

MACHINE LEARNING – WORKSHEET (CLUSTERING) SOLUTIONS

1. D
2. D
3. C
4. B
5. D
6. C
7. D
8. A
9. A
10. B
11. A
12. B

13. Clustering is important as, as a data scientist you will not get labelled data in your datasets. Most of the problems are not segregated into classes. Clustering helps to find out the patterns in the data without taking any particular variable in considerations. This is could very well help us to find the relations between the features of the data. After clustering we could treat the records based on its group. We can also find similarly patterns in the records which could be uses for recommendation system applications
14. Clustering is an unsupervised technique. Based on the data, we create cluster and represent into graphical methods for better visualizations and also for formation of optimal clusters. Once the cluster are formed we perform inferential as well as descriptive statistical analysis on the cluster and find relational patterns between them. Once this is done, we have to present and explain the cluster analysis to the client or the team, this is known as cluster profiling. We create a table of features that each of the cluster's trait. We have to describe the founding to the client of the team in a form of a story and give valuable perks which will help them to work on each cluster with respect to the traits we found in the statistical analysis steps.
15. Improving the clustering performing rather a subject and domain specific tasks. But there are some steps that we could perform to ensure betterment of the performance. First of all the more the data the better the results. Also this goes without saying that garbage in garbage out, so clean the data as much as possible before using it for analysis. Using of an appropriate clustering algorithm is also very important during cluster analysis. Choosing the optimal number of clusters is also a very important step during clustering. We should choose the algorithms for

finding the optimal number of cluster very carefully. Such as elbow method, gap stats method, silhouette method, etc.). Reducing dimensions could be one of the optimization method