LINMA 2472 ALGORITHMS IN DATA SCIENCE: HOMEWORK II, PART 2

Assignment 2 (part 2): Random Fourier features (RFF)

In this part, you shall train three different classifiers on the «MNIST» dataset. The purpose of this part is to demonstrate the speed gains on evaluating new instances when using RFFs. For this purpose we shall use 10.000 training instances and 60.000 testing instances.

Your task is to:

1. Train a linear SVM with the training data,
2. Train a kernel SVM (with the Gaussian kernel) on the same data,
3. Implement the random Fourier features (following algorithm 1 in the slides, slide 35),
4. Train a linear SVM over the transformed data.

NB: Training may take up to a few minutes on a normal computer.

Q1. (Testing time) You should use the default parameters provided in the code for this question.

With your implementation of Random Fourier Features, and of the various SVMs, compare the time taken for each method to classify the 60.000 testing instances (this does not include the training time). Do your observations match the theoretical reduction of complexity described in the lectures ?

Q2. (Training time and parameters) Try different values for the parameter D. Check the influence the size of D has over :

- the accuracy of the linear SVM with RFF,
- the time taken to create the RFF,
- the time taken to train the SVM with RFF,
- the time taken by the classifier to classify the testing instances.

If the calculations are too complex to be performed quickly enough by your computer you may (for this question only) restrict yourselves to a subset of instances for training and testing. If you decide to do so, make sure you explicitly say what you do, why you do it, and what you gain from it.