NATIONAL INSTITUTE OF TECHNOLOGY WARANGAL – 506 004

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

MCA, II Year II-Semester

Mid Examination, February 2019

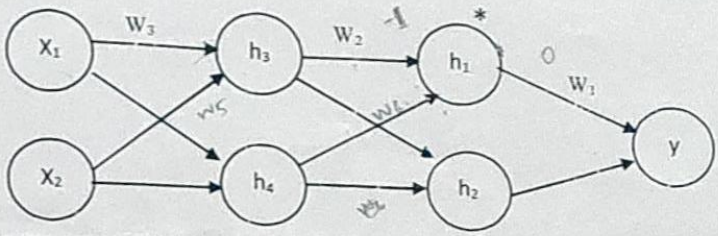
Sub : Machine Learning

Time : 2 Hours

Date: 01-03-2019

Max. Marks: 30

1. Consider the network shown in the figure. All of the hidden units use the linear rectification non linearity $h_i = max(z_i, 0)$. We are trying to minimize a cost function C which depends only on the activation of the output unit y. The unit h_1 (marked with a*) receives an input of -1 on a particular training case, so its output is 0. Based only on this information, which of the following weight derivatives are guaranteed to be 0 for this training case? Write YES or NO for each. Justify your answers informally. Hint: don't work through the backprop computations. [4M]



2. Answer the following questions

a) Consider the factorized joint probability

$$P(A,B,C,D,E,F,G) = P(G|E)P(E|B)P(F|C,D)P(C)P(D|A,B)P(B)P(A)$$

Draw the corresponding Bayesian network. [3M]

b) A drug test (random variable T) has 1% false positives (i.e., 1% of those not taking drugs show positive in the test), and 5% false negatives (i.e., 5% of those taking drugs test negative). Suppose that 2% of those tested are taking drugs. Determine the probability that somebody who tests positive is actually taking drugs (random variable D). [4M]

c) In an oral exam you have to solve exactly one problem, which might be one of three types, A, B, or C, which will come up with probabilities 30%, 20%, and 50%, respectively. During your preparation you have solved 9 of 10 problems of type A, 2 of 10 problems of type B, and 6 of 10 problems of type C. [5M]

(i) What is the probability that you will solve the problem of the exam?

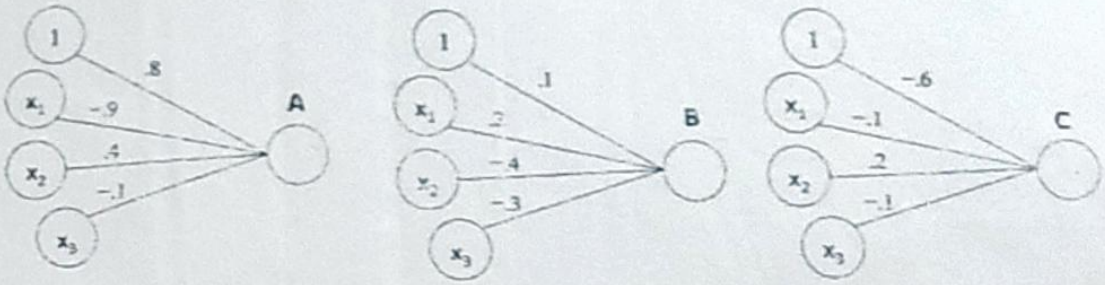
(ii) Given you have solved the problem, what is the probability that it was of type A?

3. Answer the following questions

a) A binary linear classifier to compute the NAND (not-AND) function. This function receives two binary-valued inputs x1 and x2, and returns 0 if both inputs are 1, and returns

1 otherwise. Give four constraints on the weights w_1 and w_2 and the bias b, i.e. one constraint for each of the 4 possible input configurations. [3M]

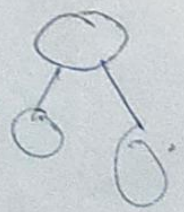
b) Consider the three perceptron's below, which respectively correspond to classes A, B, and C. For a given input x, the perceptron with the highest value of $\sum_i w_i x_i$ is the prediction of the group. [3M]



If $\mathbf{x} = (1, 1, 0)$ is the input to this group, which class is the prediction of the group?

