:≡ Course Content

FAQ - K-means Clustering

1. What are some applications of clustering in real-world scenarios?

Common applications of clustering include

Customer Segmentation
Document Clustering
Image Segmentation
Recommendation Engines

2. What is K-means clustering?

K-means is a centroid-based algorithm, or a distance-based algorithm, where we calculate the distances to assign a point to a cluster. In K-Means, each cluster is associated with a centroid.

3. What are some good things about K-means clustering?

It is very smooth in terms of interpretation and resolution.
For a large number of variables present in the dataset, K-means operates quicker than hierarchical clustering.
While redetermining the cluster center, an instance can modify the cluster.
K-means reforms compact clusters.
It can work on unlabeled numerical data.

4. What are the limitations of K-means clustering?

Sometimes, it is quite tough to figure out the appropriate number of clusters, or the value of k.
The output is highly influenced by original input, for example, the number of clusters.
It gets affected by the presence of outliers in the data set.
In some cases, clusters show complex spatial views, then executing clustering is not a good choice.

5. Is there any metric to compare clustering results?

You can compare clustering results by checking silhouette scores and by doing cluster profiling. Besides this, you should also validate the clustering result by consulting with a domain expert to see if the cluster profiles

make sense or not.

6. For K-means if there is a y-dependent variable, do we remove it before trying to group customers?

Yes, if you have a dependent variable in your dataset, you should remove that before applying clustering algorithms on your dataset.

7. I am not able to load the yellowbrick library and getting the following error:

```
ModuleNotFoundError: No module named 'yellowbrick'
```

Can you help?

If the yellowbrick library is already installed, then the following command can be run in the Anaconda prompt:

```
pip install yellowbrick --user
```

8. I have installed the yellowbrick package and updated it to the latest version. But I am still having issues importing the KElbow visualizer and getting the following error:

```
ImportError: cannot import name 'safe_indexing' from 'sklearn.utils'
```

How to fix this?

This error generally occurs due to version differences between sklearn and yellowbrick. Please update the sklearn library by running the below code in the Anaconda prompt:

```
pip install -U scikit-learn --user
```

< Previous   Next >

Privacy Terms of service Help