

Homework III: Linear Discriminant Functions and KNN

1 Background

Using scikit-learn facilities the construction of a classifier involves the following steps

- Create the classifier, i.e., initialize the object classifier. The hyper-parameter of the classifier are assigned at this phase. Note that there are default values.
- Learning. The object classifier and the training data set (features and labels) are passed to a function named *fit*
- The trained classifier can be used to predict. Function *predict* can be applied to data. This step is used to evaluate the classifier. Therefore, data which was not used in training phase should be considered. Note that there are systematic approaches to achieve this goal.

Linear Classifiers are described by discriminant functions

$$g(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b$$

The parameters are (\mathbf{w}, b) . There are different algorithms that learn the parameters given the data. For each the hyper-parameters are different.

K- Nearest Neighbor is an instance based classifier. Hyper-parameters are the similarity metric and K (number of neighbors)

2 Homework

1. Paper-and-pen task: answer to the questions published in the Exercises (Notes).
 - Choose one exercise of the list 2.5, 2.6, 2.8
 - Choose one exercise of the list 3.1, 3.2, 3.3
2. Programming task. You are expected to work with toy data and learn the main characteristics of the classifiers (linear discriminant functions and KNN). With the toy data you should also apply one evaluation methodology.
 - (a) Using toy data (*blobs*, *moons*, *circles*) and a graphical evaluation strategy completing and improving the template *templateIIGraph.ipynb*

- complete the template to include KNN.
 - comment the results of learning for the three (*blobs, moons, circles*) data sets. Suggestion: Describe the characteristics of regions and decision surfaces for the different classifiers.
- (b) Generate larger data sets. And create a new template to perform an hold out performance evaluation as it is illustrated in template *templateIIExp.ipynb*
- (c) (*bonus*) Develop a classification model for real data. Evaluate the model and comment. The group can choose the data set.

3 Useful Links

- Perceptron https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.Perceptron.html
- Gradient based approaches.
<https://scikit-learn.org/stable/modules/sgd.html>
- KNN.
<https://scikit-learn.org/stable/modules/neighbors.html>
- Evaluation.
https://scikit-learn.org/stable/modules/model_evaluation.html