4. Build a decision tree from the given tennis dataset. You should build a tree to predict Play Tennis, based on the other attributes (but, do not use the Day attribute in your tree). Show all of your work, calculations, and decisions as you build the tree. What is the classification accuracy? [8M]

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No



NATIONAL INSTITUTE OF TECHNOLOGY, WARANGAL
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
CS6375 II MCA 2nd SEMESTER END EXAMINATION
MACHINE LEARNING

Date: 3-5-2018 Time: 3.00 Hours Max.Marks: 50

1. Explain working of following with relevant formulae and suitable examples: 5
i) Bayes optimal classifier
ii) Naïve Bayes Classifier
iii) Gibbs algorithm

2. Apply Naive bayes classifier to the following ***PlayTennis*** training data from the following table and classify the following novel instance: 6
(Outlook = sunny, Temperature = cool, Humidity = high, Wind = strong). Show the Relevant formulae used with calculations.

Day	***Outlook***	***Temperature***	***Humidity***	***Wind***	***PlayTennis***
Dl	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
Dl0	Rain	Mild	Normal	Weak	Yes
Dl1	Sunny	Mild	Normal	Strong	Yes
Dl2	Overcast	Mild	High	Strong	Yes
Dl3	Overcast	Hot	Normal	Weak	Yes
Dl4	Rain	Mild	High	Strong	No

3. Full joint distribution for three Boolean variables *toothache, cavity, catch* is shown in the following table. 4

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
Cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Obtain values of following:
i) P(*Cavity*) ii) P(*Toothache*|*cavity*)
iii) P(*Cavity*|*toothache* ∨ *catch*) iv) P(*toothache*)

4a) Draw a Bayesian network showing both the topology and conditional probability tables. 4

DAG
parent

4b)	Explain the four properties which specify the Bayesian networks.	3
4c)	Show that Every entry in the full joint probability distribution can be calculated from the information in the Bayesian network with a suitable illustrative example.	4
5)	Describe the i) Locally weighted linear Regression	4
	ii) Radial basis functions with relevant formulae and explain their working. Draw also the Architecture diagram of Radial Basis Function network.	3.5
6)	a. Write the K-nearest neighbor Algorithm for i) classification and ii) prediction.	3
	b. Explain case based reasoning.	2.5
7)	Describe the Koza's method using genetic programming to solve the Block-stacking problem for creating a block of stacks on the table spelled 'UNIVERSAL' from a random initial configuration of blocks. Explain the terminal arguments and Primitive functions used.	6
8.	Describe application for face recognition of a Feed Forward Neural Network.	5

NATIONAL INSTITUTE OF TECHNOLOGY WARANGAL
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
MCA, II Year II-Semester

Sem End Examination, May 2019

Sub: Machine Learning
Time: 3 hours

Date: 10-5-2019
Max. 50

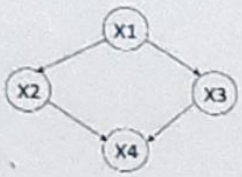
NOTE: Answer all the questions

1. Assume the probability of a certain disease is 0.01. The probability of testing positive given that a person is infected with the disease is 0.95 and the probability of testing positive given the person is not infected with the disease is 0.05. [4M]

(a) Calculate the probability of testing positive.

(b) Use Bayes' Rule to calculate the probability of being infected with the disease given that the test is positive.

2. Consider the Bayes network below, defined over four Boolean variables. [12M]



a) How many parameters are needed to define $P(X_1,X_2,X_3,X_4)$ for this Bayes Net?
b) Give the formula that calculates $P(X_1 = 1,X_2 = 0,X_3 = 1,X_4 = 0)$ using only the Bayes net parameters. Use notation like $P(X_1 = 0|X_2 = 1,X_4 = 0)$ to refer to each Bayes net parameter you use in your formula.
c) Give the formula that calculates $P(X_1 = 1,X_4 = 0)$ using only the Bayes net parameters.
d) Give the formula that calculates $P(X_2 = 1|X_3 = 0)$ using only the Bayes net parameters.

