K- Means Clustering Algorithm

1. Apply K- means algorithm on given data for k=3. Use $C_1(2)$, $C_2(16)$ and $C_3(38)$ as initial cluster centres.

   Data: 2,4,6,3,31,12,15,16,38,35,14,21,23,25,30

2. Divide the given sample data in two clusters using K- means algorithm. (Use Euclidean Distance)

HEIGHT	WEIGHT
185	72
170	56
168	60
179	68
182	72
188	77
180	71
180	70
183	84
180	88
180	67
177	76

3. Apply K-means algorithm on given data for Center points C1=16 and c2=22

   Data: 15,15,16,19,19,20,20,21,22,28,35,40,41,42,43,44,60,61,65

4. Cluster the following eight points (with (x, y) representing locations) into three clusters

   A1(2, 10)
   A2(2, 5)
   A3(8, 4)
   A4(5, 8)
   A5(7, 5)
   A6(6, 4)
   A7(1, 2)
   A8(4, 9)

   Initial cluster centers are: A1(2, 10), A4(5, 8), A7(1, 2)

5. Apply K-means algorithm on given data
   (2, 3), (5, 6), (8, 7), (1, 4), (2, 2), (6, 7), (3, 4), (3, 6)
   Where K= 2 and C1= (2, 3), C2= (5, 6)

6. What is clustering? Explain k-means clustering algorithm. Suppose the data for clustering – {2, 4, 10, 12, 3, 20, 11, 25}. Consider k=2, cluster the given data using above algorithm.

7. Calculate 2 clusters using k-means cluster algorithm. For finding the distance use Euclidian distance.

Subject	A	B
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

8. What is learning by observations? Explain k-means clustering in detail.
9. Explain k-Means clustering with Example.
10. Explain the parameter k in k-Means Clustering Algorithm. Explain k-Means Clustering Algorithm along with an example.
11. If cluster centers are P3, P4 and P9 then What will be the clusters after two iterations of k-means clustering algorithm? Use Manhattan Distance?

	X
P1	12
P2	15
P3	23
P4	4
P5	7
P6	28
P7	30
P8	35
P9	45
P10	51

12. Use k – means clustering algorithm to divide the following data into two cluster and also compute the representative data point for the cluster.

X1	1	2	2	3	4	5
X2	1	1	3	2	3	5

13. How does the K-means algorithm work?
14. What's the difference between centroid and medoids? Are centroids always data instances in the dataset? Are medoids always data instances in the dataset? Explain.
15. Discuss effective ways to determine an appropriate value of k to use
16. Discuss effective ways to choose appropriate initial centroids. Illustrate situations in which one way would be more appropriate than the others.
17. Explain why the time complexity of the k-means algorithm is O(I*K*m*n).

K- Medoids Clustering Algorithm

1. Apply K-medoids clustering algorithm on given data

	X	Y
D1	2	6
D2	3	4
D3	3	8
D4	4	7
D5	6	2
D6	6	4
D7	7	4
D8	7	3
D9	8	5
D10	7	6

2. How K-Mean clustering method differs from K-Medoid clustering method?

Hierarchical clustering
(Dendrogram, Agglomerative)

1. Apply Agglomerative Hierarchical Clustering and draw single link and average link. dendrogram for the following distance matrix.

	A	B	C	D	E
A	0	2	6	10	9
B	2	0	3	9	8
C	6	3	0	7	5
D	10	9	7	0	4
E	9	8	5	4	0

2. Plot a dendogram using complete linkage (agglomerative clustering) for the following dataset.

ITEM	**A**	**B**	**C**	**D**	**E**
A	0	9	3	6	11
B	9	0	7	5	10
C	3	7	0	9	2
D	6	5	9	0	8
E	11	10	2	8	0

3. Perform agglomerative algorithm on the following data and plot a dendogram using single link approach. The given data indicates the distance between elements.

ITEM	**E**	**A**	**C**	**B**	**D**
E	0	1	2	2	3
A	1	0	2	5	3
C	2	2	0	1	6
B	2	5	1	0	3
D	3	3	6	3	0

4. What the difference between agglomerative and divisive hierarchical clustering?
5. What's a dendrogram?
6. Describe the steps of the basic agglomerative hierarchical clustering algorithm.
7. What's a proximity matrix?
8. What's the time complexity of the basic agglomerative hierarchical clustering algorithm.

9. For the given set of points identify clusters using complete link and average link using agglomerative clustering.

	A	**B**
P1	1	1
P2	1.5	1.5
P3	5	5
P4	3	4
P5	4	4
P6	3	3.5

10. Find the clusters using single link technique. Use Euclidean distance, and draw the dendrogram.

	X	Y
P1	0.40	0.53
P2	0.22	0.38
P3	0.35	0.32
P4	0.26	0.19
P5	0.08	0.41
P6	0.45	0.30

11. For the one dimensional data set **{7,10,20,28,35}**, perform hierarchical clustering and plot the dendogram to visualize it

Association Rules

1. Apply apriori algorithm to discover the frequent item sets to the following data set.

Transaction Id	
101	Mulberry, raspberry, cherry
102	Mulberry, Papaya
103	Papaya, Mango
104	Mulberry, raspberry, cherry
105	Passion Fruit, Cherry
106	Passion Fruit
107	Passion Fruit, Papaya
108	Mulberry, raspberry, cherry, Guava
109	Guava, Mango
110	Mulberry, raspberry

2. What are the limitations of apriori algorithm? How to increase the efficiency of algorithm?
3. A database has seven transactions. Each transaction Ti is set of items purchased in a basket in a store by a customer. Let Minimum support =30% and minimum confidence= 80%Find the frequent item set using apriori algorithm.

Transaction Id	
101	Butter, Bread, Milk
102	Butter, Cheese
103	Cheese, Biscuit
104	Butter, Bread, Cheese
105	Butter, Bread, clothes, cheese, Milk
106	Bread, Clothes, Milk
107	Bread , Milk, Clothes

4. Explain What is meant by association rule mining. For the table given below perform apriori algorithm and also:
    a. Determine the frequent item set obtained.
    b. Justify the strong association rule that has been determined i.e specify which is the strongest rule obtained.

