Machine Learning with PythonCognitive Class ML0101ENv3

Techelite solutions [Follow]

Sep 15, 2019 · 4 min read

Module -1 Machine Learning :

1. Machine Learning uses algorithms that can learn from data without relying on explicitly programmed methods. — True

2. Which are the two types of Supervised learning techniques? — Classification and Regression Classification and Regression

3. Which of the following statements best describes the Python scikit library — A collection of algorithms and tools for machine learning.

Module -2 Regression :

1. Train and Test on the Same Dataset might have a high training accuracy, but its out-of-sample accuracy can be low. — True

2. Which of the following matrices can be used to show the results of model accuracy evaluation or the model's ability to correctly predict or separate the classes? — Confusion matrix

3. When we should use Multiple Linear Regression? — When we would like to identify the strength of the effect that the independent variables have on a dependent variable.

Module -3 Classification :

1. In K-Nearest Neighbors, — A very high value of K (ex. K = 100) produces an overly generalised model, while a very low value of k (ex. k = 1) produces a highly

complex model.

2. A classifier with lower log loss has better accuracy. — True

3. When building a decision tree, we want to split the nodes in a way that decreases entropy and increases information gain. — True

Module -4 Clustering :

1. This one is NOT TRUE about k-means clustering — As k-means is an iterative algorithm, it guarantees that it will always converge to the global optimum.

2. Customer Segmentation is a supervised way of clustering data, based on the similarity of customers to each other. — False

3. How is a center point (centroid) picked for each cluster in k-means? — We can randomly choose some observations out of the data set and use these observations as the initial means.

Module -5 Recommender Systems :

1. Collaborative filtering is based on relationships between products and people's rating patterns. — True

2. Which one is TRUE about Content-based recommendation systems? — Content-based recommendation system tries to recommend items to the users based on their profile.

3. Which one is correct about user-based and item-based collaborative filtering? — In user-based approach, the recommendation is based on users of the same neighborhood, with whom he/she shares common preferences.

Knowledge Test :

1. You can define Jaccard as the size of the intersection divided by the size of the union of two label sets. — True

2. When building a decision tree, we want to split the nodes in a way that increases entropy and decreases information gain. — False

3. CN

4. To calculate a model's accuracy using the test set, you pass the test set to your model to predict the class labels, and then compare the predicted values with actual values. — True

5. Which is the definition of entropy? — The amount of information disorder in the data.

6. Which of the following is true about hierarchical linkages? — Average linkage is the average distance of each point in one cluster to every point in another cluster Average linkage is the average distance of each point in one cluster to every point in another cluster

7. The goal of regression is to build a model to accurately predict the continues value of a dependent variable for an unknown case. — True

8. Which of the following statements are true about linear regression? (Select all that apply) — (a)With linear regression, you can fit a line through the data. ; (b) y=a+b_x1 is the equation for a straight line, which can be used to predict the continuous value y.

9. The Sigmoid function is the main part of logistic regression, where Sigmoid of $\theta^\wedge T.X$, gives us the probability of a point belonging to a class, instead of the value of y directly. — True

10. In comparison to supervised learning, unsupervised learning has: EITHER Less tests (evaluation approaches) OR A better controlled environment

11. The points that are classified by Density-Based Clustering and do not belong to any cluster, are outliers. — True

12. Which of the following is false about Simple Linear Regression? — It is used for finding outliers It is used for finding outliers

13. Which one of the following statements is the most accurate? — Machine Learning is the branch of AI that covers the statistical and learning part of artificial intelligence.

14. Which of the following are types of supervised learning? — (a) Classification (b)Regression (c)KNN

15. A Bottom-Up version of hierarchical clustering is known as Divisive clustering. It is a more popular method than the Agglomerative method. — False

16. Select all the true statements related to Hierarchical clustering and K-Means. — Hierarchical clustering does not require the number of clusters to be specified.;K-Means is more efficient than Hierarchical clustering for large datasets.

17. What is a content-based recommendation system? — Content-based recommendation system tries to recommend items to the users based on their profile built upon their preferences and taste.

18. Before running Agglomerative clustering, you need to compute a distance/proximity matrix, which is an n by n table of all distances between each data point in each cluster of your dataset. — True

19. CN

20. In recommender systems, "cold start" happens when you have a large dataset of users who have rated only a limited number of items. — False

Machine Learning   Python   Python Programming

About   Help   Legal

Get the Medium app