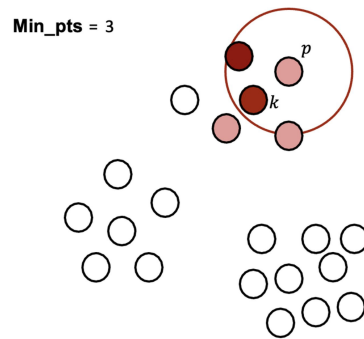# CS 506 Review Slides Questions

February 2026

Questions

1. How will point p be assigned? (Note: we are assuming that for a point to be considered a part of p's epsilon neighborhood, it must be entirely in the circle representing p's neighborhood)

   (a) p is a core point, so we will iterate on its entire epsilon neighborhood

   (b) p is a border point, so we will iterate on its entire epsilon neighborhood

   (c) p is a border point, so we will assign it to k's cluster and move onto the other points in k's neighborhood

   (d) p is a noise point, so we won't assign it to any cluster

2. Keeping epsilon the same and increasing min_pts in DBScan will likely increase the number of noise points

   (a) True

   (b) False

3. Keeping epsilon the same and increasing min_pts in DBScan will likely increase the number of core points
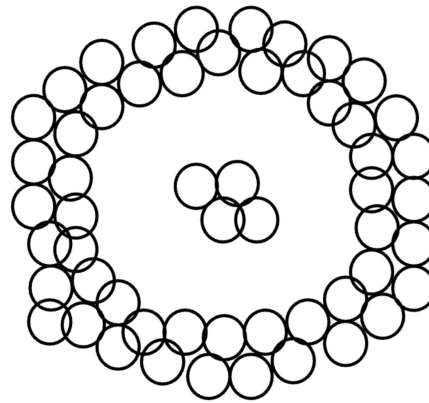
(a) True

(b) False

4. Is the following a possible output of DBScan?
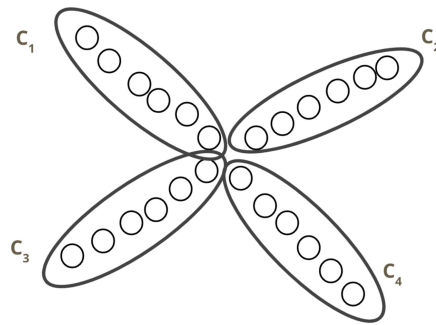
(a) Yes

(b) No

5. Is DBScan well suited for the following dataset?
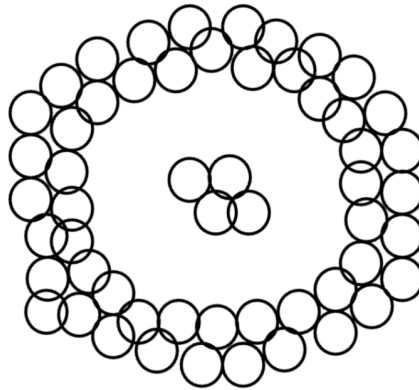

(a) Yes

(b) No

6. DBScan, like K-means, requires you to specify the number of clusters

(a) Yes

(b) No

7. Using cosine distance to calculate epsilon neighborhoods would likely enable us to detect the two natural clusters here.
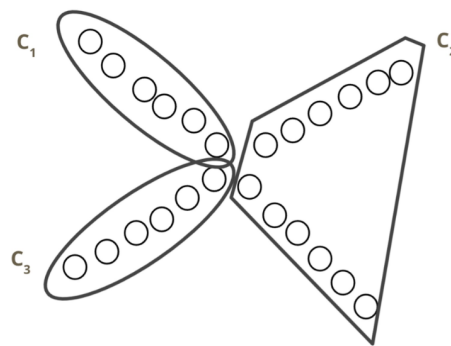
(a) True

(b) False

8. One way to initialize GMM clustering is to use K-means since it also performs well when clusters are Gaussian

(a) True

(b) False

9. In GMM, what is the likelihood of the data (with N data points) when k=N

(a) ∞

(b) $-\infty$

(c) 1

(d) 0

10. Is the following a possible output from K-means++ clustering?
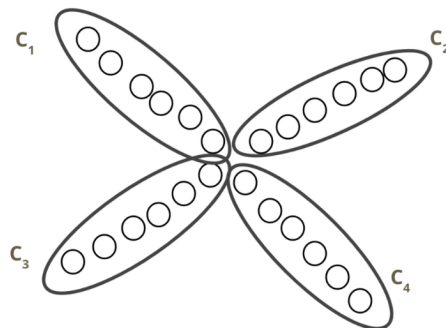
(a) Yes

(b) No

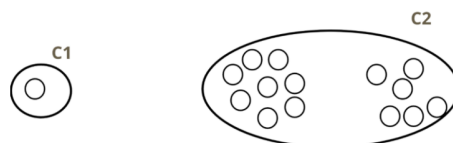11. Is the following a possible output of K-means clustering?