Transaction ID	Items
1	1,3,4,6
2	2,3,5,7
3	1,2,3,5,8
4	2,5,9,10
5	1,4

    Assume minimum support 30% and Minimum confidence of 75%
5. Consider the following dataset and find frequent item sets and generate association rules for them using Apriori Algorithm.

TID	items
T1	I1, I2 , I5
T2	I2,I4
T3	I2,I3
T4	I1,I2,I4
T5	I1,I3
T6	I2,I3
T7	I1,I3
T8	I1,I2,I3,I5
T9	I1,I2,I3

minimum support count is 2 minimum confidence is 60% .

6. Write a note on incremental Association Rule Mining.
7. Discuss hash based technique to improve the effeciency of appriori algorithm.
8. Provide formulas that define confidence, support, and lift in terms of probability.
9. How has the association rule mining problem been traditionally formulated?
10. What's the apriori principle? Provide 2 different formulations of this principle.

Decision Trees

1. Enumerate how decision trees can be made scalable?
2. Briefly outline the major steps of decision tree classification. Why tree pruning useful in decision tree induction?
3. What are tree based model? explain decision tree.
4. How Decision Tree learning is differing from Rule learning? Explain First Order Rule learning (FOIL) algorithm with its major steps.
5. Construct decision tree for the following using ID3 algorithm

Day	A1	A2	A3	Classification
1	True	Hot	High	No
2	True	Hot	High	No
3	False	Hot	High	Yes
4	False	Cool	Normal	Yes
5	False	Cool	Normal	Yes
6	True	Cool	High	No
7	True	Hot	High	No
8	True	Hot	Normal	Yes
9	False	Cool	Normal	Yes
10	False	Cool	High	No

6. Explain the concept of decision tree learning. Discuss the necessary measure required to select the attributes for building a decision tree using ID3 algorithm.
7. Discuss the issue of avoiding overfitting of data, handling continuous data and missing values in decision trees.
8. Write ID3 Decision Tree Algorithm and explain with suitable example
9. Give Decision Tree representations for following Boolean Functions (7 marks)
    a. A ᵛ (B ^ C)
    b. (A ^ B) ᵛ (C ^ D)
10. Consider the following dataset for predicting outcome of a tennis match between Fedrer and Nadale:

Time	Match Type	Court surface	Best Effort	Outcome
Morning	Master	Grass	1	F
Afternoon	Grand Slam	Clay	1	F
Night	Friendly	Hard	0	F
Afternoon	Friendly	Mixed	0	N
Afternoon	Master	Clay	1	N
Afternoon	Grand Slam	Grass	1	F
Afternoon	Grand Slam	Hard	1	F
Afternoon	Grand Slam	Hard	1	F
Morning	Master	Grass	1	F
Afternoon	Grand Slam	Clay	1	N
Night	Friendly	Hard	0	F

Night	Master	Mixed	1	N
Afternoon	Master	Clay	1	N
Afternoon	Master	Grass	1	F
Afternoon	Grand Slam	Hard	1	F
Afternoon	Grand Slam	Clay	1	F

i. Write the formula for information gain for an attribute. Compute information gain for all attributes. Which is selected as a root node?
ii. Compute the structure of the decision tree computed by ID3 algorithm.

11. For a sunburn dataset given below, Construct a decision tree:

Name	Hair	Height	Weight	Location	Class
Sunita	Blonde	Average	Light	No	Yes
Anita	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
Sweta	Blonde	Short	Light	Yes	No

12. What are the issues in decision tree induction?
13. Give decision trees to represent the following boolean functions
    a. A ^~B
    b. A V [B ^ C]
    c. A XOR B
    d. [A ^ B] v [C ^ D]
14. Greedily learn a decision tree using the ID3 algorithm and draw the tree.
    a. Write the learned concept for Martian as a set of conjunctive rules (e.g., if (green=Y and legs=2 and height=T and smelly=N), then Martian; else if ... then Martian;...; else Human).
    b. The solution of part b)i) above uses up to 4 attributes in each conjunction. Find a set of conjunctive rules using only 2 attributes per conjunction that still results in zero error in the training set. Can this simpler hypothesis be represented by a decision tree of depth 2? Justify.
15. Describe in detail 3 characteristics of decision tree construction
16. How does pre-pruning of decision trees work?
17. How does post-pruning of decision trees work?
18. What's known as the Occams' razor principle and how does it apply to the construction of a model?
19. You get the following data set
    Your task is to build a decision tree for classifying variable Y . (You can think of the data set as replicated many times, i.e. overfitting is not an issue here).

V	W	X	Y
0	0	0	0
0	1	0	1
1	0	0	1
1	1	0	0
1	1	1	0

a. Compute the information gains IG(Y |V), IG(Y |W) and IG(Y |X). Remember, information gain is defined as

$$IG(A|B) = H(A) - \sum_{b \in B} P(B = b) H(A|B = b)$$

Where

$$H(A) = -\sum_{a \in A} P(A = a) \log_2 P(A = a)$$

Is the Entropy of A and

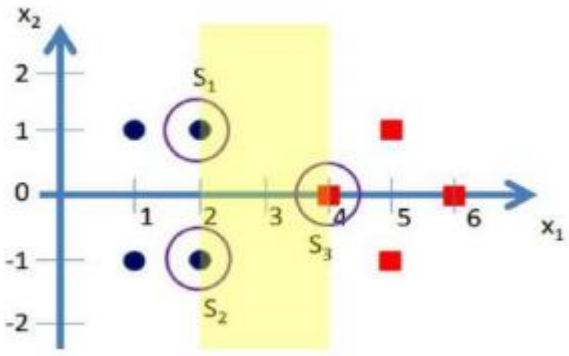
$$H(A|B = b) = -\sum_{a \in A} P(A = a|B = b) \log_2 P(A = a|B = b)$$

is conditional entropy of A given B.
Which attribute would ID3 select first?

b. Write down the entire decision tree constructed by ID3, without pruning
c. One idea for pruning would be to start at the root, and prune splits for which the information gain (or some other criterion) is less than some small ε. This is called top-down pruning. What is the decision tree returned for ε = 0.0001? What is the training set error for this tree?
d. Another option would be to start at the leaves, and prune subtrees for which
e. the information gain (or some other criterion) of a split is less than some small ε. In this method, no ancestors of children with high information gain will get pruned. This is called bottom-up pruning. What is the tree returned for ε = 0.0001? What is the training set error for this tree?
f. Discuss when you would want to choose bottom-up pruning over top-down pruning and vice versa. Compare the classification accuracy and computational complexity of both types of pruning.
g. What is the height of the tree returned by ID3 with bottom-up pruning? Can you find a tree with smaller height which also perfectly classifies Y on the training set? What conclusions does that imply about the performance of the ID3 algorithm?

