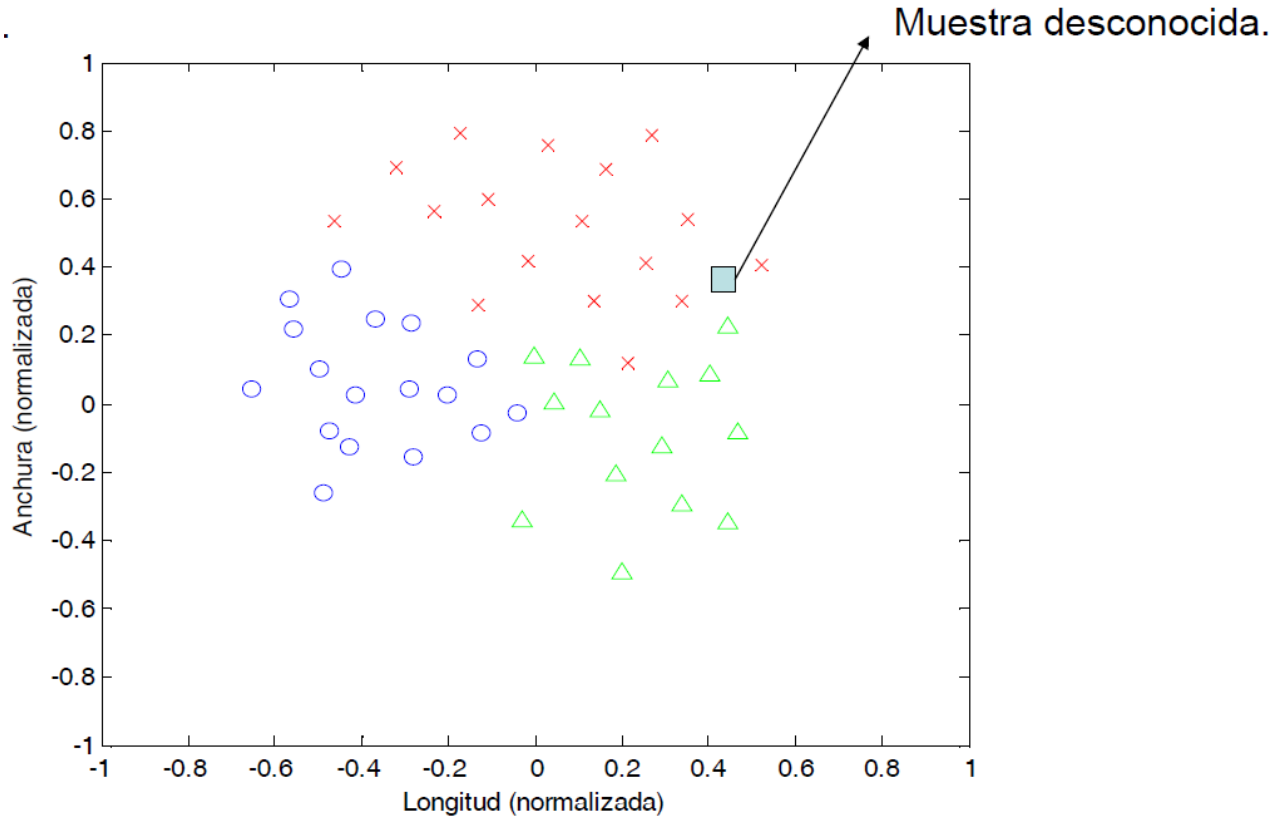


x Clase 1.

