EE 657 : Pattern Recognition and Machine Learning

End-Semester Examination

Duration: 3 hours   Date: May 5' 2016   Marks: 35

This paper comprises six questions. Ensure that you present in a logical sequence all important steps that lead to the solution of the problem.

1. Consider an HMM representation of a coin-tossing experiment. You are given two different coins with probabilities

	COIN 1	COIN 2
$P(H)$	0.7	0.4
$P(T)$	0.3	0.6

H: Head T: Tail

and with state transition probability matrix.

	COIN 1	COIN 2
COIN 1	0.2	0.8
COIN 2	0.8	0.2

The initial probabilities for tossing the coins are equally likely.

(a) What is the probability of obtaining the sequence COIN1, COIN2, COIN1 in three successive trials starting from $t = 0$.

(b) What is the probability that the above sequence generates the observation $\mathbf{O} = HTH$.

(c) What is the probability of obtaining the observation $\mathbf{O} = HTH$ from this HMM.

(d) Using the Viterbi algorithm, what is the most likely sequence of coins that can be assigned to $\mathbf{O} = HTH$? Specify the probability as well.

[1 + 1 +3 +3 = 8 marks]

2. Consider a 2 state HMM framework, where the scalar observations from the state 1 and 2 follow their respective GMMs of the form:

State 1: $\frac{7}{10}N(0,1) + \frac{3}{10}N(1,1)$ State 2 : $\frac{3}{10}N(0,1) + \frac{7}{10}N(1,1)$

It is known that the observation sequence $\{3, 1, 1.5\}$ are assigned to {State 2, State 1, State 1 } by the Viterbi algorithm.

(a) What is the responsibility of the first Gaussian component for the observation $x = 3$.

(b) What is the responsibility of the second Gaussian component for the observation $x = 1$.

The normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ in d-dimensions is given by

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{0.5}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

[2 + 1.5 = 3.5 marks]

3. Consider a multi-layer perceptron to train a set of eight-dimensional feature vectors using six `tanh` activation functions at the hidden layer. The characteristics of the ten categories are to be learnt by the network with `sigmoidal` functions at the output layer. Given an instance of the training input feature vector $\mathbf{x}$, the weights of the network are chosen to minimize the error function

$$E = \frac{1}{4}\sum_k (t_k - y_k)^4$$

where t_k and y_k denote the target and predicted values at the k^{th} output node. The target values are either 0 or 1.

(a) Draw the schematic of the Multi layer perceptron for this problem. How many weights are to be learnt in total?

(b) Derive the expression for updation of weight w_{kj}, connecting the j^{th} hidden node to the k^{th} output node.

Note:

For the sigmoidal function $g(x) = \frac{1}{1+e^{-x}}$, the derivative is $g(x)(1 - g(x))$.

For the tanh function $g(x)$, the derivative is $1 - g(x)^2$.

[1.5 + 3 = 4.5 marks]

4. Consider a one-dimensional two-category classification problem with equal priors, where the densities have the form

$$p(x|\omega_i) = \frac{2}{\theta_i}\left(1 - \frac{x}{\theta_i}\right) \qquad 0 \le x \le \theta_i$$

The parameters θ_i for $i = 1, 2$ are positive but unknown.

(i) The following two i.i.d training observations were collected: $\mathbb{D}_1 = \{2, 5\}$ and $\mathbb{D}_2 = \{3, 9\}$ for ω_1 and ω_2, respectively. Using this information, find the maximum-likelihood values for θ_1 and θ_2.

(ii) Based on your answer to part (i), determine the Bayes threshold (decision boundary) x^* for minimum classification error. You may safely assume a zero-one loss function.

(iii) Determine the Bayes error for the classifier in part (ii).

[3+2+2=7 marks]

5. The receiver of a communication system receives a random variable t which is defined as $t = x + \epsilon$ in terms of the input random variable x and the channel noise ϵ. x takes on the values $\frac{1}{4}$ and $\frac{-1}{4}$ with $P[x = \frac{1}{4}] = 0.6$ and $P[x = \frac{-1}{4}] = 0.4$. Let $p(\epsilon)$ denote the pdf of the channel noise and let x and ϵ be independent. The receiver must decide for each received signal whether the transmitted x was $\frac{1}{4}$ or $\frac{-1}{4}$.
If the channel noise is uniform in $(\frac{-1}{2}, \frac{1}{2})$,

(a) Determine and sketch the pdfs $p(t|x = \frac{1}{4})$, $p(t|x = \frac{-1}{4})$ and $p(t)$

(b) Determine the optimal Bayes threshold such that the probability of correct decision is maximised.

(c) What is the average probability of committing an error

[3+2+2=7 marks]

6. You are given a data set $\{\mathbf{x}_i, y_i\}_{i=1}^{N}$ of size N. Each input $\mathbf{x}_i$ is d-dimensional and its corresponding target y_i any real no. ~~takes one of the 2 values (+1 or -1)~~. The formulation of the SVM for the regression problem can be written as :

$$\min_{\mathbf{w},b,\xi,\hat{\xi}} \quad \frac{1}{2}||\mathbf{w}||^2 + \frac{C}{2}\sum_{i=1}^{N}(\xi_i^2 + \hat{\xi}_i^2)$$

subject to the constraints for $i = 1, 2, 3.....N$

$$\mathbf{w}^T\phi(\mathbf{x}_i) + b - y_i \leq \epsilon + \xi_i$$
$$y_i - \mathbf{w}^T\phi(\mathbf{x}_i) - b \leq \epsilon + \hat{\xi}_i$$
$$\xi_i \geq 0, \hat{\xi}_i \geq 0$$

(a) Write down the Lagrangian Function $L(\mathbf{w}, b, \boldsymbol{\xi}, \hat{\boldsymbol{\xi}})$ corresponding to this optimization problem. Here $\boldsymbol{\xi} = \{\xi_1, \xi_2, ...\xi_N\}$, and $\hat{\boldsymbol{\xi}} = \{\hat{\xi}_1, \hat{\xi}_2, ...\hat{\xi}_N\}$.

(b) State the KKT conditions, that are to be satisfied at optimality, together with expressions for weights, biases, violations ; complementary slackness conditions and other constraints (if necessary).

[1+4 = 5 marks]