Support Vector Machine (SVM)

1. What is perceptron? Difference between Perceptron and SVM.
2. What is Perceptron? What is Multi-Layer Feed Forward Networks? What is SVM? How does SVM differ from Artificial Neural Networks (ANN)?
3. What is the goal of support vector machine (SVM)? How to compute the margin?
4. Write the soft margin SVM problem. Write the KKT conditions for this problem. Derive the dual for soft-margin SVM.
5. Apply SVM algorithm for the data-points and find dimension of hyper plane to classify the data-points for the figure. (assume bias =1)



6. State whether the given statement is true or false:
   Training set error will initially decrease and then increase as we increase the parameter C in soft-margin SVM training

K - Nearest Neighbours

1. Describe K-nearest neighbor algorithm. Why is it known as instance-based Learning?
2. Which of the following will be true about k in k-NN in terms of variance? Why?
    a. When you increase the k, the variance will increase
    b. When you decrease the k, the variance will increase
    c. Cannot say
    d. None of the above
3. Explain the k-NN classification along with example. Which are the effects of k on result of k-NN classifier?
4. What is instance based learning? Explain K- nearest Neighbour.
5. Explain K-Nearest Neighbour techniques with an example.
6. In the following questions you will consider a k-nearest neighbor classifier using Euclidean distance metric on a binary classification task. We assign the class of the test point to be the class of the majority of the k nearest neighbors. Note that a point can be its own neighbor

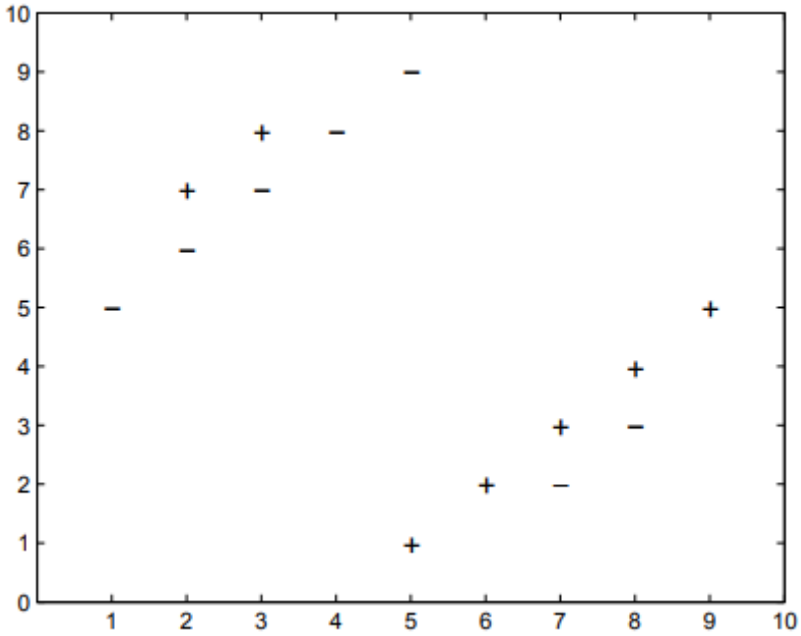


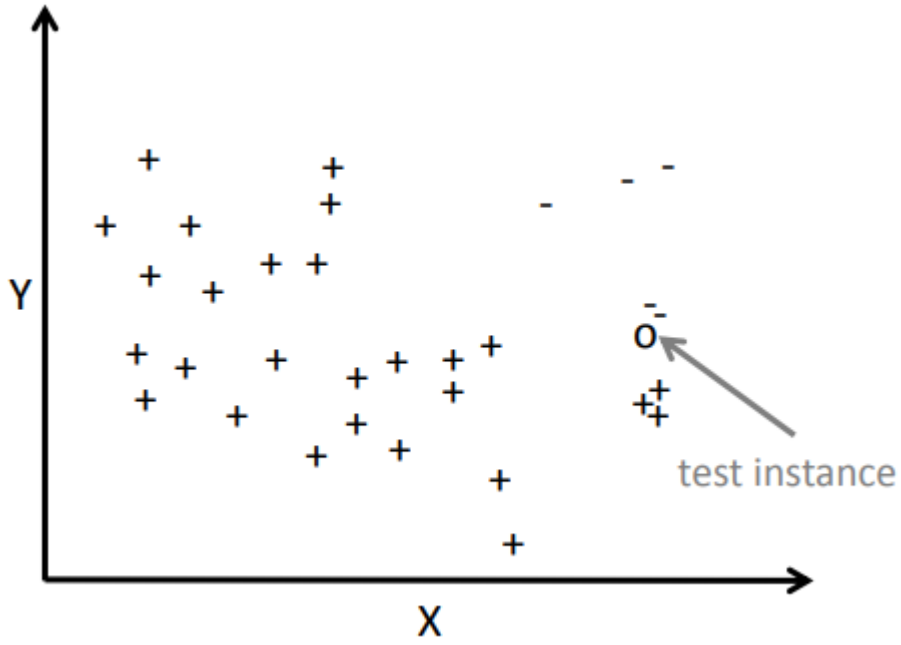
Figure shows the dataset for KNN binary classification task.

    a. What value of k minimizes the training set error for this dataset? What is the resulting training error?
    b. Why might using too large values k be bad in this dataset? Why might too small values of k also be bad?
    c. What value of k minimizes leave-one-out cross-validation error for this dataset? What is the resulting error?
    d. In Figure, sketch the 1-nearest neighbor decision boundary for this dataset.

7. [True or False] k-NN algorithm does more computation on test time rather than train time.
8. A KNN classifier assigns a test instance the majority class associated with its K nearest training instances. Distance between instances is measured using Euclidean distance.