3. Describe in brief [9M]

1. Lazy and eager learning
2. Collaborative filtering
3. Bayesian Belief networks.

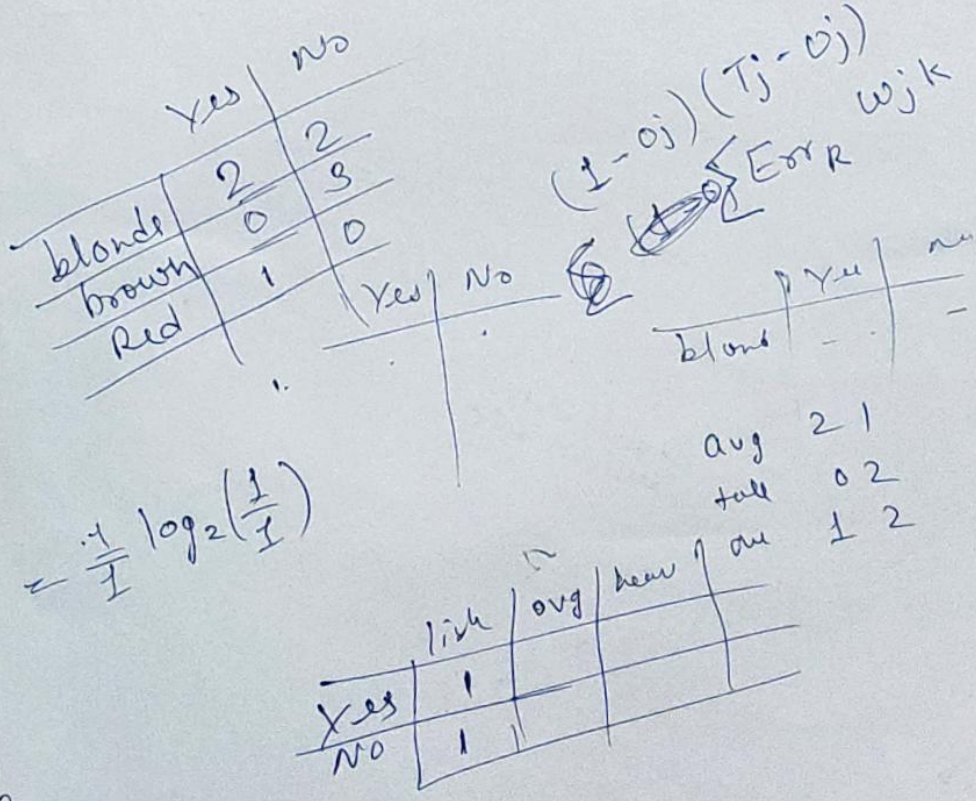
4. What is kernel? How kernel can be used with SVM to classify non-linearly separable data? Also, list standard kernel functions. [6M]

5. Determine the Principal Components for the given 2-Dimensional dataset, (1, 2), (2, 4), (3, 6). [5M]

6. For a SunBurn dataset given below, construct a decision tree [6M]

Name	Hair	Height	Weight	Location	Class
Sunita	blonde	average	light	no	yes
anit	blonde	tall	average	yes	no
kavita	brown	short	average	yes	no
sushma	blonde	short	average	no	yes
xavier	red	average	heavy	no	yes
balaji	brown	tall	heavy	no	no
ramesh	brown	average	heavy	no	no
swetha	blonde	short	light	yes	no

7. Consider a multilayer feed forward neural network. Enumerate and explain steps in back propagation algorithm use to train network. [8M]