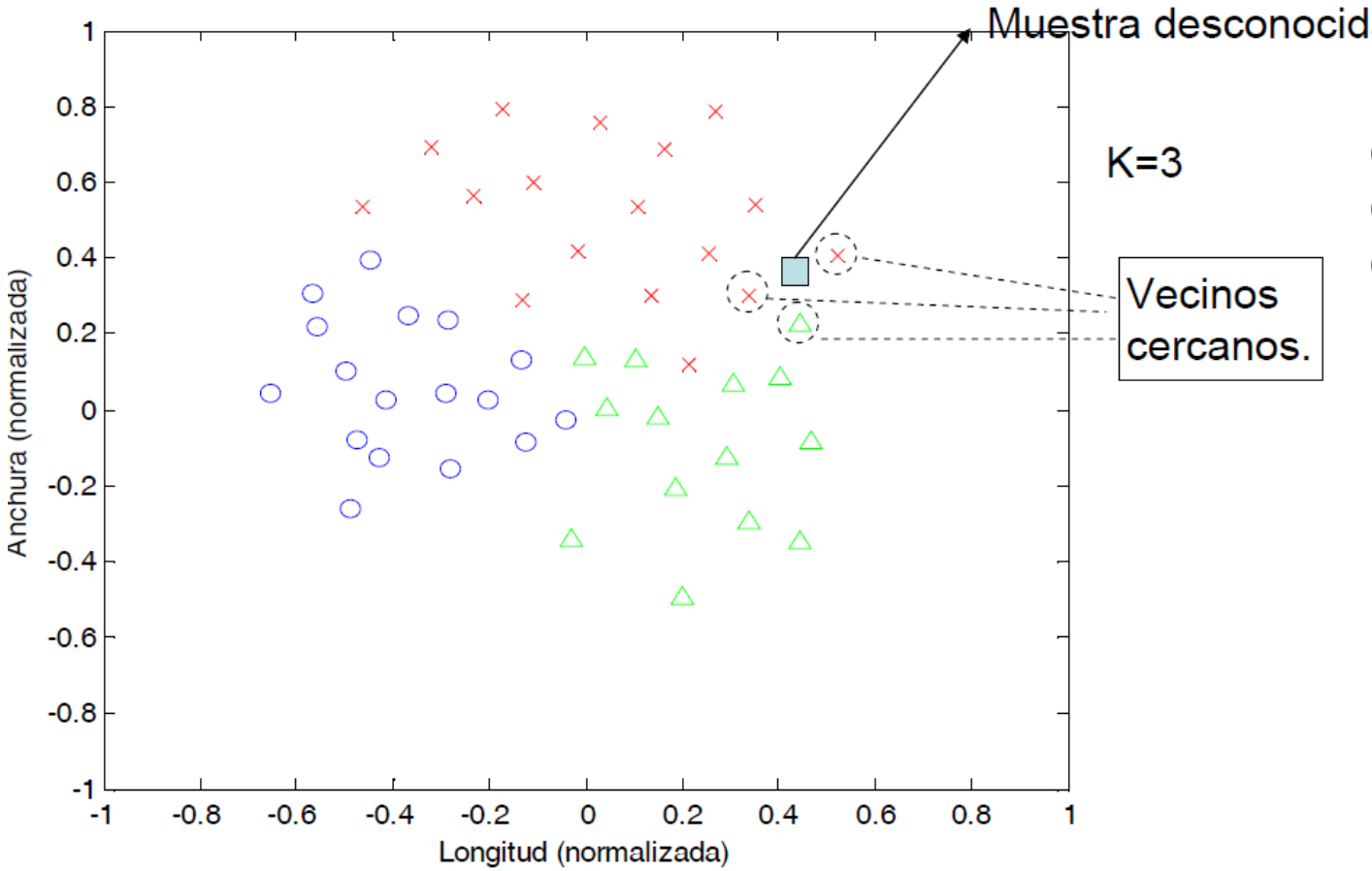
○ Clase 2.

△ Clase 3.

A new sample is classified:

