Machine Learning for Sensory Signals

Homework # 2 Due date: March 26, 2017

March 16, 2017

1. For image data, the input which has number of dimensions (D) is typically much higher than number of observations (N). Assume that data is mean removed, i.e., $\boldsymbol{\mu} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i = \mathbf{0}$. In such a case, in order to apply PCA, we need to find the eigenvectors of the data covariance matrix $\boldsymbol{S} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^T$. Let $\mathbf{X} = [\mathbf{x}_1\ \mathbf{x}_2\ ,..,\ \mathbf{x}_N]^T$ denote the data matrix of size $N \times D$ containing each image data point in the rows. Show that

   (a)

   $$\boldsymbol{S} = \frac{1}{N}\mathbf{X}^T\mathbf{X}$$

   (b) Show that any eigen vector of $\mathbf{u}$ of $\boldsymbol{S}$ can be obtained by using the eigen vector v of $\hat{\boldsymbol{S}} = \mathbf{X}\mathbf{X}^T$ as

   $$\mathbf{u} = X^T\mathbf{v}$$

   (c) How will this reduce the computational complexity of PCA computation. (**Points** 20)

2. Given the images in following database
   http://leap.ee.iisc.ac.in/sriram/teaching/MLSS/assignments/data.tar.gz
   15 subject faces with happy/sad emotion are provided in the data. Each image is of 100x100 matrix. Perform PCA on to reduce the dimension from 10000 to $K = 10$ and then perform LDA to one dimension. Plot the one dimension features for each image. Select the optimum threshold to classify the emotion and report the classification accuracy on the test data. What is the optimum value of the K ? (**Points** 30)