Department of Electrical Engineering, IIT Delhi

ELL784 Introduction to Machine Learning: Minor Examination

(Open book/Open Notes) Time: 1 hour Maximum Marks: 37

"Thou shalt not covet thy neighbour's answers"

Name: Entry No.

Assume all symbols to have their 'usual' meanings, as done in the class. No queries allowed: make suitable assumptions. Appropriate marks will be assigned according to their 'reasonability'.

1. Consider the $t_i = y(\mathbf{x}_i, \mathbf{w})$ example as done in class (and all symbols have their meanings, as done in class). Our model for explaining t_i is $\hat{y}_i = \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_i) = \sum_{j=0}^{M-1} w_j \phi_j(\mathbf{x}_i)$. Collect all such $\hat{y}_i$ to get a column vector $\hat{\mathbf{y}}$, just as we collect all t_is to get $\mathbf{t}$.
   (a) Use the expression of $\mathbf{w}_{ML}$ to derive $\mathbf{H}$, where $\hat{\mathbf{y}} = \mathbf{H}\mathbf{t}$.
   (b) Given that we usually have $N >> M$, what is the maximum rank of $\mathbf{H}$?
   (c) Show that $\mathbf{H}$ is a symmetric matrix
   (d) Show that $\mathbf{H}^n = \mathbf{H}$
   (e) Show that $\mathbf{H}$ has 1 as an eigenvalue
   (f) Show that $\mathbf{H}$ has 0 as an eigenvalue *(1+2+2+2+3+2 marks)*

2. Show that under the SVD construction, $\mathbf{u}_i$ the eigenvectors of $\mathbf{P}\mathbf{P}^T$ with the corresponding eigenvalues σ_i^2. *(5 marks)*

3. For a matrix $\mathbf{B}_{k \times k}$, the scalar $\mathbf{x}^T\mathbf{B}\mathbf{x}$ is called a quadratic form. Show that $\mathbf{B}$ is symmetric. Why is this result a bit shocking? *(2+1 marks)*

4. The Stauffer-Grimson Background subtraction method learns the mixture of Gaussians parameters possibly by an EM algorithm, from the statistics for the first given τ frames. Hence, the π_js calculated here are probabilities. For data from the $\tau + 1$th frame, the updating rule is

$$\pi_j^{\tau+1} = (1 - \alpha)\pi_j^{\tau} + \alpha\delta_{j,top\ match}^{\tau+1}$$

   All the notation used is as done in class, and all symbols have their 'usual' meanings. δ is the Kronecker Delta, and $\alpha \in [0, 1]$ is the learning rate. Is a re-normalisation necessary, to ensure that all π_js remain probabilities? *(4 marks)*

5. Consider a set of N people checked for a disease, with a screening test, and knowing the ground truth (Healthy(H) or Diseased(D)). The following table is to be interpreted as follows: n_1 is the number of people for whom the screening test returns H and they are actually healthy.

	H	D
screen=H	n_1	n_2
screen=D	n_3	n_4

   You are given $p(H) = \theta$, $p(\text{screen} = H|H) = \alpha$, $p(\text{screen} = D|D) = \beta$.
   (a) Write the relation(s) between the n_i and N
   (b) If the screening test is relatively effective, comment on the diagnals of the given matrix
   (c) Compute $p(\text{screen} = H, H)$ and the other three joint probabilites corresponding to the four entries of this table
   (d) Compute the joint likelihood of all the given data, with a suitable 'reasonable' assumption
   (e) Find the ML estimate of θ
   (f)

	H	D
screen=H	0	K_1
screen=D	K_2	0

   You are given this *Loss matrix*, where the cost of each decision is mentioned. Compute the expected loss of this screening test in terms of all given parameters *(1+1+4+2+3+2 marks)*