National University of Computer and Emerging Sciences, Lahore Campus

Course:	Statistical Pattern Recog & Learning	Course Code:	CS 557
Program:	MS(Computer Science)	Semester:	Fall 2016
Duration:	180 Minutes	Total Marks:	60
Paper Date:	29-Dec-16	Weight	40
Section:	ALL	Page(s):	5
Exam:	Final exam Solutions	Roll No:	

Instruction/Notes: **FINAL EXAM SOLUTIONS**

QUESTION 1 **(Marks: 2+2+2+2+2)**

Suppose we have the following specifications of an HMM with 4 states and 3 observation symbols x,y,z. A is the transition probability matrix, B is the emission probability matrix and π is the initial probability vector. (time starts at t=1)

$$A=\begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 1/2 & 0 & 1/4 \\ 0 & 0 & 1/4 & 3/4 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad \begin{array}{ccc} x & y & z \end{array} \begin{pmatrix} 1/3 & 2/3 & 0 \\ 1/3 & 0 & 2/3 \\ 0 & 1/3 & 2/3 \\ 0 & 0 & 1 \end{pmatrix} \qquad \pi = [1\ 0\ 0]^T$$

a. How many possible paths lead to state S_3 at time t=5. List them down.
   4 paths:
   1. S1 S1 S1 S1 S3
   2. S1 S1 S1 S3 S3
   3. S1 S1 S3 S3 S3
   4. S1 S3 S3 S3 S3

b. What is the probability of being in state S_3 at time t=5? _______$4*(1/4)^4$__________

c. What is the probability of being in state S_4 at time t=2? _______1/4______________

d. What is the state sequence that generates: xxyzx __________S1 S1 S1 S2 S2_______

e. What is $P(xxyzxyz|q_5=S_1)$? Give a one line explanation.
   Zero as state S1 is not possible at time t=5

f. Fill the table for alpha values (forward variable) when using the forward algorithm for generating xxyzx

	S1	S2	S3	S4
x (t=1)	1/3	0	0	0
x (t=2)	1/36	1/36	0	0
y (t=3)	1/216	0	1/432	0
z (t=4)	0	1/1296	1/216*1/4*2/3+ 1/432*1/4*2/3=.00157	1/216*1/4*1 + 1/432*3/4*1 =.00289
x (t=5)	0	1/7776	0	0

QUESTION 2 **(MARKS: 3)**

Give a possible mapping of the following points *a*,*b*,*c* in 2D that preserves Manhatten distance (City block distance): *a*: (1,1,2,1), *b*: (0,0,1,1), *c*: (1,0,0,0). Indicate *a,b,c* in the new mapping.

Roll Number: ___________________

ANSWER: *a* = (0,0) *b* = (3,0) *c* = (2,2)

QUESTION 3 (MARKS: 3)

Given the following training data:

Point →	0	1	2	3	4	5	6
Label →	+1	+1	-1	+1	+1	-1	-1

and the following test data:

Point →	0.5	1	2	5
Label →	+1	+1	-1	-1

Give the balanced error rate of the test data in terms of a percentage when 3-nearest neighbor is applied.

ANSWER: BER = 1/2*(0/2+1/2)*100 = 25 %

QUESTION 4 (Marks: 2+2+2)

Suppose we use the following non-parametric density function to estimate the density of any point in R^2:

$$p(x)=\frac{1}{\mathrm{Nh}}\sum_{t=1}^{N} w\left(\frac{x-x^t}{h}\right)$$, Where N = total points and h=2

The weight function w is given by:

$w(\boldsymbol{x}) = ||\boldsymbol{x}||_1$ if $||\boldsymbol{x}||_1 <= 2$ (here $||\mathbf{x}||_1$ is the L_1 norm of $\mathbf{x}$)

$w(\boldsymbol{x}) = 0$ otherwise

We have the following training points:

x_1 →	+1	4	-4	3	4	1
x_2 →	1	-2	2	3	0	2
Class →	-1	-1	-1	-1	+1	+1

Given the training points, determine the following:

a. p($\boldsymbol{x}$=(1,1) | class = +1)

1/2*1/2*(2+1/2) = 5/8

b. p($\boldsymbol{x}$=(1,1) |class = -1)

1/8*(0+0+0+0+2) = 1/4

c. Determine the classification of (1,1) using MAP and the above density function

it belongs to +1 class.

