University of Tehran

School of Electrical and Computer Engineering

Pattern Recognition

Assignment 2_1

Due Date: 31st of Farvardin

Corresponding TAs:

Niloufar Shahdoust - n.shahdoust@ut.ac.ir

Shahrzad Khorsand - shahrzad.khorsand@gmail.com

Marjan Shahi - marjan.shahi71@gmail.com

Esfand 97

PROBLEM 1

Let $\{x_k\}$, k=1, 2, …, N denote independent training samples from one of the following densities. Obtain the Maximum Likelihood estimate of θ in each case.

a. $f(x_k;\theta) = \theta \exp(-\theta x_k)$ $\quad x_k \geq 0\,,\theta > 0 \quad \mathit{Exponential\ Density}$

b. $f(x_k;\theta) = \dfrac{x_k}{\theta^2}\exp\left(-\dfrac{x_k^2}{2\theta^2}\right)$ $\quad x_k \geq 0\,,\theta > 0 \quad \mathit{Rayleigh\ Density}$

c. $f(x_k;\theta) = \sqrt{\theta}x_k^{\sqrt{\theta}-1}$ $\quad 0 \leq x_k \leq 1, \theta > 0 \quad \mathit{Beta\ Density}$

d. $f(x_k;\theta) = \theta^2 x_k \exp(-\theta x_k)$ $\quad x_k > 0\,,\theta > 0 \quad \mathit{Exponential\ Density}$

PROBLEM 2

The figure below illustrates two-dimensional training samples from two classes; where, × and ● represent classes 1 and 2, respectively. Classify the test sample $(0.5, 0)^T$ using:

a. Parzen window approximation using the uniform window function:

$$\varphi(u_1, u_2) = \begin{cases} 1 & |u_i| \leq 0.5, i = 1,2 \\ 0 & \mathit{otherwise} \end{cases}$$

i. v_n is a square with a side of h_N = 2

ii. v_n is a circle with a radius of r_N = 1

b. Unbiased K Nearest Neighbor approximation with k = 3, and

i. Euclidean distance: $d(x,y) = \sqrt{\Sigma_i (x_i - y_i)^2}$

ii. City Block distance: $d(x,y) = \max_i |x_i - y_i|$

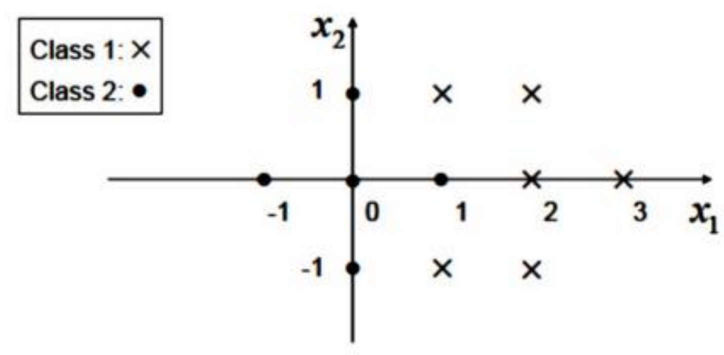

Figure 1

Hint: Any point on the border of the volume is considered an inside point.

PROBLEM 3

Let x have a uniform density

$$f_X(x|\theta) \sim U(0,\theta) = \begin{cases} \frac{1}{\theta} & 0 \le x \le \theta \\ 0 & othetwise \end{cases}$$

a. Suppose that n samples $D = \{x_1, x_2, \dots, x_n\}$ are drawn independently according to $f_X(x|\theta)$. Show that the maximum likelihood estimate for θ is $\max[D]$, i.e., the value of the maximum element in D.
b. Suppose that $n = 6$ points are drawn from the distribution and the maximum value of which happens to be $\max_k x_k = 0.7$. Plot the likelihood $f_X(D|\theta)$ in the range $0 \le \theta \le 1$.

Explain in the words why you do not need to know the values of the other five points.

PROBLEM 4

In case of Gaussian distribution using EM estimation, we have

$$f(x_k;\theta\,|\,j) = \frac{1}{2\pi\sigma_j^{0.5}} e^{-\frac{|x_k-\mu_j|^2}{2\sigma_j^2}}$$

So

$$Q(\theta;\hat{\theta}(t)) = \sum_{k=1}^{N}\sum_{j=1}^{J} p(j;\hat{\theta}(t)\,|\,x_k)\left\{-\frac{|x_k-\mu_j|^2}{2\sigma_j^2} - \frac{l}{2}\ln(2\pi\sigma_j^2) + \ln(p_j)\right\}$$

a. Find iterative estimation formulas for μ_j, σ_j^2 and p_j by maximizing $Q(\theta;\hat{\theta}(t))$.

b. How can $p(j;\hat{\theta}(t) \mid x_k)$ be computed at each iteration? Explain and derive formula.

c. Specify all the quantities that $\hat{\theta}(t)$ is an explicit function.

PROBLEM 5

Consider a D-dimensional Gaussian random variable x with distribution $N(x \mid \mu, \Sigma)$ in which, the covariance Σ is known and for which we wish to infer the mean μ from a set of observations $X = \{x_1, \ldots, x_N\}$. Given a prior distribution $p(\mu) = N(\mu \mid \mu_0, \Sigma_0)$, find the corresponding posterior distribution $p(\mu \mid X)$.

NOTES

1. Please make sure you reach the deadline because there would be no extra time available.
2. Late policy would be as bellow:
   - Every student has a budget for late submission during the semester. This budget is two weeks for all the assignments.
   - Late submission more than two weeks may cause lost in your scores.
3. Analytical problems can be solved on papers and there is no need to type the answers. The only thing matters is the quality of your pictures. Scanning your answer sheets is recommended. If you are using your smartphones you may use scanner apps such as CamScanner or google drive application.
4. Simulation problems need report as well as source codes and results. This report must be prepared as a standard scientific report.
5. You have to prepare your final report including the analytical problems answer sheets and your simulation report in a single pdf file.
6. Finalized report and your source codes must be uploaded to the course page as a ".zip" file (not ".rar") with the file name format as bellow:

   PR_Assignment #[Assignment Number]_Surname_Name_StudentID.zip

7. Plagiarisms would be strictly penalized.
8. *You may ask your questions from corresponding TAs.*