University of Tehran

School of Electrical and Computer Engineering

Pattern Recognition

Assignment 2

Due Date:

Azar 9th

Corresponding TAs:

Hossein Basafa - hbasafa@ut.ac.ir

Mohammad Dehghan Rouzi - dehghanr.mohammad@ut.ac.ir

Mahyar Khordeforush - m.khordeforush@ut.ac.ir

Suorena Saeedi - suorena.saeedi@ut.ac.ir

Aban 98

PROBLEM 1 – Posterior Estimation (MAP)

I. Consider a Gaussian distribution of known covariance Σ and unknown mean μ. Assume further that μ is normally distributed with mean m_0 and covariance Σ_0. We have calculated the MAP estimator for μ and then we have applied a linear transformation on our coordinates ($x' = Ax$). Does the prescribed MAP estimator give an appropriate estimate of μ'? Explain.

II. The random variable x is normally distributed as $N(\mu, \sigma^2)$. σ is known and μ is the unknown parameter that has Rayleigh pdf.

$$p(\mu) = \frac{\mu \exp(-\frac{\mu^2}{2\sigma_\mu^2})}{\sigma_\mu^2}$$

Show that the MAP estimator of μ is given by:

$$\hat{\mu}_{MAP} = \frac{Z}{2R}(1 + \sqrt{1 + \frac{4R}{Z}}$$

Where

$$Z = \frac{1}{\sigma^2}\sum_{k=1}^{N} x_k \quad , \quad R = \frac{N}{\sigma^2} + \frac{1}{\sigma_\mu^2}$$

PROBLEM 2 – Non-Parametric Estimations

I. a. For the set of samples X = {-7,-5,-4,-3,-2,0,3,4,5,7} taking them in sequence, find the Parzen window estimate to $P_j(x)$ for a rectangular window. Use $h_j = 1/\sqrt{j}$. Sketch the results as a function of x for $j = 1,4,11$.

b. Suppose $h_j = h/\sqrt{j}$ where h is a constant chosen at the beginning. Comment on the shape of $P_j(x)$ for various choices of h.

II. For the following samples in one dimensional space, give the values of the K-nearest neighbor estimate $P_j(x)$, for $j = 16$ and $k_j = \sqrt{j}$, at $x = 2, x = 4, x = 6, x = 8$ and $x = 10$.

$$X = \{0, 1, 3, 4.5, 5.5, 6, 6.5, 7, 7.2, 7.5, 8.0, 8.8, 9.2, 9.3, 11, 13\}$$

PROBLEM 2-1 – Non-Parametric Estimations

The smoothing parameter (aka bandwidth), h, plays an important role in kernel density estimation. A good criterion for selecting h is one that minimizes the mean-squared error. For a univariate Gaussian kernel, $h^* \approx 1.06\,\hat{\sigma}\,N^{-\frac{1}{5}}$ is the optimal choice, where $\hat{\sigma}$ is the estimate of the standard deviation of the samples and N is the number of samples.

a. Write a function, $randgen(f;\ N)$ that generates N i.i.d samples from a given probability density function f. You may find the built-in pyhthon functions for random number generation.
b. For $N = \{10;\ 100;\ 1000\}$ generate N independent samples from an exponential distribution with $\lambda = 1$. ($f(x) = e^{-x}[x \geq 0]$, where [.] is the indicator function).
c. Compute the sample standard deviation, $\hat{\sigma}$, without making any prior assumptions on the distribution (i.e., DO NOT assume that the data are drawn from an exponential distribution). For each N, estimate the optimal bandwidth, $h^*(N)$.
d. Estimate the pdf using kernel density estimation with a Gaussian kernel for each N, under three different bandwidth settings: $\left\{\frac{1}{3} \times h^*(N);\ h^*(N);\ 3 \times h^*(N)\right\}$.
e. Summarize your results by plotting the pdf estimates. You need to have 9 plots overall. Comment on the influence of h, N, and the kernel itself on the pdf estimates.

PROBLEM 3 – ML Estimation

I. Briefly describe on which condition:
   a. MAP and ML are exactly the same.
   b. Naive Bayes and MAP are corresponding together.

II. 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. (note that the $u(x)$ is the step function)

a. $f(x_k;\theta) = \frac{1}{\sigma x_k \sqrt{2\pi}} \exp\left(-\frac{(\ln x_k - \theta)^2}{2\sigma^2}\right)$ $\quad x_k > 0 \quad$ *LogNormal Density*

b. $f(x_k;\theta) = \frac{1}{\Gamma(\alpha)} \theta^\alpha x_k^{\alpha-1} \exp(-\theta x_k)$ $\quad x_k \geq 0, \theta > 0 \quad$ *Gamma Density*

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

d. $f(x_k;\theta) = \theta^2 x_k \exp(-\theta x_k)\, u(x_k)$ $\quad$ *Erlang Density*

PROBLEM 4 – Expectation Maximization

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

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

d) Suppose we seek to estimate θ describing a multidimensional distribution from data D, some of whose points are missing features. Consider an iterative algorithm in which the maximum likelihood value of the missing values is calculated, then assumed to be correct for the purposes of restinating θ and iterated.

    I. Is this always equivalent to an Expectation-Maximization algorithm, or just a generalized Expectation-Maximization algorithm?

