CS 308-644 B Pattern Recognition - Exercise 5

1. Maximum Likelihood Estimation of Parameters

(a) Let x have an exponential density function: $p(x|\theta) = \theta e^{-\theta x}$ for non-negative x and zero elsewhere. Plot $p(x|\theta)$ as a function of x for a fixed value of the parameter $\theta=1$. Plot $p(x|\theta)$ as a function of θ ($0 < \theta < 5$) for a fixed value $x=2$.

(b) Now suppose that n data samples $x_1,...,x_n$ are drawn independently according to the density $p(x|\theta)$. Prove or disprove that the maximum likelihood estimate of θ is given by:

$$\hat{\theta} = \frac{1}{\frac{1}{n}\sum_{k=1}^{n} x_k}$$

2. The 1-Nearest Neighbor and Bayes Error Rates

(a) It is easy to see that the 1-nearest-neighbor error rate P_e(1-NN) with C pattern classes can equal the Bayes error rate P_e(B) if P_e(B) = 0 (the best case) or if P_e(B) = (C-1)/C (the worst case). One may ask if there exist problems for which P_e(1-NN) = P_e(B) in between these two extremes. Show that for the one-dimensional case with equal a priori probabilities and the following class-conditional *pdf*'s, P_e(B) = R.

$$p(x/C_i) = \begin{cases} 1\,for & 0 \le x \le \dfrac{CR}{C-R} \\ 1\,for & i \le x \le i+1-\dfrac{CR}{C-R} \\ 0 & elsewhere \end{cases}$$

(b) Show that for the problem in (a) P_e(1-NN) = R also.

3. The 2-Nearest-Neighbor Decision Rule with a Reject Option

Consider the 2-Nearest Neighbor decision rule with two classes in which for a given feature vector X, its two nearest neighbors, X_1 and X_2, are determined. Furthermore, assume that if X_1 and X_2 belong to Class 1 then Class 1 is chosen, if X_1 and X_2 belong to Class 2 then

Class 2 is chosen and finally, if X_1 and X_2 belong to different classes then X is *rejected.*

Show that, under reasonable assumptions (list them), as the sample size approaches infinity, the probability of rejection, P_r(2-NN) = P_e(1-NN), i.e., the probability of error of the 1-Nearest Neighbor decision rule. Furthermore, show that the probability of error, P_e(2-NN) $\leq P_e$(B), the Bayes optimal probability of error.

4. *The k*-Nearest-Neighbor Decision Rule

Consider the *k*-nearest neighbor rule in a 2-class problem with equal *a priori* probabilities. Assume that *n* data points are generated independently in a random draw from the mixture density *p*(X). Assume further that the class-conditional probability densities $p(X/C_1)$ and $p(X/C_2)$ are uniform within unit hyperspheres a distance of ten units apart.

(a) Show that if *k* is odd the average probability of error is given by:

$$P_n(e) = \frac{1}{2^n}\left(\sum_{j=0}^{(k-1)/2}\binom{n}{j}\right)$$

(b) Show that for this case the *1*-nearest neighbor rule has a lower error probability than the *k*-nearest neighbor rule for all *k*>*1*.

5. The *j-th* Nearest-Neighbor Decision Rule

Consider the following decision rule (referred to as the *j-th nearest-neighbor rule*) for *j* even. Classify an unknown feature vector X into class C_1 if the distance (Euclidean metric) from X to the *j-th* closest vector in C_1 is less than the distance from X to the *j-th* closest vector in C_2. Otherwise classify X into C_2.

(a) Draw the discriminant functions for a 1-dimensional example with *j*=2.

(b) How is this rule related to the *k*-NN (*k*-nearest-neighbor) rule when *j* = (*k*+1)/2?