CSE471: Statistical Methods in AI

Monsoon 2016

Assignment #3: SVM, Kernel Methods

Due: Before 5:00pm on 29th October 2016   **Total Marks:** 40 (Mapped to 4% of course credits)

General Instructions:

- Assignment can be implemented in Matlab/Octave, Python, C/C++, R .
- Ensure that submitted assignment is your original work. Please do not copy any part from any source including your friends, seniors and/or the internet. If any such attempt is caught then serious actions including an **F grade in the course** is possible.
- A single pdf file needs to be uploaded to the Courses Portal. The file should contain your answers as well as the code you have written and its output (Or as directed by the TA's).
- Include the assignment number, your name and roll number at the top-left of the first page of your submission.
- Your grade will depend on the correctness of answers and output. In addition, due consideration will be given to the clarity and details of your answers and the legibility and structure of your code as well viva based oral examination done by TA's.

Problem 1 (6 Marks)

Derive the non-linear (kernelized) version of the Fisher's Linear Discriminant Analysis.

Problem 2 (16 + 18 Marks)

Implement and provide the pseudo code, assumptions made, accuracy results and your observations in a report format along with the code for the following two tasks.

a) Implement non-linear dimensionality reduction techniques, namely, Kernel Principal Component Analysis (PCA) and Kernel Linear Discriminant Analysis for UCI Arcene cancer classification data and another dataset (from UCI or anyother ML dataset listing) that have real vectorial data representation.

b) Employ SMV classifier (using any standard library) for linear and RBF kernel in the resulting K-dimensional PCA space (where K=10, 100 for the Arcene data and any two K values of choice for another dataset) as well as 1-dimensional LDA space and report the classification performance on validation data for both the datasets.