

1. Why does K-Means use centroids instead of medoids?
2. What is the Silhouette Score and how does it evaluate the quality of K-Means clusters?
3. Why is the Elbow Method used in K-Means, and what does the "elbow point" signify mathematically?
4. What is Within-Cluster Sum of Squares (WCSS), and why does K-Means try to minimize it?
5. Why is K-Means sensitive to outliers?
6. Why does K-Means fail when clusters are not spherical or have different densities?
7. How does the random initialization of centroids mathematically affect the final clusters?
8. Why does hierarchical clustering not require selecting the number of clusters at the beginning?
9. What is the intuition behind the distance matrix used in hierarchical clustering?
10. What is the intuition behind a dendrogram and what does the height represent?
11. Why is hierarchical clustering slower than K-Means?
12. How does the choice of linkage (single, complete, average) mathematically change cluster formation?
13. What is the core idea behind DBSCAN?
14. Why can DBSCAN find arbitrary-shaped clusters?
15. How does DBSCAN decide whether a point is core, border, or noise using ϵ and MinPts?
16. Why can DBSCAN detect noise better than K-Means?
17. What is the drawback of DBSCAN on high-dimensional data due to the curse of dimensionality?
18. What is the basic intuition behind PCA?
19. What mathematical idea does PCA use to find directions of maximum variance?

20. Why do we subtract the mean before applying PCA?
21. Why must we standardize or normalize data before PCA?
22. Why does PCA fail to capture nonlinear patterns in data?
23. What does an eigenvalue represent in PCA, and why are eigenvectors important?
24. How does PCA help in noise removal by dropping small-variance components?
25. What is the intuition behind dimensionality reduction, and why does it sometimes improve performance?
26. What is the intuition behind SVD, and how is it related to PCA?
27. How is SVD used to approximate matrices with reduced rank, and why does this reduce noise?
28. What does it mean when some singular values in SVD are zero, and how does this reveal rank deficiency?
29. What is the difference between eigen decomposition and SVD in terms of mathematical stability?
30. What is the intuition behind gradient descent in linear regression?
31. Why do we need a loss function in linear regression training?
32. How does learning rate affect the path and convergence of gradient descent?
33. Why do neural networks need activation functions, and how do they introduce non-linearity?
34. What is the intuition behind backpropagation, and how do partial derivatives update weights?
35. Why do sigmoid and tanh suffer from vanishing gradients, and how does ReLU solve this?