E9 205 – Machine Learning For Signal Processing

Midterm Exam

Date: Feb 28, 2022, 3:30pm

Instructions

1. This exam is open book. However, computers, mobile phones and other handheld devices are not allowed.
2. Any reference materials that are used in the exam (other than materials distributed in the course channel) should be pre-approved with the instructor before the exam.
3. No additional resources (other than those pre-approved) are allowed for use in the exam.
4. Academic integrity and ethics of highest order are expected.
5. Notation - bold symbols are vectors, capital bold symbols are matrices and regular symbols are scalars.
6. Answer all questions.
7. For online exam - name your scanned copy of answer file in pdf format as FirstName-LastName-midterm.pdf and follow the upload to the Teams channel.
8. All answer sheets should contain your name and SR number in the top.
9. Question number should be clearly marked for each response.
10. Total Duration - **90 minutes including answer upload**
11. Total Marks - **100 points**

1. **Arriving in class on time** - Prof. Ali and teaching assistant Saikat are performing a data analysis on student arrival times in the MLSP class. The students in the class belong to two categories - i. those who have a class prior to the start of MLSP course and ii. those without a class before the MLSP course. The inter-arrival time of the students is measured in seconds, starting from $3:30$pm and for each passing student, the arrival time relative to the previous student, x, is counted. Prof. Ali proposes to use a 2-mixture Poisson distribution for modeling the student arrival times, given by,

$$p_{\boldsymbol{\Theta}}(x) = \frac{\alpha_1}{x!}\lambda_1^x e^{-\lambda_1} + \frac{\alpha_2}{x!}\lambda_2^x e^{-\lambda_2}$$

where the mixture model parameters, $\boldsymbol{\Theta}$, are the mixture weights α_1, α_2 and the Poisson coefficients are λ_1, λ_2. Prof. Ali asks Saikat to,

(a) Derive EM algorithm based update equations for the parameter estimation of the model. Is your ML estimation algorithm guaranteed to converge ?

(b) Formulate an approach to initialize the model parameters to start the EM algorithm.

How will you achieve these tasks if you were Saikat. **(Points 18 + 7 = 25)**

2. Dr. House wants to design a simple detector for COVID-19 using spirometry (airflow) measurements. His design involves checking the difference between readings for the current day with the previous day (the difference is denoted as x). He is interested in a two class problem, where COVID-19 is the target class (positive) and non-COVID-19 (consisting of conditions like purely healthy as well as other respiratory diseases) is the control group (negative). His data analyst has provided him a parametric distribution for the positive and negative group population,

$$p_+(x) = \mathcal{N}(x; 1, 1)$$
$$p_-(x) = 0.4\ \mathcal{N}(x; -1, 1) + 0.6\ \mathcal{N}(x; 2, 1)$$

where $\mathcal{N}(x; \mu, \sigma)$ denotes a Gaussian distribution with mean μ and variance σ^2.

(a) Using the parametric distributions, help Dr. House find thresholds on the reading x that will provide the minimum probablity of mis-classification.

(b) How will the classifier design change if Dr. House wants to also use a Loss matrix L defined as

$$L = \begin{bmatrix} 0 & 10 \\ 1 & 0 \end{bmatrix} \tag{1}$$

where the first class is the positive class and the second class is the negative class.

(Points 10 + 10 = 20)

3. **Missing Data PCA** - While working for a startup on wearable devices, Altaf is measuring vital biological data from subjects for a month. Since the data is high dimensional, he attempts to do PCA before further processing. Let $\mathbf{X} = \{\mathbf{x}^n, n = 1, ..., N\}$ denote a set of D dimensional bio measurements from wearables for N individuals. He sets up the

PCA problem to find a set of basis vectors $\mathbf{B} = \{\mathbf{b}^j, j = 1, ..., J\}$ each of dimension D with $\mathbf{B}^T\mathbf{B} = \mathbf{I}$ such that,

$$E = \sum_{n=1}^{N}\sum_{i=1}^{D}[x_i^n - \sum_{j=1}^{J} y_j^n b_i^j]^2$$

is minimized, where $\mathbf{x}^n = [x_1^n, .., x_D^n]^T$, $\mathbf{b}^j = [b_1^j, .., b_D^j]^T$ and $\mathbf{y}^n = [y_1^n, .., y_J^n]^T$. While trying to apply PCA on the data, he finds some of the measurements are missing for some subjects. Let γ_i^n denote the binary variable which indicates the presence ($\gamma_i^n = 1$) or absence ($\gamma_i^n = 0$) of the i-th bio measurement for the n-th individual. He then modifies the error as

$$E = \sum_{n=1}^{N}\sum_{i=1}^{D}\gamma_i^n[x_i^n - \sum_{j=1}^{J} y_j^n b_i^j]^2$$

With this modification, the optimization is no longer convex and therefore he resorts to an iterative algorithm. Specfically, given a set of basis vectors $\mathbf{B} = \{\mathbf{b}^j, j = 1, ..., J\}$, he derives the solution to the projections $\mathbf{Y} = \{\mathbf{y}^n, n = 1, ..., N\}$ and given the projections $\mathbf{Y}$ he derives solution to the basis vectors $\mathbf{B}$. What are the exact set of linear equations needed he solves for this iterative optimization ? Does his choice of optimization guarantee the minimization of the modified error function ? (**Points 25**)

4. **Speech Enhancement** - Let $\mathbf{y}_t, t = 1, ..., T$ denote clean speech signal which is observed as $\mathbf{z}_t = \mathbf{y}_t + \mathbf{v_t}$, where $\mathbf{v_t}$ is the noise. Let λ_y, λ_v denote the GMM for clean speech and noise signal respectively. Let $\mathbf{l} = l_1, l_2, ...l_T$ denotes the sequence of mixture component index of probabilities $p(\mathbf{y}_t|\lambda_y)$. Each $l_t \in \{1, ..., M\}$ where M denotes the number of mixtures (both the GMMs are assumed to have the same number of mixture components). The speech enhancement task is to estimate the clean signal $\mathbf{y}_t$ by maximizing $p(\mathbf{y}|\mathbf{z})$. Show that this can be achieved by iteratively maximizing $\Phi(\mathbf{y}, \mathbf{y}')$, where

$$\Phi(\mathbf{y}, \mathbf{y}') = \sum_{\mathbf{l}} p(\mathbf{l}|\mathbf{y}', \mathbf{z}) \log p(\mathbf{l}, \mathbf{y}|\mathbf{z})$$

How can we estimate $p(\mathbf{l}|\mathbf{y}', \mathbf{z})$. Will this approach guarantee converge. Is there a relationship with the EM algorithm ?
(**Points** 20)

5. **Estimation from Noisy Data** Let v be a real noisy measurement of a real random variable u, i.e., $v = u + \epsilon$, where ϵ represents an uncorrelated zero mean random variable with variance σ_ϵ^2. Let $\mathcal{R}_{uu}$ denote the second moment of u, $E[u^2]$. From the noisy measurement v, the aim is to find an estimate of u of the form $\hat{u} = av$. Let $e = u - \hat{u}$ denote the estimation error.

    (a) Compute the value of a which minimizes the mean square error $E[e^2]$ (**Points** 5)

    (b) Find the lowest possible value of mean square error $E[e^2]$. What happens to the best estimate and the lowest error when $\sigma_\epsilon^2 \to \infty$ (**Points** 5)