CS4801: Principles of Machine Learning
Programming Assignment 2

6 points
Due on 13th September 2017
No request for change will be accepted

This homework consists of only programming assignment on SVM. A few instructions to make life easier for all of us:

- please submit your code and a short discussion on your observation (preferably PDF and latex) from your experiments. Put all codes and report in a single zipped file and name it as <First-name><Last-name>.zip. Then submit it in moodle.
- Deadline for programming assignment is 17:00 pm 13th September 2017.

1 Programming Exercises

Exercise 1 : Multi-class classification with SVM

In this exercise we are doing handwritten digit classification using multi-class SVM with a Gaussian kernel. In order to solve the optimization problem for the SVM, we are using the python interface to the LIBSVM package (http://www.csie.ntu.edu.tw/cjlin/libsvm/). Download DataSVM.zip from the course webpage in moodle. Note: You do not need to download anything from the LIBSVM webpage - everything you need is contained in the zip-file.

- Extract the files in libsvm-3.22.zip somewhere in your home directory.
- To install libsvm in your command line execute "make".
- Go inside "python" folder and read README file.
- (1 points) Write a function "getKernelSVMSolution.py"
  - Take Inputs:
    - Xtr
    - Ytr
    - C [tradeoff hyperparameter]
    - λ [kernel width]
    - Xts
  - Return Output:
    - Yprediction
- The problem deals with the classification of handwritten digits (10 classes). You are supposed to use the SVM with the Gaussian kernel:

$$k(x, y) = \exp -\lambda \|x - y\|_2^2$$

  . The training and test data is in USPSTrain.csv and USPSTest.csv and labels are in USPSTrainLabel.csv and USPSTestLabel.csv.
- (1.5 points) Write a program "OneVsOne.py" which implement the multi class classification by using OnevsOne scheme.
  - Convert both data USPSTrain.csv, USPSTrainLabel.csv and USPSTest.csv to SVM compatible format such that for each binary classification problem.
    - Create appropriate vector for class label "Y" which contains only 1 and -1.
    - Create appropriate feature X which is a list of list for example (for a feature matrix with 2 samples ans 3 features [[1,0,1],[-1,-1,0]]) the X will be defined as

$$X = [\{1 : 1, 3 : 1\}, \{1 : -1, 2 : -1\}].$$

  - Then execute binary SVM with modified data.
  - Predict multi-class class label form your binary prediction.
  - Calculate an appropriate classification error for your multi-class classification task,i.e, F1 score or AUC.

- (1.5 points) Following the similar scheme also write a program "OneVsRest.py" which implement the multi-class classification by using Onevs-Rest scheme.

- In both cases use $C = 100$ and $\lambda = 3/\gamma$, where γ is the median of all squared distances between training points, as parameters for the binary SVM.

- (2 points) Write a report "USPSreportFirstnameLastname.pdf" containing following
  - Test errors for both cases.
  - Visually inspect the digits which have been missclassified using confusion matrix for multi-class classification.
  - How do you judge the result? Compare the quality of the classification obtained by the two multi-class schemes.
  - How do the two multi-class schemes compare in terms of runtime?
  - Also generate for both cases a figure (ErrorsOneVersusOne.png and ErrorsOneVersusRest.png) containing the missclassified images in the test set

- Save your prediction on the test set in two files name as PredOneVersusOne.txt and PredOneVersusRest.txt.