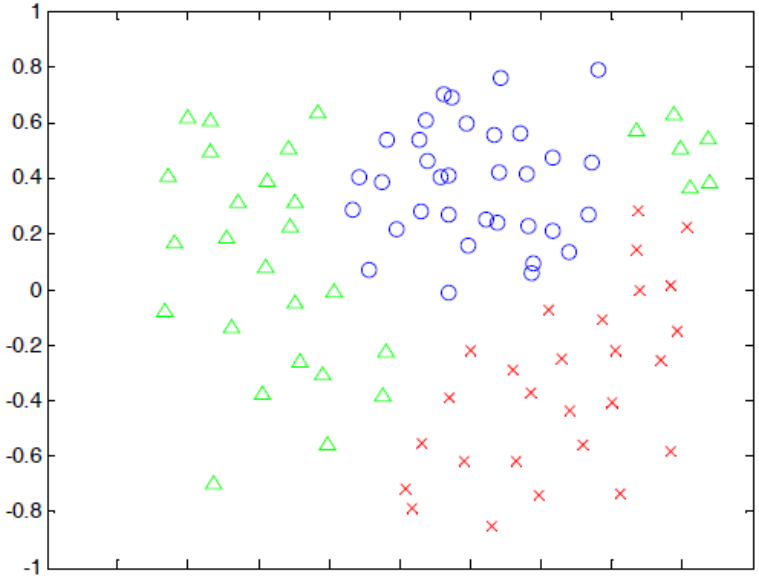


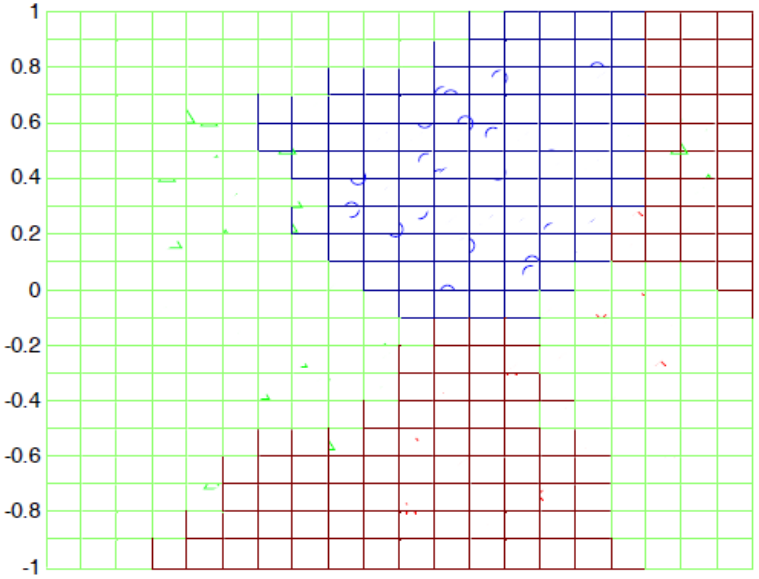
Euclidean Distance to the training samples



CLASS 1: 2 VOTOS
CLASS 2: 1 VOTO
CLASS 3: 0 VOTOS

Very simple method with good results





KNN

Pros:

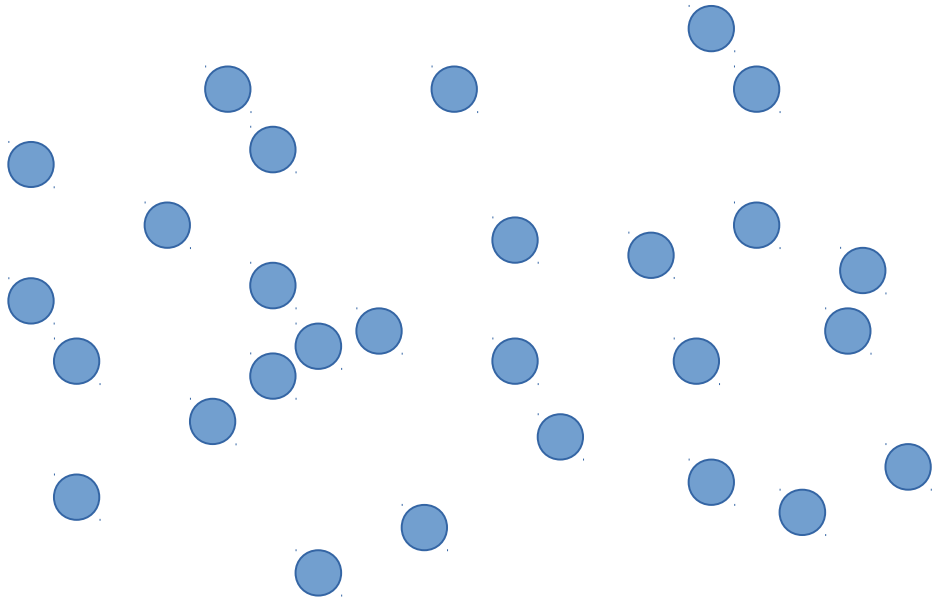
- Easy to understand and implement.
- Good results.
- Multiclass as Binary.

