Practice Problem 1 [Review of Probability Theory]

CSE 421 : Machine Learning

1. **[Probability Density Function]** Suppose that the pdf of a random variable X is as follows

$$f(x) = ce^{-x^2/2}, for - \infty < x < \infty$$

(i) Find the value of the constant c. [hint: $\int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi}$]

(ii) Find the value of P(1 < X < 2)
(iii) Find E[X] and E[X^2]
(iv) Find Var[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2

2. **[Joint Probability Distribution]** X and Y are discrete random variables which are "jointly" distributed with the following probability mass function $p_{X,Y}(x,y)$:

		X		
		-1	0	1
Y	1	1/18	1/9	1/6
	0	1/9	0	1/6
	-1	1/6	1/9	1/9

That is, for example, P(X=1, Y=0) = 1/6.

(i) The sum of the entries of the first column is P(X=-1, Y=1)+P(X=-1, Y=0)+P(X=-1, Y=1) = 6/18 = 1/3. This is equal to P(X=-1) and called the *marginal probability* of X = -1. Show that P(X=-1) can be found also using the total probability theorem. For that you will need to compute P(Y=-1), P(Y=0) and P(Y=1). To compute these use the row sums.

(ii) Show whether X and Y are independent or not (for that you will need to show the independence holds for all possible pairs of values of X and Y).

3. **[Joint Probability Distribution]** Suppose X and Y are continuous random variables, and the joint density function is given by f(x,y)

$$f(x,y) = \begin{cases} \dfrac{e^{-x/y}e^{-y}}{y}, if\ 0 < x < \infty, 0 < y < \infty \\ 0, otherwise \end{cases}$$

(i) Find the marginal probability functions of X and Y.
(ii) Are X and Y independent or not?
(iii) Find P[X > 1 | Y = y]
(iv) Find E[X | Y = 1]

4. **[Bayes Theorem]** Suppose that Ali can decide to go to work by one of three modes of transportation: car, bus, or autorickshaw. Because of high traffic and unavailability of autorickshaw in the morning, if he decides to go by autorickshaw, there is a 40% chance he will be late. If he goes by bus, the probability of being late is 25%. Using a car he is almost never late, with a probability of only 5%, but it is more expensive than the bus.
(i) Suppose that Ali is late one day, and his boss wishes to estimate the probability that he drove to work that day by car. Since he does not know which mode of transportation Ali usually uses, he gives the equal prior probability to each of the three possibilities. What is the boss' estimate of the probability that Ali commuted by car to work?

(ii) Suppose that a coworker of Ali's knows that he almost always takes the autorickshaw to work, never takes the bus, but sometimes, 10% of the time to be exact he takes the car. What is the coworker's estimate of the probability that Ali commuted by car work to that day?

5. **[Bayesian Decision Theory]** Suppose the prior probability of a fish being salmon [i.e., P(x = Salmon)] is 1/3 and that of bass is 2/3. Also assume we have two features, length and width computed for each fish. Although these are discrete features we have discretized them. The length feature is discretized into 3 levels (small, medium, and long) in the following way:

```
if length < 10in, length = small
        else if 10inch <= length < 20 inch, then length = medium
                else length = long
```

Similarly width has been discretized into two levels thin and thick using the threshold value of 5 inch.

Now the class **conditional probabilities of the discretized feature values** are given as follows in the following tables:

Distribution of length of Salmon and bass

length	Salmon	Bass
small	1/4	1/5
medium	1/2	3/5
long	1/4	2/5

that is, for example, P(length = small | x = Salmon) = 1/4, P(length = medium | x = Bass) = 3/5

and the distribution of width values according to fish category is given below:

width	Salmon	Bass
thin	1/3	3/5
thick	2/3	2/5

Now, suppose a fish x has length = 15 inch and width = 6 inch.

(i) use the length feature only to compute

a. the posterior probability of the fish being a Salmon

b. the posterior probability of the fish being a Bass

c. the class the fish belongs to according to Bayesian decision rule

(ii) use the width feature only to compute

a. the posterior probability of the fish being a Salmon

b. the posterior probability of the fish being a Bass

c. the class the fish belongs to according to Bayesian decision rule

(iii) Suppose the length feature and the width feature is independent of each other, that is P(length = small, width = thick) = P(length = small)*P(width = thick). Use this independence assumption and use both features to compute

a. the posterior probability of the fish being a Salmon

b. the posterior probability of the fish being a Bass

c. the class the fish belongs to according to Bayesian decision rule