B.

Which of the following is true about K-means?

A. K-means clustering aims to classify the objects into groups that with k objects in each group.

B. K-means is guaranteed to converge to a local minimum.

C. The initialization of the centroids has no affect on the final result of K-means.

D. K-means is guaranteed to converge to a global minimum.

D.

Which of the following statement is NOT true?

A. Naïve Bayes is a supervised learning method.

B. Naive Bayes assumes that all the features in a dataset are independent.

C. Naive Bayes assumes that all the features in a dataset are equally important.

D. Naive Bayes can be only used for binary classification problems.

A.

You've just finished training a decision tree for spam classification, and it is getting abnormally bad performance on both of your training and test sets. Suppose that your implementation has no bugs, so which of the following could be the problem?

A. The decision trees are too shallow.

B. The learning rate is too small.

C. The model is overfitting.

D. None of the above.

A.

Which of the following statement is INCORRECT about Random Forest?

A. Random Forest can not handle binary features.

B. Random Forest can be used for classification task.

C. Random Forest can be used for regression task.

D. Random Forest has multiple decision trees as base learning models.

Which of the following statements is INCORRECT about PCA?

A. We must standardize the data first.

B. We should select the principal components which explain the highest variance.

C. PCA components are not always orthogonal.

D. We can use PCA for visualizing the data in lower dimensions.

Which is not the application of language processing?

A. Textual Entailment

B. Spelling Correction

C. Information Retrieval

D. Latent Space Learning

Inappropriate selection of learning rate value in gradient descent gives rise to:

A. Local Minima

B. Oscillations

C. Slow convergence

D. All of the above

C.

Which of the following techniques is "eigenfaces" build on?

A. Non-negative matrix factorization

B. Independent Component Analysis

C. Singular Value Decomposition

D. Support Vector Machine

B.

The point where the Null Hypothesis gets rejected is called as?

A. Rejection Value

B. Critical Value

C. Significant Value

D. Acceptance Value

Consider a communication system that consists of three messages with probabilities of P1= 0.25, P2= 0.25 and P3= 0.25, P4= 0.25. The entropy of the system is (the base of the logarithm is 2):

A. 1

B. 2

C. 1.5

D. 3

A.

Which of the following does NOT use machine learning/AI?

A. Walkman (tape player)

B. Self-driving cars

C. SIRI/Alexa

D. Facial recognition on App on your phone

Which of the following is the supervised learning method?

A. Expectation Maximization

B. K-means

C. Decision Tree

D. Hierarchical clustering

Which of the following method is used in data cleaning?

A. Data munging

B. Handling missing data

C. Smooth noisy data

D. All of the above

Which of the following is NOT considered as an acoustic signal processing task?

A. Abnormality Detection

B. Audio Captioning 音频字幕

C. Automatic Music Generation (Symbolic Data)

D. Acoustic Event Detection

A.

If our data has no labels or true values associated with them, what type of learning may be used?

A. Unsupervised learning

B. Supervised learning

C. Semi-supervised

D. All of the above

B.

Which of the following statement is true for k-fold cross-validation?

A. Higher values of k will result in higher variance on the cross-validation result. ✗

B. The overall accuracy of the model is the average error across all k trials.

C. The number of data points must be larger than k. ✗

D. Every data point has the chance to be in the training set exactly once. (k-1 次) ✓

A

Consider the following data table:

X	Y
-2	0
-1	+2
+1	+4
+2	+6

$J = \sum (w_0 + w_1 \cdot x - y)^2$

$\frac{\partial J}{\partial w_1} = \sum 2x(w_0 + w_1 x - y) = 2(\sum x \cdot w_0 + \sum x^2 w_1 - \sum xy) = 0$

$\frac{\partial J}{\partial w_0} = \sum 1 \cdot (w_0 + w_1 x - y) = 0$

$\begin{cases} 10 w_1 - 14 = 0 \\ 4 w_0 + 4 = 0 \end{cases}$

$w_1 = \frac{7}{5} = 1.4$

$w_0 = 3$

A linear regression model: f(x) = ω0 + ω1 * x, is fit to the data using the least square method. What are the optimal parameters?

A. $\omega_0 = 3$ and $\omega_1 = 1.4$

B. $\omega_0 = 2$ and $\omega_1 = 1$

C. $\omega_0 = 2.4$ and $\omega_1 = 2$

D. $\omega_0 = 3$ and $\omega_1 = 2.4$