Cons:

- For big data set the computational load is high during test and the memory need is big.

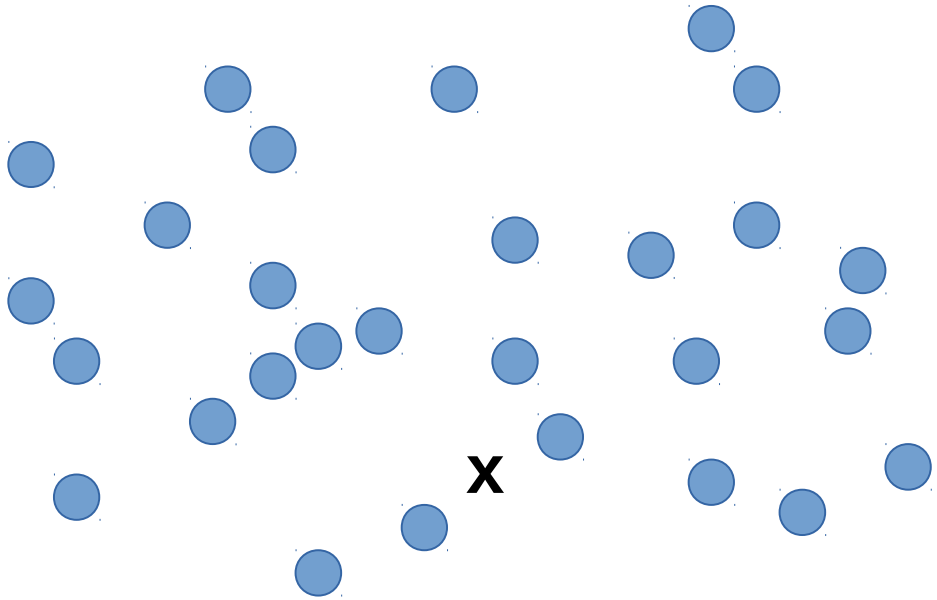
K-LAESA reduce computational load problem

