CLASS 12 07-06-2021

**QUESTIONS**

👉 What is pandas dummy variable?  
👉What is the difference between OneHotEncoder and Get\_dummies?  
👉What is Sklearn LabelEncoder?  
👉What is the difference between LabelEncoder and Get\_dummies?  
👉What is unsupervised learning?  
👉Where is unsupervised learning used?  
👉What is unsupervised learning example?  
👉What is the difference between supervised and unsupervised learning?  
👉What is meant by K-means clustering?  
👉What are the basic steps for K-means clustering?

**ANSWERS**

1. A **dummy variable** is a binary variable that indicates whether a separate categorical variable takes on a specific value. We can create dummy variables in python using get\_dummies() method.

2. **OneHotEncoder** cannot process string values directly. If your nominal features are strings, then you need to first map them into integers. pandas. **get\_dummies** is kind of the opposite.

3. Sklearn provides a very efficient tool for encoding the levels of categorical features into numeric values. **LabelEncoder** encode labels with a value between 0 and n\_classes-1 where n is the number of distinct labels. If a label repeats it assigns the same value to as assigned earlier.

4. **LabelEncoder** is used to convert a categorical variable class to numerical, for example, take one variable having multiple classes then it will convert to numeric like 0,1,2,3,4. **get\_dummies** is used to convert categorical variables to dummy variables for easy comparison between variables.

5. Unsupervised learning, also known as **unsupervised machine learning**, uses machine learning algorithms to analyze and cluster unlabeled datasets. These algorithms discover hidden patterns or data groupings without the need for human intervention.

6. Some **use cases for unsupervised learning** — clustering — include: Customer segmentation, or understanding different customer groups around which to build marketing or other business strategies. Genetics, for example clustering DNA patterns to analyze evolutionary biology.

7. **Examples of unsupervised learning**

1. Customer segmentation, or understanding different customer groups around which to build marketing or other business strategies.
2. Genetics, for example clustering DNA patterns to analyze evolutionary biology.

8. **Unsupervised learning** is a machine learning technique, where you do not need to supervise the model. **Supervised learning** allows you to collect data or produce a data output from the previous experience. Unsupervised machine learning helps you to finds all kind of unknown patterns in data.

9. **K-means clustering** is a type of unsupervised learning, which is used when you have unlabelled data (i.e., data without defined categories or groups). The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K. Data points are clustered based on feature similarity.

10. **Step of KMeans clustering**:

1. Randomly select ‘c’ cluster centers.
2. Calculate the distance between each data point and cluster centers.
3. Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers..
4. Recalculate the new cluster center
5. Recalculate the distance between each data point and new obtained cluster centers.
6. If no data point was reassigned then stop, otherwise repeat from step 3).