Date: 17-4-2018 MCA ML Time: 50 mins 2nd Minor Max.Marks: 10

1.	Explain the working of Bayes optimal classifier with necessary formulae for the following example. Assume hypothesis space contains three hypotheses h1, h2 and h3 with posterior probabilities of hypotheses being .4, .3 and .3 respectivly. For a new instance x, h1, h2 and h3 output the decision as positive, negative and negative respectively. Obtain the classification obtained by using Bayes optimal classifier showing the necessary calculations.	2.5
2.	Write the algorithm for GA	4
3.	Trace the Gabil's Genetic Algorithm for 4 generations to obtain the classifier to implement the 2 input EXOR gate.	3.5

ROLL. NO → 177936

SUDHANSHU AGRAWAL

NATIONAL INSTITUTE OF TECHNOLOGY WARANGAL – 506 004

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

MCA, II Year II-Semester

Minor-2 Examination, April 2019

Sub : Machine Learning — Date: 26-04-2019

Time : 30 Minutes — Max. Marks: 10

---

1. Consider a binary classification problem. Suppose I have trained a model on a linearly separable training set, and now I get a new labeled data point which is correctly classified by the model, and far away from the decision boundary. If I now add this new point to my earlier training set and re-train, in which cases is the learnt decision boundary likely to change? [1M]

(a) When my model is a perceptron.
(b) When my model is logistic regression.
(c) When my model is an SVM.
(d) When my model is decision tree

2. The k-means algorithm for clustering is guaranteed to converge to a local optimum. [1M]

A) True B) False

3. The singular value decomposition of a real matrix is unique. [1M]

A) True B) False

4. Assume we use radial basis kernel function. Thus there is some implicit unknown mapping function $\Phi(x)$. Prove that for any two input instances x_i and x_j, the squared Euclidean distance of their corresponding points in the feature space Q is less than 2. [2M]

5. Explain about expectation and maximization algorithm with equations [2M]

6. You are given the following five training instances [2M]

1. x1 = 2; x2 = 1; y = 4
2. x1 = 6; x2 = 3; y = 2
3. x1 = 2; x2 = 5; y = 2
4. x1 = 6; x2 = 7; y = 3
5. x1 = 10; x2 = 7; y = 3

We want to model this function using the K-nearest neighbor model. When we want to predict the value of y corresponding to (x1; x2) = (3; 6)

(a) For K = 2; y = 3
(b) For K = 2; y = 2:5
(c) For K = 3; y = 2:33
(d) For K = 3; y = 2:666

7. Explain about Soft margin in SVM [1M]

NATIONAL INSTITUTE OF TECHNOLOGY WARANGAL – 506 004

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

MCA, II Year II-Semester

Minor-2 Examination, April 2019

Sub : Machine Learning — Date: 26-04-2019

Time : 30 Minutes — Max. Marks: 10

---

1. Consider a binary classification problem. Suppose I have trained a model on a linearly separable training set, and now I get a new labeled data point which is correctly classified by the model, and far away from the decision boundary. If I now add this new point to my earlier training set and re-train, in which cases is the learnt decision boundary likely to change? [1M]

(a) When my model is a perceptron.
(b) When my model is logistic regression.
(c) When my model is an SVM.
(d) When my model is decision tree

2. The k-means algorithm for clustering is guaranteed to converge to a local optimum. [1M]

A) True B) False

3. The singular value decomposition of a real matrix is unique. [1M]

A) True B) False

4. Assume we use radial basis kernel function. Thus there is some implicit unknown mapping function $\Phi(x)$. Prove that for any two input instances x_i and x_j, the squared Euclidean distance of their corresponding points in the feature space Q is less than 2. [2M]

5. Explain about expectation and maximization algorithm with equations [2M]

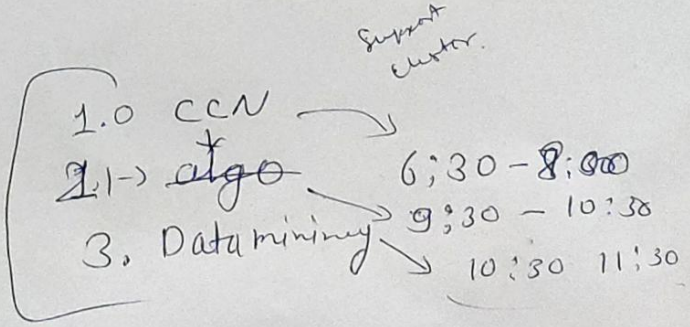
6. You are given the following five training instances [2M]