K-LAESA



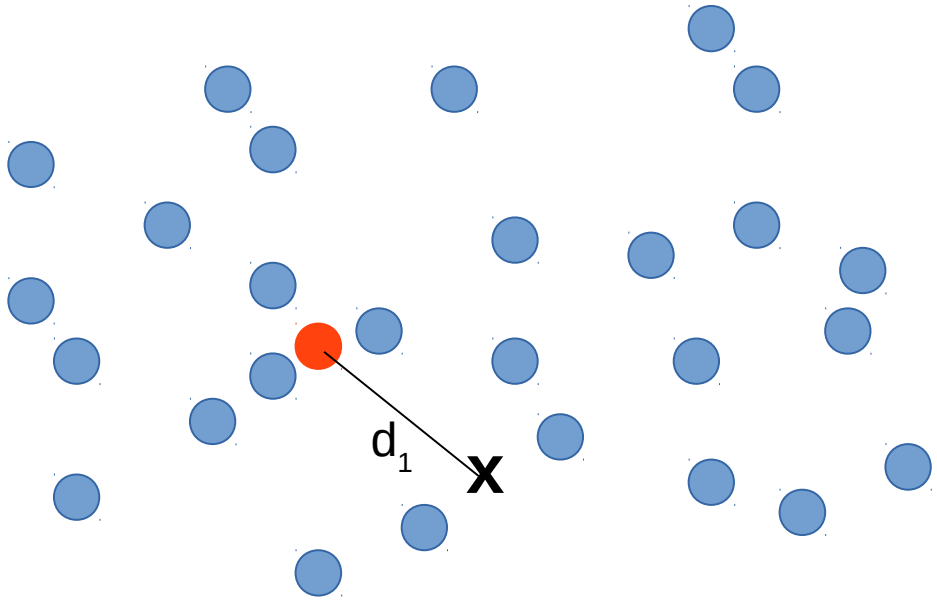
Training set

K-LAESA



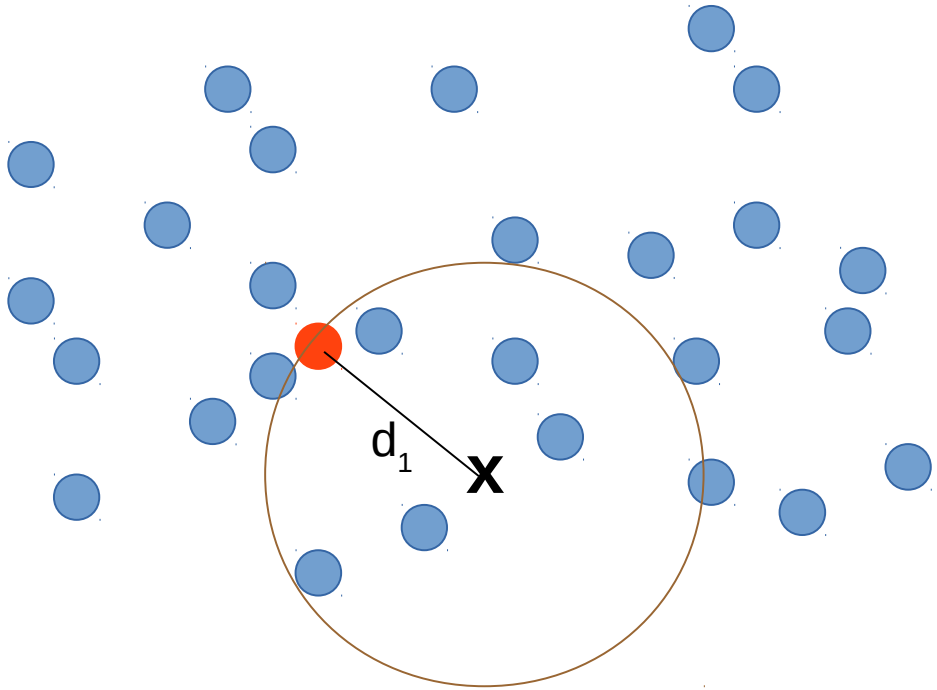
Input vector to be classified

K-LAESA



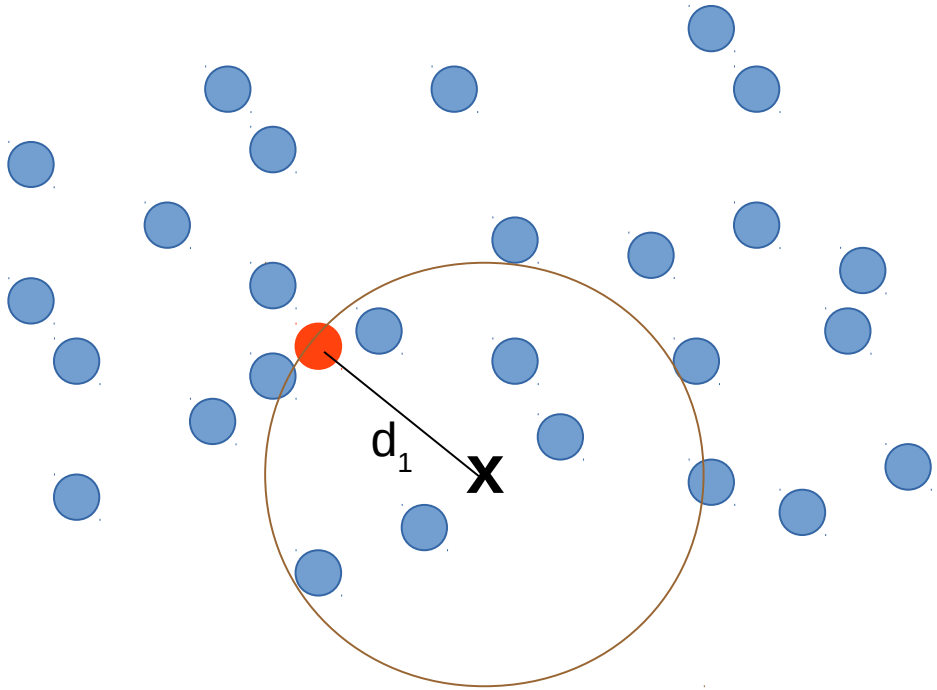
The first calculated distance

