Random Fourier features (RFF)

Please prepare a written report (**max** 5 pages) with appropriate figures based on the results from the assignments. The deadline for this project is on the 13th of November at 23h59 (along with the other part of the homework). Hand in all your files in a **single** zip archive (so two reports and two jupyter notebooks), following the format: "group_#_Name1_Name2_Name3_HW2.zip".

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 with default parameters)   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. (Varying Parameters)   Try different values for the parameter D, we suggest to range D from the order of 10 to the order of 1000. 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.

Comment your results, do not hesitate to use figures to illustrate your statements. If you do use graphs, make sure they are easy to read, with large labels and clear captions.