   Suppose we have the following training set of positive (+) and negative (-) instances and a single test instance (o).

   All instances are projected onto a vector space of two real-valued features (X and Y). Answer the following questions. Assume "unweighted" KNN (every nearest neighbor contributes equally to the final vote).



   a) What would be the class assigned to this test instance for K=1
   b) What would be the class assigned to this test instance for K=3
   c) What would be the class assigned to this test instance for K=5
   d) Setting K to a large value seems like a good idea. We get more votes! Given this particular training set, would you recommend setting K = 11? Why or why not?

9. Perform KNN Classification algorithm on following dataset and predict the class for X(P1 = 3 and P2=7). K=3
   (Use Euclidean Distance)

P1	P2	Class
7	7	False
7	4	False
3	4	True
1	4	True

10. Below given is a dataset of students result.

Maths	**CS**	**Result**
4	3	Fail
6	7	Pass
7	8	Pass
5	5	Fail
8	8	Pass

Student X got 6 marks in Maths and 8 marks in CS. State the result of student X. Use K – nearest neighbor classification. (K=3)

11. Check whether Anni likes cricket, football or neither of the sports. Use K – nearest neighbor classification. (K=3).
(Assume: Person cannot like both)

Name	**Age**	**Gender**	**Sport**
Ajay	32	Male	Football
Mark	40	Male	Neither
Sara	16	Female	Cricket
Zaira	34	Female	Cricket
Sachin	55	Male	Neither
Rahul	40	Male	Cricket
Pooja	20	Female	Neither

Smith	15	Male	Cricket
Laxmi	55	Female	Football
Michael	15	Male	Football
Anni	05	Female	???

12. Use K-nearest neighbor classification to solve following problem (Use Euclidean distance measure) (k=3)

Name	**Acid Durability**	**Strength**	**Class**
Type-1	7	7	Bad
Type-2	7	4	Bad
Type-3	3	4	Good
Type-4	1	4	Good
Type - 5	3	7	???

13. Consider a dataset having two variables: Height (cm) & Weight (kg) and each point is classified as Normal or Underweight. Suppose there is a person X having weight 57 kg and Height 170 cm then find the class of person X. Use K- nearest neighbor classification. (K=3)

Weight(X2)	**Height(Y2)**	**Class**
51	167	Underweight
62	182	Normal
69	176	Normal
64	173	Normal
65	172	Normal
56	174	Underweight
58	169	Normal
57	173	Normal
55	170	Normal

Naïve Bayes Classifier

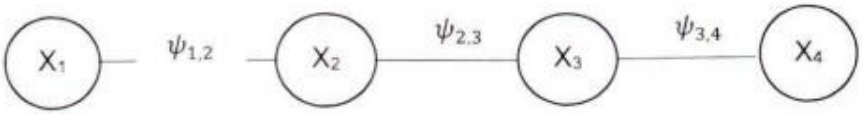
1. Explain role played by the Bayes theorem in Naïve Bayes algorithm What is naïve Bayes' classifier? Explain the prior probability, posterior probability and maximum likelihood probability.
2. Explain the role played by the Bayes' Theorem in Naïve Bayes classifier. What is Naïve Bayesian Classifier? Explain prior Probability, Posterior Probability and Maximum Likelihood with respect to Naïve Bayesian Classifier.
3. In which cases Naive Bayes is useful in Classification? Why?
4. Consider the following balloons dataset:

Colour	Size	Act	Age	Inflated
YELLOW	SMALL	STRETCH	ADULT	T
YELLOW	SMALL	STRETCH	CHILD	T
YELLOW	SMALL	DIP	ADULT	T
YELLOW	SMALL	DIP	CHILD	T
YELLOW	LARGE	STRETCH	ADULT	T
YELLOW	LARGE	STRETCH	CHILD	F
YELLOW	LARGE	DIP	ADULT	F
YELLOW	LARGE	DIP	CHILD	F
PURPLE	SMALL	STRETCH	ADULT	T
PURPLE	SMALL	STRETCH	CHILD	F
PURPLE	SMALL	DIP	ADULT	F
PURPLE	SMALL	DIP	CHILD	F
PURPLE	LARGE	STRETCH	ADULT	T
PURPLE	LARGE	STRETCH	CHILD	F
PURPLE	LARGE	DIP	ADULT	F
PURPLE	LARGE	DIP	CHILD	F

   Compute the parameters of Naïve Bayes classifier for predicting inflated; and the training set error.
5. How to use naive or general Bayesian nets to classify a test instance?
6. How to determine the topology (graph edges) and CPTs of a Bayesian model?

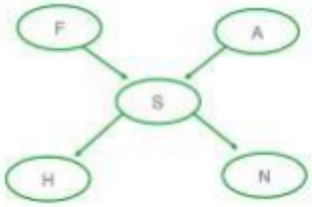
Reinforcement learning

1. Explain Q Learning algorithm
2. Define the term Reinforcement Learning. Also enlist and explain the elements of Reinforcement Learning.
3. Explain Reinforcement learning problem with neat diagram.
4. Write the steps for Reinforcement Learning & also enlist the salient features of it.
5. Explain the Q function and Q Learning Algorithm.
6. With a well labeled diagram, explain various elements of Reinforcement Learning (RL).
7. Explain The Markov Property.
8. Differentiate between Model-Free and Model-based Learning.
9. Give a brief overview of various Problem solving methods for RL.
10. Define the following terms:
    a. Reward Shapping.
    b. Credit Assignment Problem.
    c. Policy Network.
    d. Policy Gradient.
11. Explain the Q function and Q Learning Algorithm assuming deterministic rewards and actions with example.
12. Differentiate between Associative and Non Associative Tasks.
13. Write Reinforcement Learning problem characteristics.
14. Narrate the concept learning in brief. List down the application areas where concept of learning works better than other approaches. Explain Supervised, Unsupervised, Semi-supervised and Reinforcement learning using an example.
15. What do you mean by reinforcement learning? How reinforcement learning problem differ from other function approximation tasks.
16. Explain reinforcement learning in detail along with the various elements involved in forming the concept. Also define what is meant by partially observed state.
17. Explain how reinforcement learning problem differs from other function approximation.
18. Develop a Q learning task for recommendation system of an online shopping website. What will be the environment of the system? Write the cost function and value function for the system.
19. Consider the following Markov Random Field with the corresponding potentials:



$(X_i, X_{i+1}) = 1 - X_i X_{i+1}$ for odd i, and $\psi_{i,i+1}(X_i, X_{i+1}) = 1 + X_i X_{i+1}$ for even i; where $X_i \in \{+1, -1\}$. Find $P(X_2)$ and $x_3^* = \max_{x_3} P(X_3|X_2 = 1)$ using an appropriate message-passing algorithm.