K-LAESA



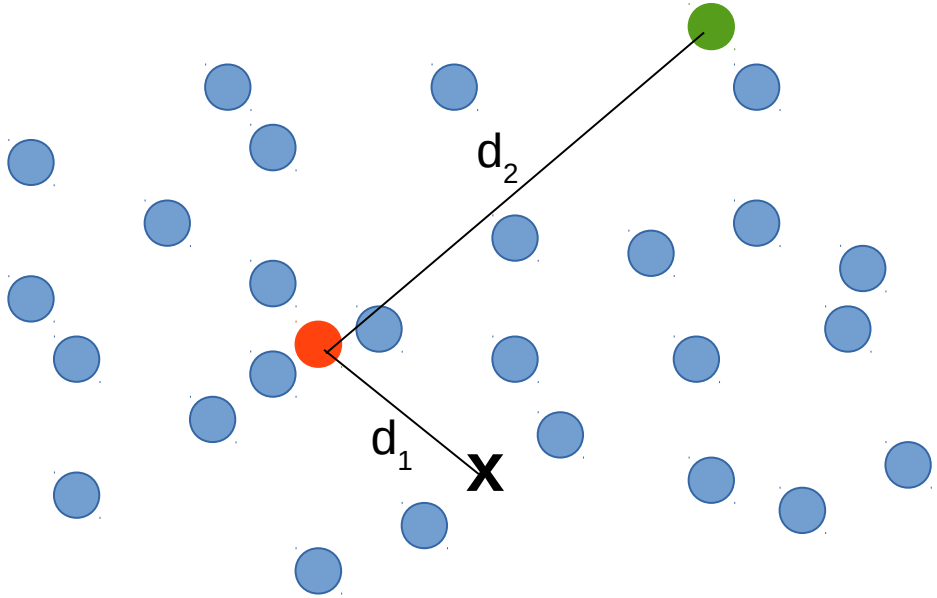
Every vector outside the circle is not closer the the first vector

K-LAESA



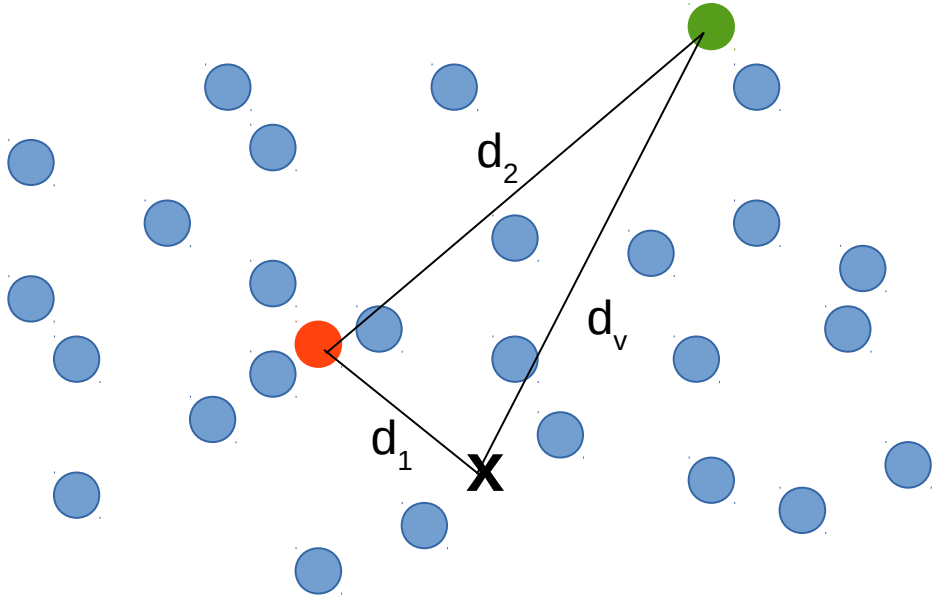
We do not know which vector are inside the circle

