Week_7_mlt_questions

2024 Mar 24 Quiz 2

1. **Comprehension:**
Consider a binary classification problem with a training dataset of 80 points, evenly distributed between two classes (40 points in each class). You decide to train a k-NN algorithm with $k = 3$. Each point is considered its own neighbor during classification.
Based on the above data, answer the given subquestions.

i. What is the minimum number of misclassifications that can occur in the training dataset when using this k-NN algorithm?

[Ans: 0]

ii. Assuming there are outliers, the decision boundary becomes smoother with increasing value of k in a k-NN algorithm (Fill in 1 for yes and 0 for no).

[Ans: 1]

2. Select all true statements. (MSQ)

(a) In Decision tree, if a question Q_1 is "better" than question Q_2, then information gains for Q_1 is greater than information gains Q_2 always.

(b) The training dataset is required while predicting the label of a test-point in the k-NN algorithm.

(c) A question of the form $f_k \leq \theta$ always partitions the dataset into two non-empty sets.

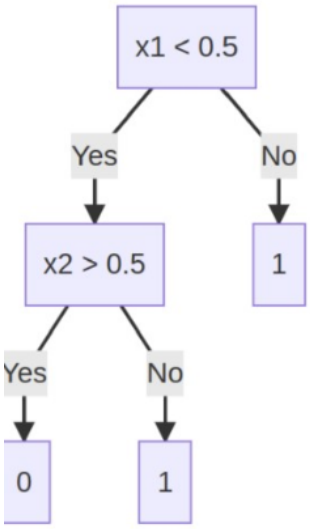
(d) The depth of the tree is a hyperparameter and has to be chosen using cross validation.

(e) Decision trees are prone to overfit if the maximum depth is set too low.

[Ans: a, b, d]

2024 Apr 28 ET

3. Consider the following decision tree for a classification problem in which all the data-points are constrained to lie in the unit square in the first quadrant. That is $0 \leq x_1, x_2 \leq 1$. If a point is picked uniformly at random from the unit square, what is the probability that the decision tree predicts this point as belonging to class 1?



[Ans: 0.75]

2024 Aug 04 Quiz 2

4. Which of the following classifiers are certainly capable of achieving zero training error for any binary classification problem with features in $\mathbb{R}^2$? Note: while computing the training error for KNN, each data-point is its own neighbor. (MSQ)

(a) Decision tree

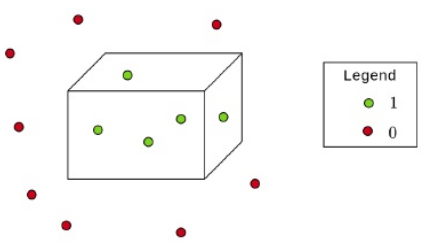
(b) KNN with $k = 1$

(c) KNN with $k = 5$

(d) Naive Bayes

[Ans: a, b]

5. Consider the following decision boundary obtained for a decision tree trained on a dataset in $\mathbb{R}^3$ for a binary classification problem. All green points (label 1) lie inside the box and the red points (label 0) are outside it. What is the minimum depth of this tree? Recall that the depth is the number of edges on the longest path from the root to a leaf.



[Ans: 6]

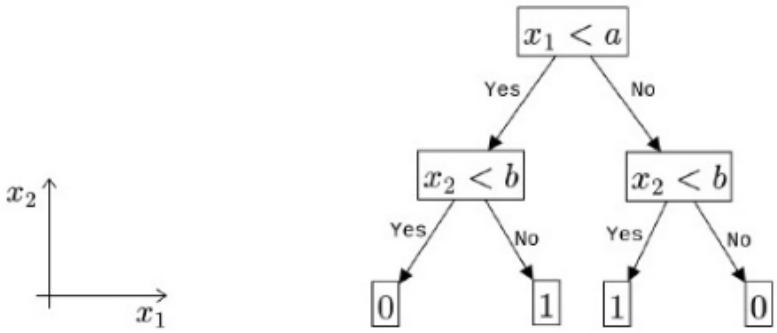
6. **Comprehension:**

If a and b are integers, answer the given subquestions that follow. Hint: Plotting the dataset helps.