20. Consider the following graphical model, which defines a joint probability distribution over five Boolean variables. Apply Expectation Maximization to train this Bayesian network, given training data in which the variables F, S, H and N are fully observed, and where the variable A is sometimes unobserved.

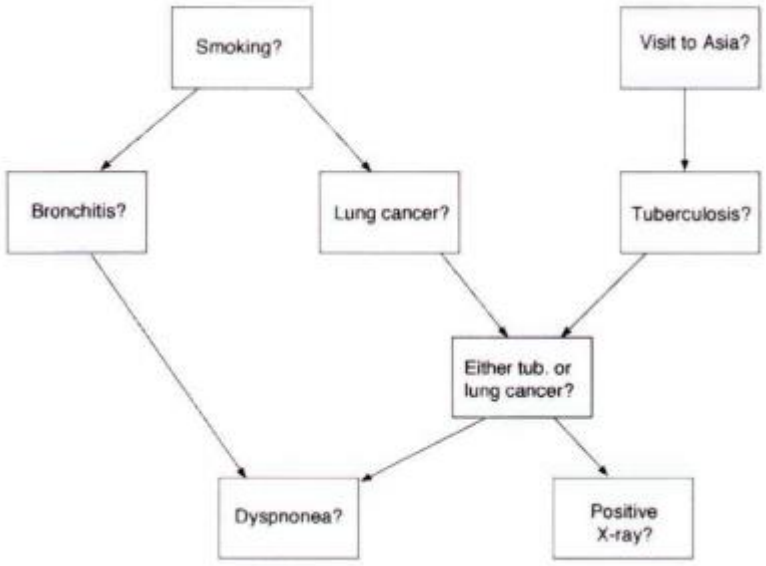


a. What are the conditional probability distributions associated with each of the five random variables in this network?
b. What variables are in the Markov blanket for variable A?
c. Given that all variables in the Markov blanket of A are observed, it should be possible to compute the distribution over A based on only these variables. Simplify your expression from part (2) so that it uses only the variables in the Markov blanket of A.

Bayesian Learning

(Bayes Theorem, Maximum Likelihood and Least Squared Error Hypothesis, Maximum likelihood hypothesis for Predicting probabilities, Bayesian Belief Network.)

1. Which are the difficulties faced in *Hypothesis Estimation*, when the dataset is too small? Explain each of them in brief.
2. Explain maximum a posteriori (MAP) hypothesis using Bayes theorem.
3. Explain Naïve Bayes Classifier and Bayscian Belief Network
4. Prove that how maximum likelihood (Bayesian learning) can be used in any learning algorithms that are used to minimize the squared error between actual output hypothesis and predicted output hypothesis.
5. Explain following (2 marks each)
   i. Bayesian Belief Networks
   ii. MAP Hypothesis
6. Consider the following Bayesian network connection various factors to chest disease:



   iii. Are Bronchitis and Tuberculosis independent when nothing is observed?
   iv. Are they independent after observing x-ray?
   v. Are they independent after observing both x-ray and smoking?
7. What do you mean by hypothesis space, instance space and version space?
8. Find the maximum general hypothesis and maximally specific hypothesis for the training examples given in the table using candidate elimination algorithm.
9. Explain MAP Hypothesis to predict probability.
10. Why it is necessary to estimate the accuracy of hypothesis. Explain procedure to estimate difference in error between two learning methods.

Deep Neural Network

(Introduction to Deep Learning, Deep Neural Network, Restricted Boltzmann machine, Convolution Neural Network, AutoEncoders, Deep Belief Network, Recurrent Neural Network, Transfer learning.)

1. What is the difference between a Feedforward Neural Network and Recurrent Neural Network?
2. What is Boltzmann Machine?
3. What is deep learning?
4. What is perceptron? Write the differences between Single Layer Perceptron(SLP) and Multilayer Perceptron(MLP).
5. What Are the Applications of a Recurrent Neural Network (RNN)?
6. What Are the Softmax and ReLU Functions?
7. What Are Hyper parameters?
8. What Will Happen If the Learning Rate Is Set Too Low or Too High?
9. What Is Dropout and Batch Normalization?
10. What is Pooling on CNN, and How Does It Work?
11. What Are the Different Layers on CNN?
12. How Does an LSTM Network Work?
13. What Is the Difference Between Epoch, Batch, and Iteration in Deep Learning?
14. Why Is Tensorflow the Most Preferred Library in Deep Learning?
15. What Do You Mean by Tensor in Tensorflow?
16. Explain Generative Adversarial Network in brief.
17. What Is an Auto-encoder?
18. What Is Bagging and Boosting?
19. How Are Weights Initialized in a Network?
20. Differentiate between AI, Machine Learning and Deep Learning.
21. Do you think Deep Learning is Better than Machine Learning? If so, why?
22. What are the Hyper parameters? Name a few used in any Neural Network.
23. Which is Better Deep Networks or Shallow ones? and Why?
24. Name a few deep learning frameworks
25. Explain the different Layers of CNN.

