Practice Problem Set- I

1. Consider a one-dimensional two-category classification problem with priors, $P(\omega_1) = 2/5$ and $P(\omega_2) = 3/5$, where the continuous class conditional densities have the form

$$p(x|\omega_i) = \theta_i \exp^{(-\theta_i x)} \qquad x \geq 0$$

The parameters θ_i for $i = 1, 2$ are positive but unknown. In addition, the misclassification of samples of ω_1 to ω_2 incurs a loss twice to that of ω_2 to ω_1. Correct classifications are assigned zero loss.

   (i) Compute the mean and variance corresponding to the distributions $p(x|\ \omega_1)$ and $p(x|\ \omega_2)$.

   (ii) The following two i.i.d training observations were collected: $\mathbb{D}_1$ ={1, 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.

   (iii) Based on your answer to part (ii), determine the Bayes threshold (decision boundary) x^*. Use this threshold to classify the test point x=4.

   (iv) Write down the expression of the Bayes error for the classifier in part (iii).

2. Repeat the parts 1(iii) and (iv) assuming the case of zero-one loss function, wherein the optimal Bayes classifier gives the minimum probability of error.

3. A researcher has developed a Bayes belief net that expresses the dependencies among the age of a person (A), his or her nationality (B), whether he or she likes cricket (C), and how much he or she watches sports TV during the Cricket World Cup season (D). Use the conditional probability tables to answer the following.

   (a) What is the probability we find a non-Indian who is younger than 30 who likes cricket and watches a "lot" of sports TV?

   (b) Suppose we find someone who is an Indian between 31-40 years of age who watches "some" sports TV. What is the probability that this person likes cricket?

   (c) Suppose we find someone over 40 years of age who "never" watches sports TV. What is the probability that this person likes cricket?

   The Bayes net and associated probability tables are given in the next page.

$P(a_1)$	$P(a_2)$	$P(a_3)$
0.3	0.6	0.1

$P(b_1)$	$P(b_2)$
0.2	0.8

	$P(c_1 \mid a_i, b_j)$	$P(c_2 \mid a_i, b_j)$
a_1, b_1	0.5	0.5
a_1, b_2	0.6	0.4
a_2, b_1	0.6	0.4
a_2, b_2	0.2	0.8
a_3, b_1	0.5	0.5
a_3, b_2	0.1	0.9

	$P(d_1 \mid c_i)$	$P(d_2 \mid c_i)$	$P(d_3 \mid c_i)$
c_1	0.4	0.2	0.4
c_2	0.3	0.5	0.2

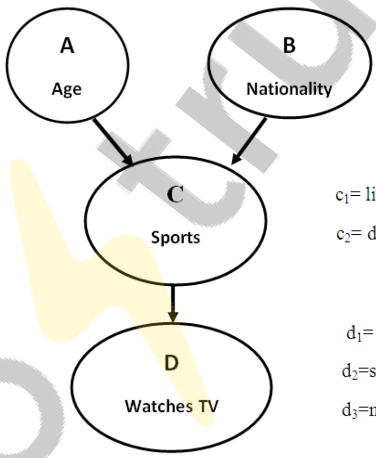
a_1 =0-30 years

a_2 =31-40 years

a_3 = 41-100 years

b_1= Indian

b_2= non-Indian



c_1= likes cricket

c_2= does not like cricket

d_1= a lot

d_2=some

d_3=none