Department of Electrical Engineering, IIT Delhi

EEL709 Pattern Recognition: Major Examination

(Closed book/Closed Notes) Time: 2 hours Maximum Marks: 50

"Thou shalt not covet thy neighbour's answers"

```
''We used your Minor II question paper on our detainees from the
Universities in Kiev and Crimea. Most confessed, within five minutes
of reading it. Whether it was solely due to the content, or Bing's
translation of the same, I cannot say, but nevertheless...spasiba.''
do svidaniya,
Col. Vladimir Illiych Pugachev,
Head, Secret Police, Russian Federation (Ukraine cadre).
```

1. (a) It is sometimes said that the Bayesian view of probability is subjective, whereas the frequentist view is objective. Assess this statement; to what extent do you agree with it? Can you give an example of a pair of probabilistic statements to illustrate the difference? *(3 marks)*

   (b) A coin is tossed 5 times, and 5 heads are observed. Would either a frequentist or a Bayesian infer that the coin has heads on both sides? Why or why not? *(2 marks)*

2. Suppose we have a regression problem with 2D inputs, the two dimensions being denoted by x_1 and x_2. We know that $x_i \in [-1, 1], i = 1, 2$. Given the inputs, the target output t is generated via the following model: $t \sim \mathcal{N}(x_1^3 x_2^5 - 10x_1x_2 + 7x_1^2 + 5x_2 - 3, 1)$.
   We generate 20 training points sampled uniformly at random from the input space, and attempt to fit least-squares polynomial regression models of order 1, 2, 8, and 10 to this data set. An independent set is generated for testing the models.
   Of the 4 fitted models, which do you think will have:
   (a) The lowest training error?
   (b) The highest training error?
   (c) The lowest typical test error?
   If you think multiple models will be equivalent on any of these counts, list all of them. Justify your answers. *(3 marks)*

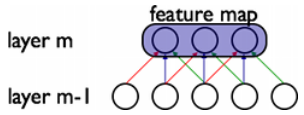
3. (a) In class we saw how the kernel trick can be used to learn non-linear decision boundaries in a given feature space, by implicitly mapping the data to a d-dimensional space. Further, let the feature mappings corresponding to these kernels be $\boldsymbol{\phi}_1 : \mathbb{R}^n \to \mathbb{R}^d$, $\boldsymbol{\phi}_2 : \mathbb{R}^n \to \mathbb{R}^d$. Let c be a real-valued constant. Show how $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ can be used to compute the following kernels: (i) $K(\mathbf{x}, \mathbf{x}') = cK_1(\mathbf{x}, \mathbf{x}')$; (ii) $K(\mathbf{x}, \mathbf{x}') = K_1(\mathbf{x}, \mathbf{x}')K_2(\mathbf{x}, \mathbf{x}')$. *(2 marks)*

   (b) Suppose we have a two-class data set in 2D-space, generated as follows: positive samples taken from points on the curve $x_1^2 + x_2^2 = 5$, and negative samples taken from points on the curve $x_1^2 + x_2^2 = 10$. The number of samples in both classes are equal. Show visually the kind of decision boundary that would be obtained by training an SVM with a (i) linear kernel; (ii) polynomial kernel of order 2; (iii) RBF kernel. *(2 marks)*

Sumeet Agarwal/Sumantra Dutta Roy, EE {sumeet,sumantra}@ee.iitd.ac.in

(c) You are given a small data set with just 4 points in 2D-space. Two positive examples, with coordinates $(1, 4)$ and $(2, 3)$; and two negative examples, with coordinates $(4, 5)$ and $(5, 6)$. Find the weight vector $\mathbf{w}$ (including the bias term w_0) corresponding to the maximum-margin decision boundary learnt by an SVM on this data set. Give justification/derivation for your answer. Also draw a plot showing the data points (with support vectors circled) and the decision boundary learnt. *(3 marks)*

4. Consider a specific variety of neural network, known as *convolutional neural networks*, where multiple weights are constrained to have the same value. For example: Here all links with the same colour/orientation have



the same weight. The nodes in the upper layer all have the same pattern of incoming weights, which is known as a *feature map*.

How must the standard backpropagation algorithm be modified in order to satisfy such constraints when evaluating the derivatives of an error function with respect to the network weights? Please explain your modified algorithm clearly, with proper notation. *(10 marks)*

5. **K-Means, means K***nowing K?*

(a) **Being different: an alternative formulation** Give a very short description of an alternative formulation to the K-Means algorithm, which does not involve K as a parameter. Clearly specify the important parameter here, and outline the basic algorithm. *(4 marks)*

(b) **बड़े धोखे हैं इस राह में, बाबूजी धीरे चलना**?

i. Why does K-Means converge, at all? Formulate an argument involving an objective function, and justify the convergence.

ii. Why may K-Means converge to a local optimum?

iii. Given an example of a 2-D distribution of points for which K-Means will fail *(3+1.5+1.5 marks)*

6. **मिलावट: Gaussian Mixture Models, Seedy Question** How can you use K-Means to seed the EM Algorithm, for a Gaussian mixture? *(3 marks)*

7. **Knowing when to draw a line, and where!** Given a set of 2-D points $[x_i\ y_i]^T$, $0 \leq i \leq n-1$, we define the following terms:
$\mu_x = (1/n)\ \sum x_i$, $\mu_y = (1/n)\ \sum y_i$,
${\sigma_x}^2 = (1/n)\ \sum (x_i - \mu_x)^2$, ${\sigma_y}^2 = (1/n)\ \sum (y_i - \mu_y)^2$,
${\sigma_{xy}}^2 = (1/n)\ \sum (x_i - \mu_x)(y_i - \mu_y)$.
Specifically consider the following form of the equation of a line $y = \frac{x}{a} + \frac{y}{b}$. We want to fit a line to the set of n points in find the 'best-fit' line (in the Least-Squares sense) which minimises the perpendicular distance of all points from the line. Define a suitable error function, and minimise it to find out the values of a and b as functions of μ_x, μ_y, ${\sigma_{xy}}^2$, ${\sigma_x}^2$, and ${\sigma_y}^2$. How is this related to the eigenvalues and eigenvectors of the covariance matrix of the n points? Find these, and explain your answer. *(12 marks)*