Evaluation

(Cross-Validation, Measures of Performance for Classification (Accuracy, Confusion Matrix, Precision, Recall, F1-Score), Measures of Performance for Clustering (Homogeneity, Completeness, V-Measure))

1. Which are the two measures of rule interestingness? Explain with example.
2. List and explain distance measures to compute the distance between a pair of points and find 7out the distance two objects represented by attribute values {1,6,2,5,3} and {3,5,2,6,6} by using any two of the distance measures.
3. What are the methods of evaluating classifier accuracy?
4. How to measure the performance of classifier? What is a confusion matrix? Which are the other similar terms associated with it?
5. How to measure the performance of a Classifier? How to compare the performance of one Classifier with the other? What is a Confusion Matrix?
6. Which of the following is true or false? Explain with reasons (max 2-3 lines)
    a. Root means square error is a good performance measure for multi class classification problem.
    b. Cross- validation is expected to reduce the variance in the estimate of error rate of classifier.
7.
    a. Draw a contingency table for each of the rules using the transactions shown in Table

Rules: [A] {b} → {c}; [B] {a} → {d}; [C] {b} → {d}; [D] {e} → {c}; [E] {c} → {a};

Transaction ID	Item Brought
0001	{a, b, d, e}
0002	{b, c, d}
0003	{a, b, d, e}
0004	{a, c, d, e}
0005	{b, c, d, e}
0006	{b, d, e}
0007	{c, d}
0008	{a, b, c}
0009	{b, d, e}
0010	{b, d}

    b. Use the contingency tables obtained in part (i) to compute and rank the rules in decreasing order according to the following measures: [A] Support; [B] Confidence; [C] Interest (X → Y) = P(X, Y) / (P(X)*P(Y))
8. Explain why k-fold cross validation does not work well with time series model. What can you do about it?
9. Explain Confusion Matrix with respect to detection of “Spam e-mails”.
10. Explain following (2 marks each)
    a. Confusion Matrix
    b. Root Mean Square Error
11. Define training set, testing set, and validation set
12. Define accuracy, error rate, and confusion matrix.
13. Discuss disadvantages of testing a model over its training set.
14. Define the Percentage Split approach for training and testing a model. Discuss its

advantages and disadvantages.

15. Explain how n-fold cross-validation works. Illustrate with an example.
16. What's the leave-one-out evaluation method?
17. What's bootstrap? What's the difference between bootstrap and random subsampling?
18. How is sum of squared error (SSE) defined?
19. How can SSE (see above) be used to evaluate clusters? Discuss whether or not SSE is useful in evaluating clusters results of each of the following methods: k-means, basic agglomerative, DBSCAN.
20. What are cluster cohesion and separation? Provide formulas to calculate them. How can these measures be used to evaluate a clustering?
21. What is the silhouette coefficient? Provide a formula to calculate it. How can this measure be used to evaluate a clustering?
22. How can correlation between the similary (proximity) matrix and the cluster-based incidence matrix be used to evaluate a clustering?

Fundamental Concepts

(Introduction to Data science, Theory and practices in machine learning, Designing a Learning System, Issues in Machine Learning, Applications of ML, Global Developments of ML, Key challenges to adoption of ML in India.)

1. What are the key challenges to adaption of Machine Learning in India?
2. Explain Global Development of Machine Learning?
3. Explain the concept learning in brief. List down the application areas where concept learning works better than other approaches.
4. Give three computer applications for which Machine Learning approaches seem appropriate and three applications for which they seem inappropriate.
5. Explain in brief Descriptive and Predictive Tasks of Machine Learning.
6. What is a Recommender System? How Machine Learning is useful in Recommender Systems?
7. Explain the issues in machine learning.
8. List Applications of machine learning.

Statistical Learning Techniques

(Descriptive statistics, Simple Linear Regression, ANOVA, Logistic Regression, Multi Linear regression, Correlation, Moving Average, Random Number Generation, Histogram Smoothing, Sampling, Rank Percentile.)

1. Briefly explain Linear and Non-linear regression.
2. Discuss Correlation analysis.
3. Eleven students were asked to measure their pulses for 30 seconds and multiply by two to
4. get their 1-minute pulse rate. The results were 62,30,60,66,70,72,74,74,78,84,80. Create five number summary for the pulse rates and draw histogram.
5. Compare and contrast quantile plot and quantile-quantile plot taking below example:

Level	Remember	Understand	Apply	Evaluate	Create
Marks(50%)	4	24	14	-	-
Marks(%)	8	48	28	-	-

6. What is regression? How to measure performance of regression? Explain the regression along with equation.
7. Explain the Regression along with an equation. How to measure performance of Regression? Describe error calculation in Regression.
8. Write down the basics of *sampling theory*. Explain *k-fold cross-validation* and *stratified sampling* in brief.
9. Explain the Regression along with an equation. How to measure performance of Regression? Describe error calculation in Regression.
10. Explain locally weighted linear regression
11. Calculate the cost value of linear regression for the following dataset.
    X= [1,2,3,4,5]
    Y= [3,6,7,11,15]
    Slope of hypothesis (Θ1) = 7
    Constant/ intercept (Θ0) = 5
    Epoch = 3
    Learning rate (α) = 0.5
12. The values of independent variable X and dependent value Y are given below:

X	Y
0	2
1	3
2	5
3	4
4	6

Find the least square regression line y= a x +b. Estimate the value of Y when X is 10.

13. The values of independent variable x and dependent value y are given below:

X	Y
0	2
1	3
2	5
3	4
4	6

Find the least square regression line y=ax+b. Estimate the value of y when x is 10.

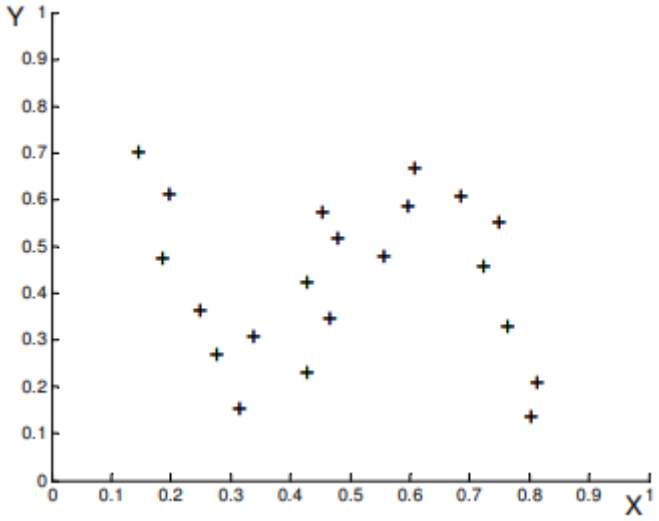
14. Explain sample error, true error, confidence intervals
15. What is histogram? Explain with example.
16. Explain importance of histogram.
17. Draw a histogram for following data.

