

MACHINE LEARNING

In Q1 to Q7, only one option is correct, Choose the correct option:

1. What is the advantage of hierarchical clustering over K-means clustering?
A) Hierarchical clustering is computationally less expensive
B) In hierarchical clustering you don't need to assign number of clusters in beginning
C) Both are equally proficient D) None of these
Answer – B) **In hierarchical clustering you don't need to assign number of clusters in beginning**
2. Which of the following hyper parameter(s), when increased may cause random forest to over fit the data?
A) **max_depth** B) n_estimators
C) min_samples_leaf D) min_samples_splits
Answer – A) **max_depth**
3. Which of the following is the least preferable resampling method in handling imbalance datasets?
A) SMOTE B) RandomOverSampler
C) **RandomUnderSampler** D) ADASYN
Answer – A) **RandomUnderSampler**
4. Which of the following statements is/are true about "Type-1" and "Type-2" errors?
 1. Type1 is known as false positive and Type2 is known as false negative.
 2. Type1 is known as false negative and Type2 is known as false positive.
 3. Type1 error occurs when we reject a null hypothesis when it is actually true.A) 1 and 2 B) **1 only**
C) 1 and 3 D) 2 and 3
Answer – B) **1 only**
5. Arrange the steps of k-means algorithm in the order in which they occur:
 1. Randomly selecting the cluster centroids
 2. Updating the cluster centroids iteratively
 3. Assigning the cluster points to their nearest centerA) **3-1-2** B) 2-1-3
C) 3-2-1 D) 1-3-2
Answer – A) **3-1-2**
6. Which of the following algorithms is not advisable to use when you have limited CPU resources and time, and when the data set is relatively large?
A) Decision Trees B) **Support Vector Machines**
C) K-Nearest Neighbors D) Logistic Regression
Answer – B) **Support Vector Machines**

MACHINE LEARNING

7. What is the main difference between CART (Classification and Regression Trees) and CHAID (Chi Square Automatic Interaction Detection) Trees?
- A) CART is used for classification, and CHAID is used for regression.
 - B) CART can create multiway trees (more than two children for a node), and CHAID can only create binary trees (a maximum of two children for a node).
 - C) CART can only create binary trees (a maximum of two children for a node), and CHAID can createmultiway trees (more than two children for a node)**
 - D) None of the above

Answer – C) **CART can only create binary trees (a maximum of two children for a node), and CHAID can createmultiway trees (more than two children for a node)**

In Q8 to Q10, more than one options are correct, Choose all the correct options:

8. In Ridge and Lasso regularization if you take a large value of regularization constant(λ), which of the following things may occur?
- A) Ridge will lead to some of the coefficients to be very close to 0
 - B) Lasso will lead to some of the coefficients to be very close to 0**
 - C) Ridge will cause some of the coefficients to become 0
 - D) Lasso will cause some of the coefficients to become 0.

Answer – B) **In Ridge and Lasso regularization if you take a large value of regularization constant(λ), whichof the following things may occur**

MACHINE LEARNING

9. Which of the following methods can be used to treat two multi-collinear features?

- A) remove both features from the dataset
- B) remove only one of the features
- C) **Use ridge regularization**
- D) use Lasso regularization

Answer – C) **Use ridge regularization**

10. After using linear regression, we find that the bias is very low, while the variance is very high. What are the possible reasons for this?

- A) **Overfitting**
- B) Multicollinearity
- C) Underfitting
- D) Outliers

Answer – A) **Overfitting**

Q10 to Q15 are subjective answer type questions, Answer them briefly.

11. In which situation One-hot encoding must be avoided? Which encoding technique can be used in such a case?

Answer - One-hot encoding creates d-dimensional vectors for each instance where d is the unique number of feature values in the dataset. For a feature having a large number of unique feature values or categories, one-hot encoding is not a great choice.

To fight the curse of dimensionality, binary encoding might be a good alternative to one-hot encoding because it creates fewer columns when encoding categorical variables. Ordinal encoding is a good choice if the order of the categorical variables matters.

12. In case of data imbalance problem in classification, what techniques can be used to balance the dataset? Explain them briefly.

Answer – Imbalanced datasets are a special case for classification problem where the class distribution is not uniform among the classes. Typically, they are composed by two classes: The majority (negative) class and the minority (positive) class

1. Random Undersampling and Oversampling - A widely adopted and perhaps the most straightforward method for dealing with highly imbalanced datasets is called resampling. It consists of removing samples from the majority class (under-sampling) and/or adding more examples from the minority class (over-sampling).

2. Undersampling and Oversampling using imbalanced-learn

imbalanced-learn(imblearn) is a Python Package to tackle the curse of imbalanced datasets.

3. Class weights in the models

MACHINE LEARNING

13. What is the difference SMOTE and ADASYN sampling techniques between?

Answer - The key difference between ADASYN and SMOTE is that the former uses a density distribution, as a criterion to automatically decide the number of synthetic samples that must be generated for each minority sample by adaptively changing the weights of the different minority samples to compensate for the skewed .

14. What is the purpose of using GridSearchCV? Is it preferable to use in case of large datasets? Why or why not?

Answer – GridSearchCV is a technique for finding the optimal parameter values from a given set of parameters in a grid. It's essentially a cross-validation technique. The model as well as the parameters must be entered. After extracting the best parameter values, predictions are made.

15. List down some of the evaluation metric used to evaluate a regression model. Explain each of them in brief.

Answer - There are three error metrics that are commonly used for evaluating and reporting the performance of a regression model; they are:

 FLIP ROBO

1. Mean Squared Error (MSE) - the mean squared error (MSE) or mean squared deviation (MSD) of an estimator (of a procedure for estimating an unobserved quantity) measures the average of the squares of the errors—that is, the average squared difference between the estimated values and the actual value. MSE is a risk function, corresponding to the expected value of the squared error loss. The fact that MSE is almost always strictly positive (and not zero) is because of randomness or because the estimator does not account for information that could produce a more accurate estimate. In machine learning, specifically empirical risk minimization, MSE may refer to the *empirical* risk (the average loss on an observed data set), as an estimate of the true MSE (the true risk: the average loss on the actual population distribution)
2. Root Mean Squared Error (RMSE) - Root mean square error or root mean square deviation is one of the most commonly used measures for evaluating the quality of predictions. It shows how far predictions fall from measured true values using Euclidean distance.
3. Mean Absolute Error (MAE) - Mean absolute error (MAE) is a measure of errors between paired observations expressing the same phenomenon. Examples of Y versus X include comparisons of predicted versus observed, subsequent time versus initial time, and one technique of measurement versus an alternative technique of measurement.