E1 213 Pattern Recognition and Neural Networks

Practice Problems: Set 1

1. Consider a 2-class problem with one dimensional feature space. Let the class conditional densities be: $f_0(x) = e^{-x},\ x > 0$, and $f_1(x) = 1/2a,\ x \in [-a,\ a],\ a > 0$. The prior probabilities are equal. Assume we are using 0–1 loss. Find the Bayes classifier. For the case when $a = 0.25$, find Bayes error.

2. Consider a 2-class PR problem with feature vectors in $\Re^2$. The class conditional density for class-I is uniform over $[1, 3] \times [1, 3]$ and that for class-II is uniform over $[2, 4] \times [2, 4]$. Suppose the prior probabilities are equal and we are using 0–1 loss. Consider line given by $x + y = 5$ in $\Re^2$. Is this a Bayes Classifier for this problem? Is Bayes Classifier unique for this problem? If not, can you specify two different Bayes classifiers? Suppose the class conditional densities are changed so that the density for class-I is still uniform over $[1, 3] \times [1, 3]$ but that for class-II is uniform over $[2, 5] \times [2, 5]$. Is the line $x + y = 5$ a Bayes classifier now? If not, specify a Bayes classifier now. Is the Bayes classifier unique now? For this case of class conditional densities, suppose that wrongly classifying a pattern into class-I is 10 times more expensive than wrongly classifying a pattern into class-II. Now, what would be a Bayes classifier?

3. Consider a general K-class problem with a general loss function. Let $h(X)$ denote the output of the classifier on X. Let $R(\alpha_i|X)$ denote the expected loss when classifier says α_i and conditioned on X. That is, $R(\alpha_i|X) = E[L(h(X), y(X))|h(X) = \alpha_i, X]$, where, as usual, $y(X)$ denotes the 'true class'. We had only considered deterministic classifiers where h is a function that assigns a unique class label for any given X. Suppose we use a stochastic classifier, h, which, given X, outputs α_i with probability $p_h(\alpha_i|X)$. (Note that we would have $p_h(\alpha_i|X) \geq 0$ and $\sum_i p_h(\alpha_i|X) = 1$). For this classifier, show that the risk is given by
$$R(h) = \int \left[\sum_{i=1}^{K} R(\alpha_i|X) p_h(\alpha_i|X) \right] f(X)\ dX$$
where $f(X)$ is the density of X. Using the above expression, find the best choice of values for all the $p_h(\alpha_i|X)$ and hence conclude that we do not gain anything by making the classifier stochastic.

4. Let $x_1, \cdots, x_n$ be *iid* data drawn according to exponential density with parameter λ. Derive the ML estimate for λ. (The exponential density is given by $f(x) = \lambda e^{-\lambda x},\ x > 0$).

5. Suppose X is uniformly distributed over $[0,\ \theta]$, with $\theta > 0$ being the unknown parameter. (The uniform density is given by $f(x) = 1/\theta$, if $0 \leq x \leq \theta$ and $f(x) = 0$ otherwise). Suppose we have three *iid* samples, $1.75, 0.5, 2.2$. What is the value of the likelihood function $L(\theta|\mathcal{D})$ for (i). $\theta = 10$, (ii). $\theta = 1.9$? Now consider the general case where we represent the three *iid* samples as x_1, x_2, x_3. Plot the likelihood function (that is, plot $L(\theta|\mathcal{D})$ versus θ). Now, consider the case where we have n *iid* samples, what is the ML estimate for θ.

6. Suppose you have n samples from a normal density with mean μ and variance 1. You estimated the mean using the sample mean. Then you discover that your friend had m samples from the same density and has estimated the mean using sample mean. How should you combine your estimates to get a better estimate.