    Data: 7, 7, 7, 8, 8, 8, 8, 8, 9, 10, 10, 10, 11, 11, 12, 12, 12, 13

18. Here are the grades of 15 students.
    88, 48, 60, 51, 57, 85, 69, 75, 97, 72, 71, 79, 65, 63, 73
    Build a histogram from the data
19. Below given data is ages of people present at a restaurant
    Ages: 1, 3, 27, 32, 5, 63, 26, 25, 18, 16, 4, 45, 29, 19, 22, 51, 58, 9, 42, 6
    Show distribution of ages and draw a histogram.
20. Draw histogram of the following data:

Class Interval (Price range of pen)	10-20	20-30	30-40	40-50	50-60
Frequency (No of pens sold)	15	20	30	25	5

21. For each of the following questions, you are given the same data set. Your task is to fit a smooth function to this data set using several regression techniques. Please answer all questions qualitatively, drawing the functions in the respective figures.
    a. Show the least squares fit of a linear regression model Y = aX + b.
    b. Show the fit using Kernel regression with Gaussian kernel and an appropriately chosen bandwidth.
    c. Show the fit using Kernel local linear regression for an appropriately chosen bandwidth
    d. Suggest a linear regression model $Y = \sum_I \phi_i(X)$ which fits the data well. Why might you prefer this model to the kernel local linear regression model from part 3)?


Y
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
X

Neural Networks

(Neurons and biological motivation. Linear threshold units. Perceptrons: representational limitation and gradient descent training, Perceptron learning rule, Hebbian learning rule, Delta Learning rule, Loss Functions (L1 loss, L2 loss, Cross-Entropy), Multilayer networks and Backpropagation Learning Algorithm, Feed Forward, Activation Functions, Types of Neural Network Architecture, Bias-Variance trade-off. Regularization and model/feature selection, Optimizers (Gradient Descent, Adagrad, RMSProp, Adam), Learning Rate)

1. What are neural network? How does they behave like classifier? Explain with example.
2. What is the biological motivation behind using the artificial neural network? Define artificial neural networks and list down the applications in which ANN is use.
3. Define perceptron with a well labelled diagram. Why the perceptron cannot be used to implement the EXCLUSIVE-OR function?
4. List and explain the perceptron training rules.
5. What is linear separability? Give example of linearly separable and inseparable problem?
6. Explain gradient descent algorithm and also derive Gradient Descent rule
7. Explain the Limitation with Back Propagation Algorithm
8. Explain back propagation algorithm and derive expressions for weight update relations
9. Explain Hebbian Learning rule.
10. Explain Neural Network architecture based on Learning Method.
11. Explain various activation functions of neural network.
12. Mention the characteristics of problems suitable for ANNs.
13. List some applications of ANNs.
14. What are the design parameters of ANN?
15. Mention the linear and nonlinear activation functions used in Artificial Neural Networks.
16. What are the design steps to be followed for using ANN for your problem?
17. Explain the three classifications of ANNs based on their functions. Explain them in brief.
18. Develop simple ANNs to implement the three input AND, OR and XOR functions.
19. Draw the architecture of a Multilayer perceptron (MLP) and explain its operation. Mention its advantages and disadvantages.
20. A two-layer network is to have four inputs and six outputs. The range of the outputs is to be continuous between 0 and 1. What can you tell about the network

architecture? Specifically,

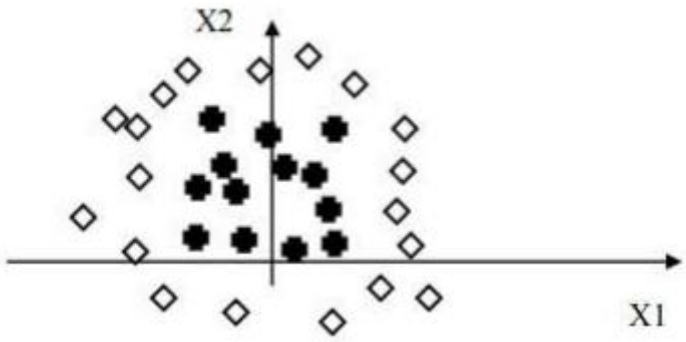
I. How many neurons are required in each layer?

II. What are the dimensions of the first-layer and second layer weight matrices? (Hidden layer neurons are 5)

III. What kinds of transfer functions can be used in each layer?

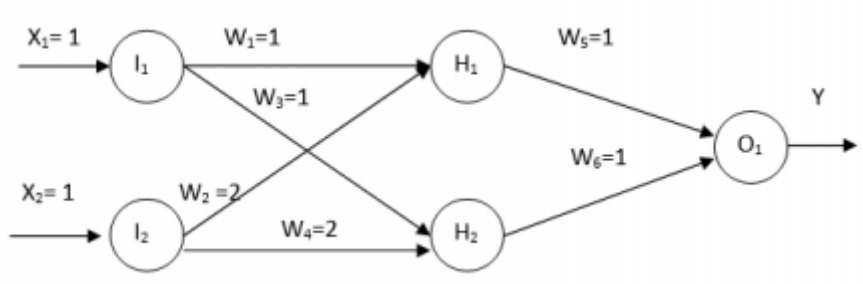
21. Assume we have a set of data from patients who have visited UPMC hospital during the year 2011. A set of features (e.g., temperature, height) have been also extracted for each patient. Our goal is to decide whether a new visiting patient has any of diabetes, heart disease, or Alzheimer (a patient can have one or more of these diseases).
22. We have decided to use a neural network to solve this problem. We have two choices: either to train a separate neural network for each of the diseases or to train a single neural network with one output neuron for each disease, but with a shared hidden layer. Which method do you prefer? Justify your answer.
23. What is Machine Learning? What is the need of it? Briefly explain the Machine Learning characteristic.
24. Explain in brief Descriptive and Predictive Tasks of Machine Learning.
25. How machine learning uses computer algorithms to search for patterns in data?
26. Discuss the BP (Back Propagation) NN (Neural Network) learning algorithm along with its mathematical derivation. Discuss briefly about the limitations of Back Propagation learning Algorithm.
27. What is Artificial Neural Network? What are the benefits of Artificial Neural Network? What are various Network Architectures?
28. What is the effect of learning rate α in the performance of gradient descent? Also justify the statement: "Gradient Descent can convergence to a local mimimum, with the learning rate α is fixed."
29. What is over-fitting and under-fitting? Which are the reasons for over-fitting and under-fitting?
30. What is overfitting catalyst? Reasons for Overfitting.
31. What is Classification? Which are the phases of Classification process? Differentiate between Binary Classification and Multiclass Classification
32. Suppose $\theta 1$is at local optimum of $J(\theta 1)$ such as shown in the figure. What will one step of gradient descent will do? Justify your answer.
33. Differentiate: Supervised Learning V/S Unsupervised Learning.
34. Define Machine Learning. Describe the steps in designing learning system.
35. Explain artificial neural network based on perceptron concept with diagram.
36. What is gradient descent and Delta Rule? Why stochastic approximation to gradient descent is needed?
37. Describe the multilayer Neural Network. Explain why back propagation algorithm is required?

