

Lecture 9: "*Unsupervised Learning Techniques*"

Book Chapter

Please read the *Chapter 9 "Unsupervised Learning Techniques"* till page 258 ('Other Clustering Algorithm') and answer the following questions.

Keep in mind: If you answer these questions and write a detailed summary, you won't need to read these chapters again while preparing for the exam

Video Nugget

There is a nice video on [K-Means clustering](#) and also on [DBSCAN](#).

Questions

Clustering

1. What is the major benefit of *unsupervised learning*?
2. Describe the tasks *Clustering*, *Anomaly Detection* and *Density Estimation* in one sentence.
3. In which applications is *clustering* used?

K-Means

4. How can K-Means be used for dimensionality reduction?
5. Explain the difference between hard and soft clustering.
6. Describe the K-Means algorithm in your own words.
7. What can happen if we get bad random initial centroids?
8. How can we initialize the centroids better?
9. What is the idea behind K-Means++?
10. How did accelerated K-Means and mini-batch K-Means improve training time?
11. How can we measure the performance of K-Means?
12. Why is minimizing inertia a bad metric if we try to get the best number of clusters?
13. How can we find a good guess for k ?
14. How does the *silhouette score* work?
15. How can we interpret the *silhouette diagrams* on page 248?
16. What are the limitations of K-Means?

Applications of K-Means

17. What is the idea behind clustering for image segmentation?
18. How can clustering be used for semi-supervised Learning?
19. What is the idea behind label propagation?
20. What does *Active Learning* mean?

DBSCAN

21. Which instances are taken together in clusters using the DBSCAN algorithm?
22. What is a core instance in DBSCAN?
23. With which parameters is the density in DBSCAN defined?
24. List the advantages and disadvantages of the DBSCAN algorithm.

Homework Assignment

Please work on the exercises given in [Unsupervised Learning task.ipynb](#)