E0 270: MACHINE LEARNING (JAN-APRIL 2022)

PROBLEM SHEET #1
INDIAN INSTITUTE OF SCIENCE

1. Suppose we have two features $x = (x_1, x_2)$ and the two class-conditional densities, $P(x|\omega = 1)$ and $P(x|\omega = 2)$, are 2D Gaussians distributions centered at points $(4, 11)$ and $(10, 3)$ respectively with same covariance $\Sigma = 3I$ (where I is the identity matrix). Suppose the priors are $P(\omega = 1) = 0.6$ and $P(\omega = 2) = 0.4$. Using bayes rule find the two discriminant functions $g_1(x)$ and $g_2(x)$? Derive the equation for decision boundary?
2. In a two class, two dimensional classification task the feature vectors are generated by two normal distributions sharing the same covariance matrix:

$$\Sigma = \begin{bmatrix} 1.1 & 0.3 \\ 0.3 & 1.9 \end{bmatrix}, \ \Sigma^{-1} = \begin{bmatrix} 0.95 & -0.15 \\ -0.15 & 0.55 \end{bmatrix}, \ |\Sigma| = 2$$

   and the mean vectors $\mu_1 = [0, 0]^T$ and $\mu_2 = [3, 3]^T$ respectively. Classify the vector $[1.0, 2.2]^T$ according to bayes classifier? (assume uniform prior) (from https://www.cse.unr.edu/ bebis/CS479/Handouts/)
3. Consider a linear model of the form

$$y(x, w) = w_0 + \sum_{i}^{D} w_i x_i$$

   together with a sum-of-squares error function of the form

$$E_D(w) = 0.5 * \sum_{n=1}^{N} [y(x_n, w) - t_n]^2$$

   Now suppose that Gaussian noise η_i with zero mean and variance σ^2 is added independently to each of the variables x_i. By making use of $\mathcal{E}[\eta_i] = 0$ and $\mathcal{E}[\eta_i \eta_j] = \delta_{ij}\sigma^2$, show that minimizing E_D averaged over the noise distribution is equivalent to minimizing the sum-of-squares error for noise-free input variables with the addition of a weight-decay regularization term, in which the bias parameter w_0 is omitted from the regularizer. (Bishop 3.4)
4. A student needs to achieve a decision on which courses to take, based only on his first lecture. From previous experience he knows the following

Quality of Course	Good	Fair	Bad
Probability ($P(\omega_j)$)	0.2	0.4	0.4

These are the priors. The student also knows the class conditionals

$P(x\|\omega_j)$	Good	Fair	Bad
Interesting Lecture	0.8	0.5	0.1
Boring Lecture	0.2	0.5	0.9

He also knows the loss function for the actions

$\lambda(a_i\|\omega_j)$	Good	Fair	Bad
Taking the course	0	5	10
Not taking the course	20	5	0

What is the optimal decision by minimizing the risk if he found the lecture for a course interesting? ($http://www.cs.haifa.ac.il/\ rita/ml\_course$)

5. In many pattern classification problems one has the option either to assign to one of *c classes*, or to *reject* it as being unrecognizable. If the cost for rejects is not too high, rejection may be a desirable action. Let

$$\begin{aligned}\lambda(\alpha_i|\omega_j) &= 0\ if\ i=j\ \&\ i,j=1,....,c\\ &= \lambda_r\ i=c+1\\ &= \lambda_s\ otherwise\end{aligned}$$

where λ_r is the loss incurred for choosing the $(c+1)$th action, rejection, and λ_s is the loss incurred for making a substitution error. Show that the minimum risk is obtained if we decide ω_i if $P(\omega_i|x) \geq P(\omega_j|x)$ for all j and if $P(\omega_i|x) \geq 1-\lambda_r/\lambda_s$, and reject otherwise. What happens if $\lambda_r = 0$? What happens if $\lambda_r > \lambda_s$?

6. Let the conditional densities for a two-category one-dimensional problem be given by the Cauchy distribution (Duda Hart Prob 7 & 8)

$$P(x|\omega_i) = \frac{1}{\pi b}\frac{1}{1+(\frac{x-a_i}{b})^2} \text{ for } i=1,2.$$

   (a) Find the minimum error decision boundary for 0-1 loss assuming uniform prior ?
   (b) Show the the minimum probability of error is given by

$$P(error) = \frac{1}{2} - \frac{1}{\pi}\tan^{-1}|\frac{a_1-a_2}{2b}|.$$

7. Find the discriminant function for two class classification problem where the feature vectors are binary and independent given the class? (Assume 0-1 loss)
8. Let $\omega_{max}(x)$ be the state of nature for which $P(\omega_{max}|x) \geq P(\omega_i|x)$ for all i, $i=1,...,c$. (Duda Hart Prob 12)
   (a) Show that $P(\omega_{max}|x) \geq \frac{1}{c}$?

(b) Show that for the minimum-error-rate decision rule the average probability of error is given by

$$P(error) = 1 - \int P(\omega_{max}|x)p(x)dx.$$

9. Consider a simple linear regression model in which y is the sum of a deterministic linear function of x, plus random noise η.

$$y = wx + \eta$$

where x is the real-valued input; y is the real-valued output; and w is a single real-valued parameter to be learned. Here η is a real-valued random variable that represents noise, and that follows a Gaussian distribution with mean 0 and standard deviation σ; that is, $\eta \sim N(0, \sigma)$.
Find the MAP estimate for parameter w assuming a gaussian prior with variance τ
$http : //www.cs.cmu.edu/\ tom/10701\_sp11/midterm.pdf$