QUESTION 5 (Marks: 4+4)

a. Suppose that the probability of getting a job when being a good student (CGPA>= 3.0) is 0.9 and in case the student's performance is not very good (CGPA < 3.0) then the probability of getting a job is 1/*d*. In a situation where a student gets a job, what is the probability that the student is a good student? The probability of getting a job is assumed to be *q*. Give an expression in terms of *q* and *d*.

We need P(good student) which can be computed from the above
P(good) = (q-1)/(.9d-1)
P(good|job) = 0.9/1*(q-1)/(.9d-1)

b. For the above scenario make the confusion matrix for the predictions regarding getting a job when the student is a good student and find the precision and recall.

Roll Number: ____________________

	good	~good
job	.9	1/d
~job	.1	1-1/d

recall = 0.9
precision = 0.9/(.9+1/d) = .9d/(.9d+1)

QUESTION 6 (MARKS 3)

Given the following data in 2D:

x_1 →	0	0	0	1	1	1
x_2 →	2	1	0	2	1	0
label	+1	+1	+1	-1	-1	-1

Using LDA, give **w** that separates the two labels.

ANSWER **w** = (1,0)

QUESTION 7 (MARKS 3)

Given the following data for 3 attributes A,B,C

A →	0	1	0	1	0
B →	0	1	1	1	1
C →	1	0	1	0	1

a. Give P(A=0 and B=1 and C=0) when Bernoulli distribution is applied to the above data

ANSWER: 3/5*4/5*2/5

QUESTION 8 (Marks: 3.5+3.5+3)

Given the following initial SOM grid **at iteration 0**:

	0	1	2
0	$w_{00}=[-1,-1]^T$	$w_{01}=[1,1]^T$	$w_{02}=[1,5]^T$
1	$w_{10}=[0,0]^T$	$w_{11}=[2,2]^T$	$w_{12}=[0,0]^T$
2	$w_{20}=[4,1]^T$	$w_{21}=[1,1]^T$	$w_{22}=[3,1]^T$

Suppose the learning rate =1 and the neighborhood function e is given by:
e(k,l,i,j) = 0.5 if $|k-i|<1$
e(k,l,i,j) = 0, otherwise
In the above (i,j) are the coordinates of the best matching unit and (k,l) are the coordinates of the unit to update. Use **Euclidean distance** to determine BMU and show working for all parts along with the mathematical expressions you are using.

a. Show the updated grid when the training point (3,3) is presented to the above grid **at iteration 0**.

b. Show the updated grid when the training point (-1,0) is presented to the grid **at iteration 0**.

c. Show the mapping of the test points (4.5,1), (1,4.5),(-1,-5) when the grid **at iteration 0** is used.

Roll Number: ____________________

a.

	0	1	2
0	$w_{00}=[-1,-1]^T$	$w_{01}=[1,1]^T$	$w_{02}=[1,5]^T$
1	$w_{10}=[3/2,3/2]^T$	$w_{11}=[3,3]^T$	$w_{12}=[3/2,3/2]^T$
2	$w_{20}=[4,1]^T$	$w_{21}=[1,1]^T$	$w_{22}=[3,1]^T$

b. If BMU is (0,0)

	0	1	2
0	$w_{00}=[-1,0]^T$	$w_{01}=[0,1/2]^T$	$w_{02}=[0,5/2]^T$
1	$w_{10}=[0,0]^T$	$w_{11}=[2,2]^T$	$w_{12}=[0,0]^T$
2	$w_{20}=[4,1]^T$	$w_{21}=[1,1]^T$	$w_{22}=[3,1]^T$

c. The mapping would be: (2,0), (0,2), (0,0)

QUESTION 9 (Marks: 4)

Suppose we have 5 points in R^3 given by (1,0,1),(1,2,0),(1,1,1),(1,1,3),(2,0,3):

Suppose the data is projected onto $z = \mathbf{w}^T\mathbf{x}$ with $\mathbf{w} =$ 1/sqrt(6)*[1 1 2]^T . Find the data points after projection and what is the variance of z?

z = [1/sqrt(6) * 3 1/sqrt(6) * 3 1/sqrt(6) * 4 1/sqrt(6) * 8 1/sqrt(6) * 8]^T
var(z) = 1/6*134/25