Consider the following training dataset for a binary classification problem with the data-points in $\mathbb{R}^2$, the labels in the set $\{0, 1\}$ and a decision tree grown from it that achieves zero training error:

$$\mathbf{X} = \begin{bmatrix} -2 & -0.9 & 0.9 & 2 & -2 & -0.9 & 0.9 & 2 \\ 1 & 0.9 & 0.9 & 1 & -1 & -0.9 & -0.9 & -1 \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}^T$$



i. What is the value of a?

[Ans: 0]

ii. What is the value of b?

[Ans: 0]

iii. Find the entropy of the root node. Use $\log_2$.

[Ans: 1]

iv. If E_R is the entropy of the root and E_L is the weighted entropy of all the leaves, what is $E_R - E_L$? Use $\log_2$.

[Ans: 1]

v. What is the predicted label of the test data-point $\begin{bmatrix} 3 \\ -4 \end{bmatrix}$?

[Ans: 1]

2024 Dec 01 Quiz 2

7. Consider the following feature vectors in $\mathbb{R}^3$:

$$x_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad x_2 = \begin{bmatrix} -1 \\ -2 \\ 0 \end{bmatrix}, \quad x_3 = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}, \quad x_4 = \begin{bmatrix} 0 \\ 1 \\ 4 \end{bmatrix}, \quad x_5 = \begin{bmatrix} 2 \\ -1 \\ -3 \end{bmatrix}$$

The labels of these points are $y_1 = 0$, $y_2 = 1$, $y_3 = 1$, $y_4 = 0$, $y_5 = 1$.
If we use a k-NN algorithm with $k = 3$, what would be the predicted label for the test point $x_{\text{test}} = \begin{bmatrix} 1 & 1 & 2 \end{bmatrix}^T$?

[Ans: 0]

8. Consider a decision stump (parent with two children). The parent node has 200 data-points out of which 50 belong to the positive class. The left child has 100 data-points out of which 50 belong to the positive class. Find the information gain. Use $\log_2$. Enter your answer correct to three decimal places.

[Ans: 0.26 to 0.37]

9. **Comprehension:**
Consider a binary classification problem with a training dataset of 100 points, evenly distributed between two classes (50 points in each class). You decide to train a k-NN algorithm with $k = 3$. Each point is considered its own neighbor during classification.
Based on the above data, answer the given subquestions.

i. What is the minimum number of misclassifications that can occur in the training dataset when using this k-NN algorithm?

[Ans: 0]

ii. Assuming there are outliers, the decision boundary becomes smoother with decreasing value of k in a k-NN algorithm.

(a) True
(b) False

[Ans: b]

10. **Comprehension:**
Consider a binary classification problem in which a decision tree is classifying data points into two classes, A and B. In a particular node of the tree, 60% of the data points belong to class A, while the remaining 40% belong to class B.
Based on the above data, answer the given subquestions.

i. Do you have enough information to find the entropy of this node?

(a) Yes
(b) No

[Ans: a]

ii. If the answer to the previous question is "Yes," calculate the entropy of this node to three decimal places. If the answer to the previous question is "No," enter -1 as your answer.

[Ans: 0.94 to 1.01]

2025 Mar 16 Quiz 2

11. p is the proportion of points with label 1 in some node in a decision tree. Which of the following statements are true?

(a) As the value of p increases from 0 to 1, the impurity of the node increases.

(b) As the value of p increases from 0 to 1, the impurity of the node decreases.

(c) The impurity of the node does not depend on p.

(d) $p = 0.5$ correspond to the case of maximum impurity.

[Ans: d]

12. Consider the following feature (x) along with its label (y):

x	y
1	1
-1	1
2	1
-2	0
3	0
4	1

If you are provided with the following two questions, which one should you choose for the first (top) node of the decision tree?
Q_a: $x \leq -2$
Q_b: $x \leq -1$
Recall: Information gain = Entropy(D) $-$ γEntropy(D_{Yes}) $-$ $(1 - \gamma)$Entropy(D_{No}).

Use the following values if required:
$\log_2(3/4) = -0.41$, $\log_2(4/5) = -0.32$, $\log_2(1/5) = -2.32$, $\log_2(1/4) = -2$, $\log_2(1/2) = -1$.

(a) $x \leq -2$

(b) $x \leq -1$

(c) Insufficient information

[Ans: a]

13. **Comprehension:**
Consider the following 2D data points along with their labels:

x	$[1\ \ 1]^T$	$[1\ \ -1]^T$	$[2\ \ 1]^T$	$[2\ \ 2]^T$	$[2\ \ -2]^T$	$[3\ \ -3]^T$	$[4\ \ -1]^T$
y	1	1	1	1	0	0	1

Based on the above data, answer the given subquestions.

i. If we use a k-NN algorithm with $k = 3$, what would be the predicted label for $x_{\text{test}} = [4\ \ 0]^T$?

