Practical session 7 SVM

February 12, 2019

Some website :

- `http://scikit-learn.org/stable/modules/svm.html`
- `http://en.wikipedia.org/wiki/Support_vector_machine`

This lab aims at introducing SVM and use them on real and synthetic datas using the package `scikit-learn`.

Exercise 1

1. Download the digit datasets : `http://yann.lecun.com/exdb/mnist/`. The images are stored in the file `t10k-images-idx3-ubyte`. The labels are stored in the file `t10k-labels-idx1-ubyte`
2. Install the library `idx2numpy` using the command
   `pip3 install idx2numpy`
3. Import the idx file and convert it in `np.array` using the command `idx2numpy.convert_from_file`
4. Define and fit the SVM linear model on the train set
5. What is the accuracy of the method on the test set?

Exercise 2 We shall use the object `sklearn.svm.SVC` :
`from sklearn.svm import SVC`

1. Use the website :
   `http://scikit-learn.org/stable/modules/svm.html`
   and the dataset `Iris`. Implement a classifier which classifies class 1 against class 2 of the dataset `iris` using the two first variables and a linear kernel. Use half of the dataset for training and half of the dataset for validation. To import the dataset `iris`, type

```
from sklearn import datasets
iris = datasets.load_iris()
X = iris.data
y = iris.target
X = X[y != 0, :2]
y = y[y != 0]
```

2. Compare the performance of the linear SVM with a SVM based on polynomial kernel