CS511 : Learning With Kernels
Assignment 2 and 3
Submission Date: 5-April-2014@17:00

Question 1: (10 points)

PCA: You are given two datasets namely `iris.data` and `optdigit.data`. `iris.data` consists of 150 data points. Each of the data point has four features. There are a total of 3 classes in this dataset (last column of the dataset). You need not take into account class label while performing PCA (or KPCA). Similarly `optdigit.data` consists of 3823 data points. Each data point is has 64 features. Last column of this dataset represent the class label to which the data point in question belongs. Datasets can be obtained from:

- `http://jatinga.iitg.ernet.in/~saradhi/iris.data`

- `http://jatinga.iitg.ernet.in/~saradhi/optdigit.data`

(a) Obtain principal directions and principal components for each of the dataset.

(b) For each of the dataset obtain biplot for different values of $\alpha = \{0, 0.5, 1\}$. List your observations.

(c) Re-construct data points from the obtained principal directions. Report the deviation in the reconstruction error (through a plot) when you use varying principal directions ($1 \leq d \leq 4$ *or* 64). The plot should be: x axis represent data point number and y axis represent the reconstruction error ($\|\mathbf{x}_i - \widetilde{\mathbf{x}}_i\|$; where $\widetilde{\mathbf{x}}_i$ represent the reconstructed $\mathbf{x}_i$ using chosen number of principal directions).

(d) Plot the eigenvalue variations for these datasets.

Question 2: (10 points)

KPCA: Repeat above steps for kernel principal component analysis for two varying kernel functions namely:

- Polynomial kernels: $k(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j + 1)^p$; p is a specified value.

- Radial basis function: $k(\mathbf{x}_i, \mathbf{x}_j) = exp^{\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right)}$; σ is a specified value.

Write your observations on differences between linear projections