[Ans: 1]

ii. Compute the Leave-one-out cross validation (LOOCV) error for the 3-nearest neighbour (3-NN) algorithm on this dataset. Enter the answer correct to two decimal places.

[Ans: 0.41 to 0.45]

iii. Is it possible to reduce the cross-validation error obtained in the previous question by changing the label of certain points?

(a) Yes

(b) No

(c) Cannot conclude from the given information

[Ans: a]

2025 Apr 13 ET

14. Which of the following conditions must be satisfied for a decision tree to be pure at a given node? (MSQ)

(a) The entropy of the node is maximum.

(b) The information gain at the node is negative.

(c) All samples at a node belong to the same class.

(d) The entropy of the node is 0.

[Ans: c, d]

2025 Aug 03 Quiz 2

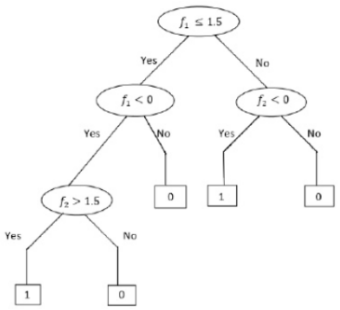
15. Perform a 3-NN regressor for the following dataset:

x_1	x_2	y
1	2	3.5
2	1	2.8
3	3	5.1
4	2	4.7
5	3	6.2
6	1	5.9

where, $x = [x_1 \quad x_2]^T$ is a feature vector and y is the label. What will be the predicted value for the test data point $[3 \quad 2]^T$? Enter the answer correct to one decimal place.

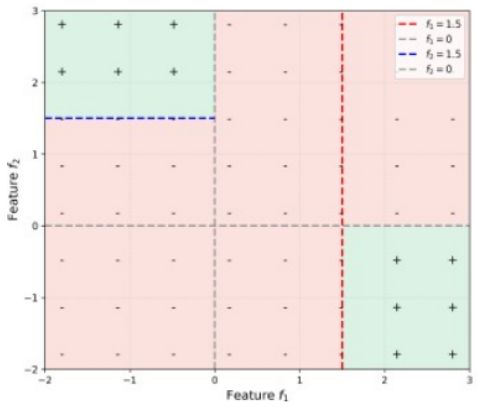
[Ans: 4.2]

16. **Comprehension:**
Consider the following decision tree for a binary classification problem with two features f_1 and f_2:



Based on the above data, answer the given subquestions.

i. Which among the following represents the correct decision boundary for the decision tree? Note: Here + represents the label 1 and - represents the label 0.



[Ans: b]

ii. What will be the prediction for the new datapoint $(2, 2)$?

[Ans: 0]

2025 Aug 31 ET

17. Which of the following statements are true about the decision tree algorithm? (MSQ)

(a) Decision trees are prone to overfitting if the maximum depth is set too shallow.

(b) Decision trees are not affected by the order of features in the dataset.

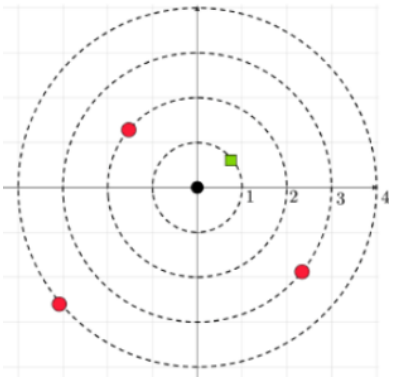
(c) Decision trees are sensitive to small perturbations in the dataset and can result in different tree structures.

(d) Decision trees can handle both numerical and categorical features.

[Ans: c, d]

2026 Aug 14 Quiz 2

18. Consider a training dataset for a classification problem with four data-points. Red circles represent data-points with label -1 and green squares represent data-points with label $+1$. A test-point is located at $(0, 0)$. Dotted circles centered at $(0, 0)$ with radii $1, 2, 3$ and 4 are drawn for your convenience.



KNN with Euclidean distance and $k = 3$ is the classifier. Now, a new training data-point with label $+1$ is added somewhere inside the outermost circle of radius 4 such that it leaves the predicted label of the test data-point unchanged. Let S be the set of all points that could be the potential location of this new training data-point. Find the area of the region corresponding to S.

