UniGe | MaLGa

Lab 1 – KNN

K-NN

- It is a local method that follows the idea of predicting the output of a new input reasoning on the outputs of the K-closest points in the input space

Regression

$$\hat{f}(x) = \frac{1}{K} \sum_{j=1}^{K} y_l$$

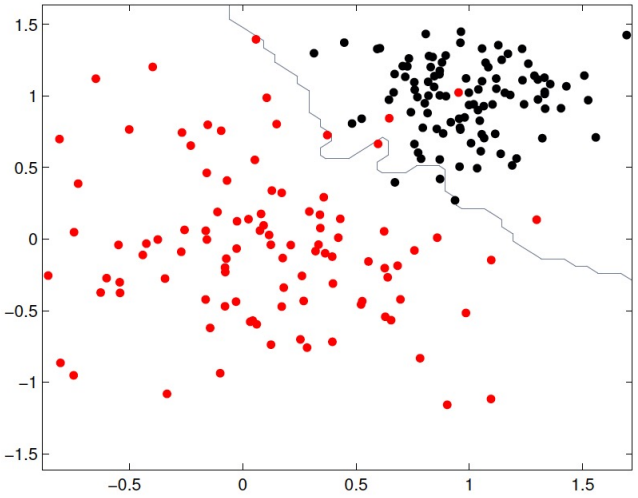
Binary classification

$$\hat{f}(x) = sign \sum_{j=1}^{K} y_l$$

K-NN: the algorithm

Given x (the new input), S (the training set), and K

- Compute the distances between x and all the points in S
- Sort the distances in increasing order
- Take the outputs of the K closest points
- Compute the predicted output for x according to one of the rules (depending on the task)



Parameter K, noise and number of samples

- K controls the fit and the stability of the function estimated by the KNN algorithm
- We discussed the fact the choice of K influences the “quality” of the estimator, also depending on the amount of noise and the number of samples in the training set
- Today we try and appreciate the effect of its value on the behavior of the K-NN algorithm

Objectives for today

- In the hands-on activity you are asked to provide a (guided) implementation and analysis of K-NN as you change K, the amount of samples in the training set and the amount of noise with specific reference to the properties of fitting and stability
- NOTE
  - To evaluate the fitting, you may consider the prediction ability on the training set
  - To evaluate the stability, we simulate the availability of future data, generating a new test set

UniGe