1. x1 = 2; x2 = 1; y = 4
2. x1 = 6; x2 = 3; y = 2
3. x1 = 2; x2 = 5; y = 2
4. x1 = 6; x2 = 7; y = 3
5. x1 = 10; x2 = 7; y = 3

We want to model this function using the K-nearest neighbor model. When we want to predict the value of y corresponding to (x1; x2) = (3; 6)

(a) For K = 2; y = 3
(b) For K = 2; y = 2:5
(c) For K = 3; y = 2:33
(d) For K = 3; y = 2:666

7. Explain about Soft margin in SVM [1M]



NATIONAL INSTITUTE OF TECHNOLOGY

WARANGAL – 506 004

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

MCA, IV Semester

Model Minor Examination, Feb 2019

Sub : Machine Learning

Date: 21-02-2019

Time : 50 Minutes

Max. Marks: 10

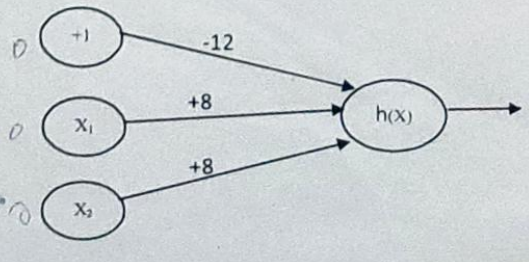
1. Which of the following is true?
(i) On average, neural networks have higher computational rates than conventional computers.
(ii) Neural networks learn by example.
(iii) Neural networks mimic the way the human brain works.
(a) all of them are true
(b) (ii) and (iii) are true
(c) (i), (ii) and (iii) are not true
(d) (i) and (iii)

2. The back-propagation learning algorithm applied to a two-layer neural network
A) always finds the globally optimal solution.
B) finds a locally optimal solution which may be globally optimal.
C) never finds the globally optimal solution.
D) finds a locally optimal solution which is never globally optimal

3. In a neural network, which one of the following techniques is NOT useful to reduce overfitting?
A) Dropout
B) Regularization
C) Batch normalization
D) Adding more layers

4. For an image recognition problem (such as recognizing a cat in a photo), which architecture of neural network has been found to be better suited for the tasks?
A) Multi-layer perceptron
B) Recurrent neural network
C) Convolutional neural network
D) Perceptron

5. The neural network given below takes two binary valued inputs x., x2 ∈{0, 1} and the activation function is the binary threshold function *{h(x)=1 if x>0; 0 otherwise}*. Which of the following logical functions does it compute?



A) OR
B) AND
C) NAND
D) None of these

6. Consider the following Bayesian network, where F stands for Flu and C stands for Coughing. Find P(C).

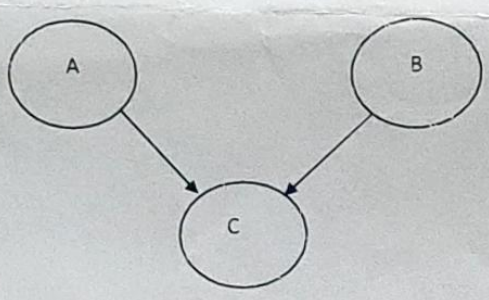
P(F)= 0.1

P(C/F)= 0.8

P(C/$\overline{F}$)=0.3

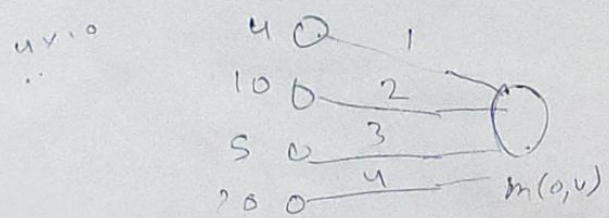
7. Diabetic Retinopathy is a disease that affects 80% people who have diabetes for more than 10 years. 5%-of the Indian population has been suffering from diabetes for more than 10 years. Answer the following questions. What is the joint probability of finding an Indian suffering from Diabetes for more than 10 years and also has Diabetic Retinopathy? [2M]