    II. If it is an Expectation-Maximization algorithm, what is $Q(\theta,\hat{\theta}(t))$, as described by $Q(\theta;\ \hat{\theta}(t))=\epsilon_{D_b}[ln\,p(D_g,D_b;\ \theta)|D_g;\ \hat{\theta}(t)]$? [1]

PROBLEM 5 (25% Bonus)

Assume that you want to model the future probability that your dog is in one of three states given its current state considering its health and sickness. Compute the probability of the observation sequences PESESEPEPSPEPS and SESPSEPSPSESPE using **Forward Algorithm** in the following HMM. Which one is more likely?

---

[1] Given such a candidate θ, this equation calculates the likelihood of the data, including the unknown feature D_b marginalized with respect to the current best distribution, which is described by $\hat{\theta}(t)$; and the D_g is the known feature that $D\ =\ D_g\ \cup D_b$. for further information about this equation refer to the Eq 78. in Duda Note book chapter 3.

S	Sleeping
E	Eating
P	Pooping

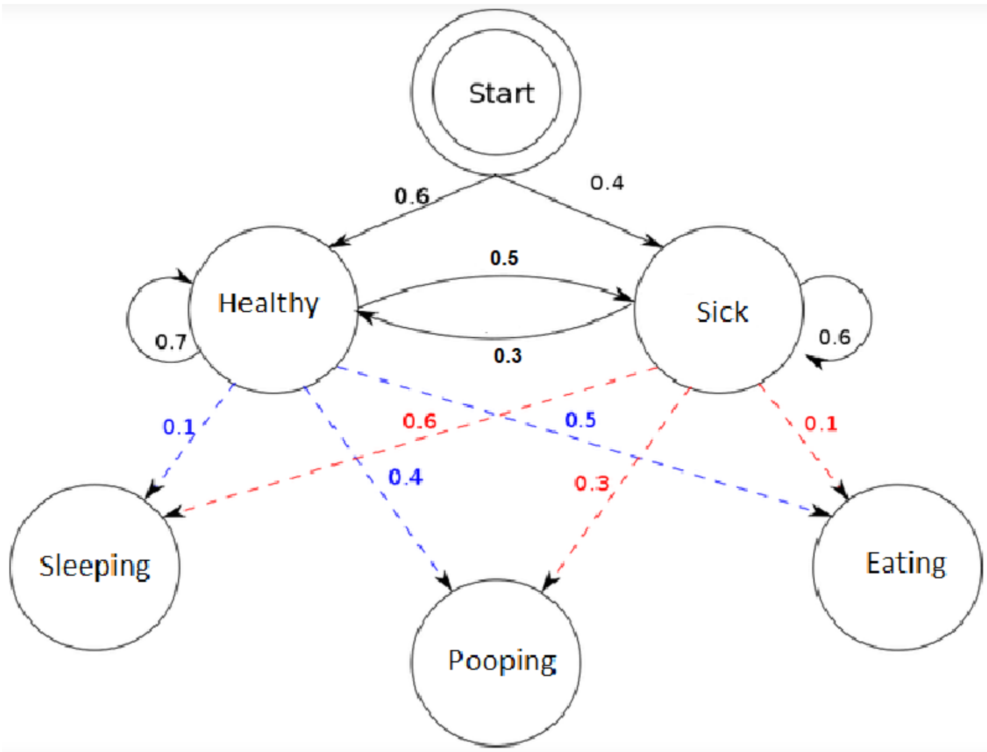

Figure 2- A hidden Markov model for relating numbers of dog behavior to the health level of it (Sick(S) or Healthy(H) are the hidden variables).

** Figure 2 shows a sample HMM for the dog typical tasks when it's healthy or sick. The two hidden states (H and S) correspond to Health and Sickness of dog, and the observations are sleeping, pooping and eating.

PROBLEM 6 (25% Bonus)

I. Show that for a Markov Chain, $P(q_t|q_{t+1}, \dots, q_T) = P(q_t|\, q_{t+1})$.

II. Consider the following 2-state Hidden Markov Models where both states have two possible output symbols A and B.

Model 1:
Transition probabilities: $a_{11} = 0.6,\ a_{12} = 0.4,\ a_{21} = 0.0,\ a_{22} = 1.0$
(a_{ij} is the probability of going from state i to state j)
Output probabilities: $b_1(A) = 0.45,\ b_1(B) = 0.55,\ b_2(A) = 0.5,\ b_2(B) = 0.5$
Initial probabilities: $\pi_1 = 0.5,\ \pi_2 = 0.5$

Model 2:
Transition probabilities: $a_{11} = 0.15,\ a_{12} = 0.85,\ a_{21} = 0.0,\ a_{22} = 1.0$
Output probabilities: $b_1(A) = 0.4,\ b_1(B) = 0.6\ ,\ b_2(A) = 0.7,\ b_2(B) = 0.3$
Initial probabilities: $\pi_1 = 0.5,\ \pi_2 = 0.5$

a. Sketch the state diagram for two models.
b. Which model is more likely to produce the observation sequence {A, B, A}?

Hint: *For more information on Hidden Markov Models refer to the* ***third chapter*** *of* **this** *reference.*

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.