38. Derive the back propagation rule considering the output layer and training rule for output unit weights.
39. List out and explain in briefly representation power of feed forward networks
40. Explain following (2 marks each)
    I. Machine Learning
    II. Information Gain
    III. Back Propagation
41. What is a Neural Network (NN)? What types of problems are suitable WITH NN? Explain Hidden Layer with suitable example.
42. With a suitable example explain back propagation in Neural Network?
43. With a suitable example, explain Face Recognition using Machine Learning.
44. With a suitable method, identify handwritten characters [0-9] using appropriate machine learning technique.
45. A dealer has a warehouse that stores a variety of fruits and vegetables. When fruit is brought to the warehouse, various types of fruit may be mixed together. The dealer wants a model that will sort the fruit according to type. Justify with reasons how machine learning model is efficient compared to feature based classification technique.
46. Suppose you are only allowed to use binary logistic classifiers to solve a multi-class classification problem. Given a training set with 2 classes, this classifier can learn a model, which can then be used to classify a new test point to one of the 2 classes in the training set. You are now given a 6 class problem along with its training set, and have to use more than one binary logistic classifier to solve the problem, as mentioned before. Propose the following scheme - you will first train a binary logistic classifier for every pair of classes. Now, for a new test point, you will run it through each of these models, and the class which wins the maximum number of pairwise contests, is the predicted label for the test point. How many binary logistic classifiers will you need to solve the problem using your proposed scheme?
47. Explain the effect of following factors in achieving global minima with gradient descent algorithm.
    I. Epochs
    II. Learning rate
    III. Weights of hypothesis (Θ0 and Θ1)
48. Suppose that we want to build a neural network that classifies two dimensional data (i.e., X = [x1, x2]) into two classes: diamonds and crosses. We have a set of training data that is plotted as follows:



Draw a network that can solve this classification problem. Justify your choice of the number of nodes and the architecture. Draw the decision boundary that your network can find on the diagram.

49. Consider the following Neural Network with alpha = 0.5, eta=0.24, desired output = 1 and sigmoid activation function.
    I. Perform one forward pass and calculate the error
    II. Calculate the updated weights for w5 and w6 using back- propagation



50. What is linearly in separable problem? Design a two layer network of perceptron to implement XOR and AND gates.
51. Explain how an agent can take action to move from one state to other state with the help of rewards.
52. NASA wants to be able to discriminate between Martians (M) and Humans (H) based on the following characteristics: Green ∈{N, Y} , Legs ∈{2,3} , Height ∈{S, T}, Smelly ∈{N, Y} Our available training data is as follows:

	Species	Green	Legs	Height	Smelly
1	M	N	3	S	Y
2	M	Y	2	T	N
3	M	Y	3	T	N
4	M	N	2	S	Y
5	M	Y	3	T	N
6	H	N	2	T	Y
7	H	N	2	S	N
8	H	N	2	T	N
9	H	Y	2	S	N
10	H	N	2	T	Y

53. Define Machine learning? Briefly explain the types of learning.
54. What is the role of radial basis function in separating nonlinear patterns?
55. What are the steps in designing a machine learning problem? Explain with the checkers problem.
56. Show 5 updates of perceptron algorithm for the following data, starting W = (0,0): {(-, (1, -4)), (+1, (5, -1)), (+, (3,3)), (+, (-1,5)), (-, (-5, 2)), (-, (-2, -2))}
57. Discuss the two approaches to prevent over fitting of data.
58. What is the role of weights and bias?
59. What are the benefits of mini-batch gradient descent?
60. Why is Weight Initialization important in Neural Networks?
61. In training a neural network, you notice that the loss does not decrease in the few starting epochs. What could be the reason?

62. Consider the following set of training examples:

Instance	Classification	a1	a2
1	+	T	T
2	+	T	T
3	-	T	F
4	+	F	F
5	-	F	T
6	-	F	T

I. What is the entropy of this collection of training examples with respect to the target function classification?
II. What is the information gain of a2 relative to these training examples?

63. Consider the set of training examples:

Instance	Classification	A1	A2
1	1	1	1
2	1	1	1
3	0	1	0
4	1	0	0
5	0	0	1
6	0	0	1

I. What is the entropy of this collection of training examples with respect to the target function classification?
II. What is the information gain of a2 relative to these training examples?

64. What are the requirements of clustering algorithms?
65. What is independent component analysis?
66. What Is the Difference Between Batch Gradient Descent and Stochastic Gradient Descent?
67. Consider the given below training example which finds malignant tumors from MRI scans.

Example	Shape	Size	Color	Surface	Thickness	Target Concept
1	Circular	Large	Light	Smooth	Thick	Malignant
2	Circular	Large	Light	Irregular	Thick	Malignant
3	Oval	Large	Dark	Smooth	Thin	Benign
4	Oval	Large	Light	Irregular	Thick	Malignant
5	Circular	Small	Light	Smooth	Thick	Benign

Show the specific and general boundaries of the version space after applying candidate elimination algorithm. (Note: Malignant is +ve, Benign is -ve)

Density Based Clustering

(DBScan)

1. What is density based clustering?
2. What are the advantages of DBSCAN?
3. What are the disadvantages of DBSCAN?
4. What is DBSCAN? When to use it?

Prepared by:
Priyal Vaghela
Assistant Professor
DEPSTAR – CSE Dept.