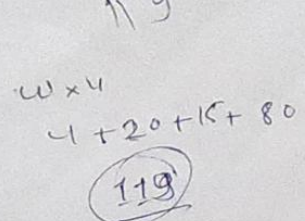
8. In the following Bayesian network A, Band C are Boolean random variables taking values in {True, False}. Which of the following statements is true?



A) The value of C is not given. If the value of B changes from True to False, the conditional probability of A, P(AIB) changes.
B) The value of C is given to be True. If the value of B changes from True to False the conditional probability of A, P (A| B) changes.
C) Neither A nor B
D) Both A and B

9. A 4-input neuron has bias of 0 and weighs 1, 2, 3 and 4. The transfer function is given by f(v)= max(0,v). The inputs are 4, 10, 5 and 20 respectively. The output will be?





NATIONAL INSTITUTE OF TECHNOLOGY WARANGAL – 506 004

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

MCA, II Year II-Semester

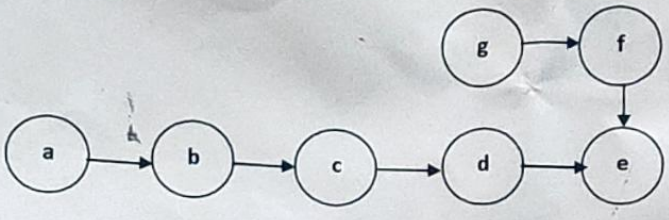
Mid Examination, February 2020

Sub : Machine Learning (CS6375) Date: 17-02-2020

Time : 2 Hours Max. Marks: 30

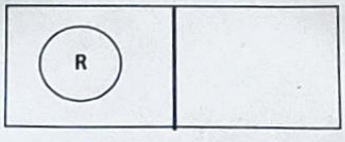
1. Given 6 data points in 5-d space, (1, 1, 1, 0, 0), (−3, −3, −3, 0, 0), (2, 2, 2, 0, 0), (0, 0, 0, −1, −1), (0, 0, 0, 2, 2), (0, 0, 0, −1, −1). We can represent these data points by a 6×5 matrix X, where each row corresponds to a data point. Determine the Principal Components for the given dataset. [5M]

2. Consider the Bayesian network shown in below figure. All the variables are boolean. [3M]



Write the expression for the joint likelihood of the network in its factored form.

3. Consider a robot operating in the two-cell grid world shown below. Suppose the robot is initially in the cell C_1. At any point of time the robot can execute any of the two actions: A_1 and A_2. A_1 is "to move to a neighboring cell". If the robot is in C_1 the action A_1 succeeds (moves the robot into C_2) with the probability 0.9 and fails (leaves the robot in C_1) with the probability 0.1. If the robot is in C_2 the action A1 succeeds (moves the robot into C_1) with the probability 0.8 and fails (leaves the robot in C_2) with the probability 0.2. The action A_2 is "to stay in the same cell", and when executed it keeps the robot in the same cell with probability 1. The first action the robot executes is chosen at random (with an equal probability between A_1 and A_2). Afterwards, the robot alternates the actions it executes. (for example, if the robot executed action A_1 first, then the sequence of actions is $A_1, A_2, A_1, A_2, \ldots$). Answer the following questions. [2X6=12M]

