E9 205 – Machine Learning for Signal Processing

Homework # 1

January 25, 2025

Due date: February 4, 2025

Analytical part, prepared in writing, can be scanned. This should be attached to the report on the coding part. Finally, a single pdf file containing the response (no in-person handouts) is to be submitted.
Source code also needs to be included.
Name of file should be "Assignment1_FullName.pdf" submitted to teams channel.
Assignment should be solved individually without collaboration with other human or online (GPT-like) resources..

1. Prove the following two matrix derivative properties for square symmetric matrices $\mathbf{A}, \mathbf{B}$,

$$\frac{\partial}{\partial \mathbf{A}} \log(|\mathbf{A}|) = 2\mathbf{A}^{-1} - diag(\mathbf{A}^{-1})$$

$$\frac{\partial}{\partial \mathbf{A}} tr(\mathbf{AB}) = 2\mathbf{B} - diag(\mathbf{B})$$

(**Points** 10)

2. **Fisherfaces -** Sagar is a data scientist who analyzes face images for detecting emotions. In his course, he has leant about LDA and wants to use it to reduce the dimensionality before training a classifier. However, he is faced with a situation where he has N face images each of dimension D with $N << D$. As he knows to apply PCA for high dimensional data, he uses whitening to reduce the dimensionality to $d < N$. The whitenning process can be described as,

$$\boldsymbol{y} = \boldsymbol{\Lambda}^{-\frac{1}{2}} \boldsymbol{W}^T (\boldsymbol{x} - \boldsymbol{\mu})$$

where $\boldsymbol{x}$ is the input D dimensional image, $\boldsymbol{\mu}$ is the sample mean of input images, $\boldsymbol{W}$ is the PCA projection matrix of dimension $D \times d$, $\boldsymbol{\Lambda}$ is $d \times d$ diagonal matrix containing d largest eigenvalues of sample covariance and $\boldsymbol{y}$ is the whitenned output of dimension d. Given a set of N data points, $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N$ and the corresponding class labels $t_1, t_2, ..., t_N$, (where $t_n = \{1, 2, ..., K\}$, is one of the K-class labels), he tries to learn Fisher LDA projection on the whittened outputs, $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N$. Here, let $\boldsymbol{\mu}$ denote the sample mean for the N samples.

   (a) As a first step, Sagar tries to find the total covariance (sample covarince) of whitenned outputs y, given as,

$$\boldsymbol{S}_T^y = \frac{1}{N} \sum_{n=1}^{N} \mathbf{y}_n \mathbf{y}_n^T$$

Show that for this case, $\boldsymbol{S}_T^y = \boldsymbol{I}$ where $\boldsymbol{I}$ is the $d \times d$ identity matrix. (**Points** 10)

(b) Assuming that $\boldsymbol{S}_w^y$ is invertible, show that the first LDA projection vector $\boldsymbol{w}$ is given by the eigenvector of $\boldsymbol{S}_w^y$ with minimum magnitude of eigen value.

(**Points** 10)

3. **Maximum Likelihood Linear Regression** - Kiran is doing his PhD on acoustic channel estimation. In order to estimate the channel characteristics, he designs an experiment in which he records the ultra sound signal at the source as well as at the output of the acoustic channel. Let $\mathbf{x}_i, i = 1, \ \ldots \ , N$ and $\mathbf{y}_i, i = 1, \ \ldots \ , N$ denote the feature sequence corresponding to the source and channel outputs. He begins with an assumption of a linear model for the channel,

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b} + \boldsymbol{\epsilon}$$

where the source features $\mathbf{x}$ are assumed to be non-random and $\boldsymbol{\epsilon}$ represents i.i.d. channel noise $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$. Given this model, his advisor Mohan recommends the maximum likelihood (ML) method to estimate the parameters of the channel $\big(\mathbf{A}, \mathbf{b}, \sigma\big)$. How will you solve the problem if you were Kiran ? (**Points** 10)

4. **Maximum Likelihood Classification** Consider a generative classification model with K classes defined by prior probabilities $p(C_k) = \pi_k$ and class-conditional densities $p(\boldsymbol{\phi}|C_k)$ where $\boldsymbol{\phi}$ is the input feature vector. Suppose that a training data is given $\{\boldsymbol{\phi}_n, \boldsymbol{t}_n\}$ for $n = 1, .., N$ and $\boldsymbol{t}_n$ denotes a binary target vector of dimension K with components $t_{nj} = \delta_{j,k}$ if input pattern $\boldsymbol{\phi}_n$ belongs to class k. Assuming that the data points are drawn independently, show that the ML solution for prior probabilities is given by,

$$\pi_k = \frac{N_k}{N}$$

where N_k is the number of points belonging to class k. (**Points** 10)

Coding

5. **PCA + LDA** - Data is posted here
http://leap.ee.iisc.ac.in/sriram/teaching/MLSP25/assignments/data/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 (using PCA for high dimensional data) 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 best choice of K which gives the maximum separability ? (**Points** 25)

6. **Speech spectrogram** - We have clean and noisy speech files here
http://leap.ee.iisc.ac.in/sriram/teaching/MLSP25/assignments/data/speech.zip
Pick any single wav form.

The files are in wav format sampled at $16kHz$. Use library function (librosa mel spectrogram function from librosa tool) to compute the spectrogram features of clean and noisy files (use 20 ms window length with a hop length of 10 ms). The spectrogram should be a matrix of size, $D \times T$, where D is mel dimension of the spectrogram and T is the duration of audio in ms by 100.

(a) Assume each time frame of spectrogram is independent of each other. From the clean files, compute the whitening transform. Apply the transform on spectrogram feature of a noisy file. Compute the covariance of the whitened spectrogram features of the noisy file. Is the covariance matrix identity ? Argue the reason for your result. Repeat the procedure by reversing the roles of clean and noisy files. (**Points** 10)

(b) On the clean speech file, using the T features from the clean file, train a 2 mixture Gaussian model from scratch. Train the model for 20 iterations. Plot the training log-likelihood function for 20 epochs for two cases i) random initialization, ii) k-means initialization. (**Points** 15)