K-LAESA



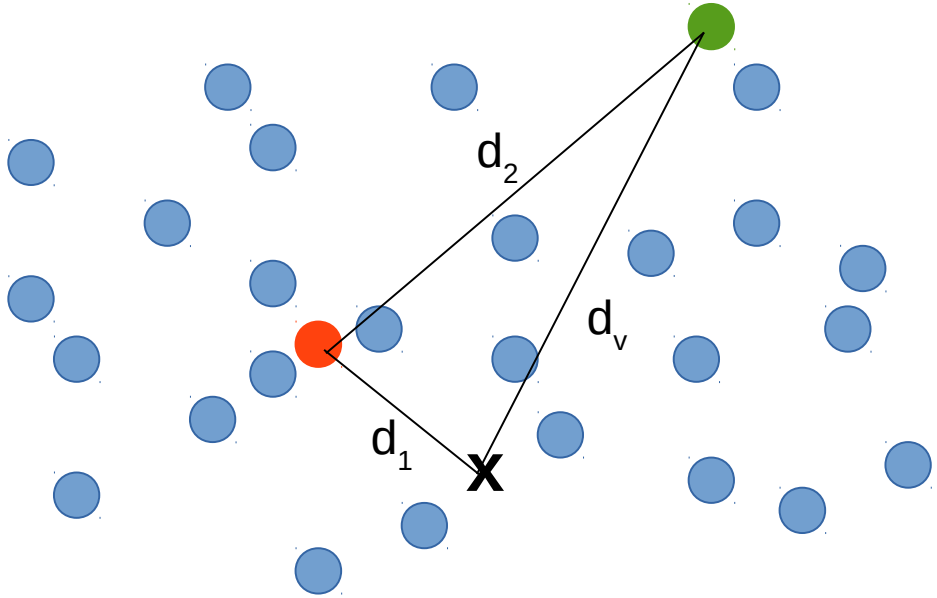
If the distances to a reference vector are known

K-LAESA



Si conocemos las distancias al primer vector con el que hemos probado, $d_2 \leq d_1 + d_v$ ==> $d_2 - d_1 \leq d_v$

K-LAESA



If we have an ordered list with the distances when a dy largaer than the minimum is obtained the process can be finished

Exercise

Ejercicio 1:

>> Representar_Datos(‘data1.mat’);

>>Clasificar_Nueva_Muestra('data1.mat');

Ejercicio 2:

>>data=Obtener_Datos( );

>> save p.mat data

>>Clasificar_Nueva_Muestra('data1.mat');

Ejercicio 3:

Modificación del código para 5 vecinos

Exercise

Exercise 4:

>> load datachar.mat

>> visualizar_datos

>> model=Train_kNN(data,2);

>>
[ACC,CM,Labels]=Test_kNN(dataTest,model);

ACC: Accuracy

CM: Confusión

Exercise

Confusion Matrix

	C1	C2	C3	C4	C5
C1	0.98	0.01	0.00	0.00	0.01
C2	0.00	0.97	0.02	0.01	0.00
C3	0.00	0.00	1.00	0.00	0.00
C4	0.00	0.01	0.00	0.97	0.02
C5	0.01	0.00	0.00	0.02	0.97

Exercise

Confusion Matrix

98% of the samples from class 1
Are correctly classified

	C1	C2	C3	C4	C5
C1	0.98	0.01	0.00	0.00	0.01
C2	0.00	0.97	0.02	0.01	0.00
C3	0.00	0.00	1.00	0.00	0.00
C4	0.00	0.01	0.00	0.97	0.02
C5	0.01	0.00	0.00	0.02	0.97


Universidad
de Alcalá


Teoría de la Señal y Comunicaciones