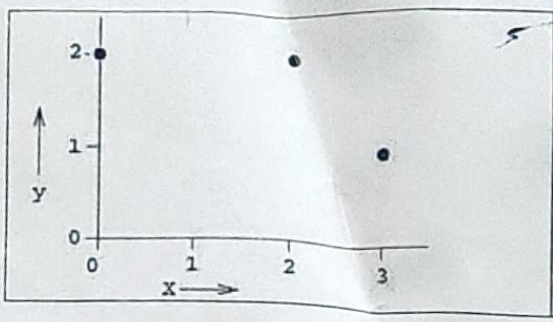


C_1 C_2

A. Draw the Bayes net that represents the cell the robot is in during the first two actions the robot executes (e.g, initial cell, the cell after the first action and the cell after the second action) and fill in the probability tables. (Hint: The Bayes net should have five variables: q_1 -the initial cell, q_2, q_3 - the cell after the first and the second action, respectively, a_1, a_2 - the first and the second action, respectively).

B. Suppose you were told that the first action the robot executes is A_1. What is the probability that the robot will appear in cell C_1 after it executes close to infinitely many actions?

4. Suppose you have this data set with one real-valued input and one real-valued output. [3M]



What is the mean squared leave one out cross validation error of using linear regression?
(i.e. the mode is $y = \beta_0 + \beta_1 x + \text{noise}$)

$\beta_0 = \bar{y} - \beta_1 \bar{x}$

$\beta_1 = \frac{\sum (x - x_1)(y - y_1)}{(x - x_1)^2}$

5. The following dataset will be used to learn a decision tree for predicting whether a person is happy or sad based on the color of their shoes, whether they wear a wig and the number of ears they have. [4M]

Color	Wig	Num. Ears	Output
G	Y	2	S
G	N	2	S
G	N	2	S
B	N	2	S
B	N	2	H
R	N	2	H
R	N	2	H
R	N	2	H
R	Y	3	H

a. What is H(Emotion|Wig=Y)

b. Draw the full decision tree (with proper steps) that would be learned for this data. (Note: Direct answer not acceptable)

6. A RGB Color 512 x 512 image is input into an algorithm which outputs a color 64 x 64 image representing some important portions of the original image. For example, the input could be a suspicious tumor like portion, etc. If we write a model of the relation between the input vector x and output vector y as $y = Ax + b$. then the total number of elements in the matrix A is? [3M]

NATIONAL INSTITUTE OF TECHNOLOGY WARANGAL – 506 004

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

MCA, II Year II-Semester

Minor-2 Examination, April 2019

Sub : Machine Learning — Date: 26-04-2019

Time : 30 Minutes — Max. Marks: 10

---

1. Consider a binary classification problem. Suppose I have trained a model on a linearly separable training set, and now I get a new labeled data point which is correctly classified by the model, and far away from the decision boundary. If I now add this new point to my earlier training set and re-train, in which cases is the learnt decision boundary likely to change? [1M]

(a) When my model is a perceptron.
(b) When my model is logistic regression.
(c) When my model is an SVM.
(d) When my model is decision tree

2. The k-means algorithm for clustering is guaranteed to converge to a local optimum. [1M]

A) True B) False

3. The singular value decomposition of a real matrix is unique. [1M]

A) True B) False

4. Assume we use radial basis kernel function. Thus there is some implicit unknown mapping function $\Phi(x)$. Prove that for any two input instances x_i and x_j, the squared Euclidean distance of their corresponding points in the feature space Q is less than 2. [2M]

5. Explain about expectation and maximization algorithm with equations [2M]

6. You are given the following five training instances [2M]

1. x1 = 2; x2 = 1; y = 4
2. x1 = 6; x2 = 3; y = 2
3. x1 = 2; x2 = 5; y = 2
4. x1 = 6; x2 = 7; y = 3
5. x1 = 10; x2 = 7; y = 3

We want to model this function using the K-nearest neighbor model. When we want to predict the value of y corresponding to (x1; x2) = (3; 6)

(a) For K = 2; y = 3
(b) For K = 2; y = 2:5
(c) For K = 3; y = 2:33
(d) For K = 3; y = 2:666

7. Explain about Soft margin in SVM [1M]