Note: We use KNN with $k = 3$ for both datasets.

(a) 7π

(b) 5π

(c) 16π

(d) 9π

[Ans: a]

19. **Comprehension:**
Consider a node in a decision tree that has 5 data-points. The i^{th} data-point is represented by a feature-label pair (x_i, y_i), where $y_i \in \{-1, 1\}$.

i	x_i	y_i
1	4	-1
2	10	1
3	13	1
4	25	1
5	38	-1

Consider the split $x \leq 7$ at the parent node resulting in left and right child nodes. The left child node has all data points for which $x \leq 7$ is true. Entropy is used to determine the purity of the nodes. Recall that the formula for the entropy of a node is:

$$E(p) = -p \log_2 p - (1-p) \log_2 (1-p)$$

where p is the fraction of positive data-points in the node. Use at least three digits after the decimal point in all intermediate calculations. Based on the above data, answer the given subquestions.

i. Find the entropy of the parent node. Enter your answer correct to two decimal places.

[Ans: 0.87 to 1.07]

ii. Find the entropy of the right child. Enter your answer correct to two decimal places.

[Ans: 0.71 to 0.91]

iii. Find the information gain. Enter your answer correct to two decimal places.

[Ans: 0.22 to 0.42]

20. Perform a 3-NN regressor for the following dataset:

x_1	x_2	y
2	3	4.1
1	4	3.2
4	2	5.5
3	5	6.8
5	2	7.1
2	1	3.9

where, $x = [x_1 \quad x_2]^T$ is a feature vector and y is the label. What will be the predicted value for the test data point $x_{\text{test}} = [2 \quad 2]^T$? Enter the answer correct to one decimal place.

[Ans: 4.5]

21. Consider a binary classification problem in $\mathbb{R}^2$. The feature space is partitioned into the following four regions:

$$R_1 = \{(x_1, x_2) \mid x_1 \geq 0,\, x_2 \geq 0\}, \quad R_2 = \{(x_1, x_2) \mid x_1 < 0,\, x_2 \geq 0\},$$

$$R_3 = \{(x_1, x_2) \mid x_1 < 0,\, x_2 < 0\}, \quad R_4 = \{(x_1, x_2) \mid x_1 \geq 0,\, x_2 < 0\}.$$

Each region contains the same number of observations. All observations in regions R_1 and R_3 belong to the positive class, while all observations in regions R_2 and R_4 belong to the negative class. A decision tree is constructed using axis-aligned binary splits of the form $x_1 \leq t$ or $x_2 \leq t$, where t is any real number. At each node, the split is chosen to maximize the information gain based on entropy. What is the minimum number of binary questions (splits) that the decision tree must ask to obtain leaf nodes with zero entropy?

(a) 3

(b) 2

(c) 4

(d) It is impossible using only axis-aligned splits.

[Ans: a]

22. A node in a decision tree contains an equal number of positive and negative training examples. What is the entropy of this node?

(a) 0

(b) 0.5

(c) 1

(d) It depends on the total number of training examples.

[Ans: c]

23. Which of the following statements about the K-Nearest Neighbours (KNN) algorithm are correct? (MSQ)

(a) The choice of distance metric can affect the predicted class.

(b) Once trained, KNN makes predictions without referring to the training data.

(c) Every prediction does not require computing distances from the test point to all training data.

(d) The computational cost of prediction generally increases with the size of the training dataset.

[Ans: a, d]

2026 Sep 13 ET

24. Identify the correct relationship(s) between bias, variance, and the hyperparameter k in the k-Nearest Neighbours algorithm. (MSQ)

(a) Increasing k leads to an increase in bias.

(b) Decreasing k leads to an increase in bias.

(c) Increasing k leads to an increase in variance.

(d) Decreasing k leads to an increase in variance.

[Ans: a, d]

25. Which of the following statements about entropy and information gain in decision-tree learning are correct? (MSQ)

(a) If a feature splits a node into perfectly pure children, its information gain at that node equals the entropy of that node.

(b) A node whose examples all belong to a single class has entropy 0 and needs no further splitting.

(c) Information gain can be negative if a split increases impurity.

(d) After a parent node is split into child nodes, no individual child node can have entropy greater than the parent node's entropy.

[Ans: a, b]