(a) Yes

(b) No

12. The difference between K-means and K-means++ is that K-means++ is faster, hence the ++

(a) True

(b) False

13. Consider the 1-dimensional dataset $\{4,5,x\}$. With $k = 2$ and $x > 5$, what is the smallest value of x BEYOND which x will always be in its own cluster?
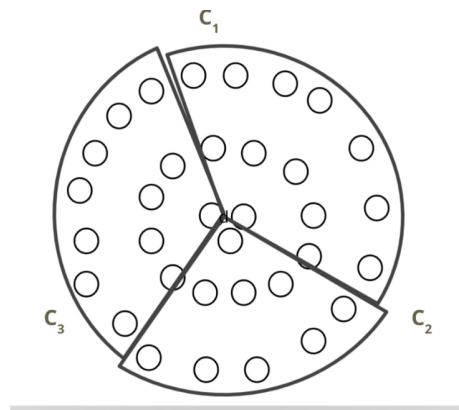
(a) 7

(b) 9

(c) 6

(d) 8

14. Is the following a possible output from k-means++ clustering?

(a) Yes

(b) No

15. Is the following a possible output of k-means clustering?

(a) Yes

(b) No

16. One way to divide clusters in Hierarchical Clustering's Divisive Algorithm (as opposed to agglomerative) is to use a partitional clustering method like K-means.
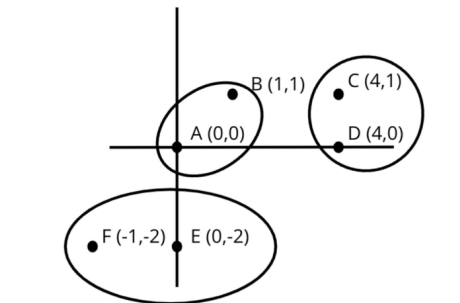
(a) True

(b) False

17. Using Euclidean and complete-link distance, what is the merging order for $A = (0, 0), B = (1, 1), C = (3, 0), D = (0, -2)$?

(a) $AB \rightarrow ABD \rightarrow ABCD$

(b) $AB \rightarrow CD \rightarrow ABCD$

(c) $AB \rightarrow ABC \rightarrow ABCD$

18. How many clusters would be created by cutting the dendrogram below as shown?
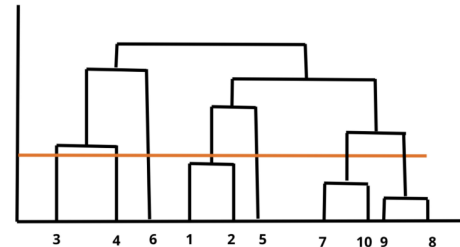
(a) 5

(b) 4

(c) 6

(d) 7

19. Which clusters get merged next if using Euclidean and complete-link distances?

(a) AB and EF

(b) AB and CD

(c) CD and EF

20. The Manhattan distance is always less than or equal to the Euclidean distance between two points in a 2D space.

(a) True

(b) False

21. In n-dimensional space the Manhattan distance equals the Euclidean distance for

    (a) Points where all the coordinates are the same except for one

    (b) Points where all the coordinates are different except for one

    (c) Points where all the coordinates are zero except for one

    (d) all of the above

22. Consider arbitrary points A and B in $\mathbb{R}^n$ s.t. $A \neq B$. For which distance function d, is $d(A, B)$ invariant to scaling? i.e., $c * d(A, B) = d(A, B)$ for any constant c

    (a) Euclidean distance

    (b) Chebyshev Distance

    (c) Cosine Distance

    (d) Manhattan Distance

23. What is the Jaccard distance between two completely different documents (i.e. no words in common at all)?

    (a) -1

    (b) 1

    (c) 0

    (d) Depends on the size of the documents

24. How many parameters does GMM need to estimate for k clusters and 1D data?

    (a) $3k$

    (b) k

    (c) $4k$

    (d) $\sqrt{k}$

25. Since outliers most likely don't belong to any cluster, the probabilities of belonging to each cluster, assigned by a GMM, will all be small

    (a) True

    (b) False

26. What type of distribution does each cluster follow in a GMM?

    (a) Normal

    (b) Geometric

    (c) Hypergeometric

    (d) Poisson

Solutions

1. (c) p is a border point, so we will assign it to k's cluster and move onto the other points in k's neighborhood

2. (a) True

3. (b) False

4. (b) No

5. (a) Yes

6. (b) No

7. (b) False

8. (a) True

9. (a) ∞

10. (a) Yes

11. (a) Yes

12. (b) False

13. (a) 7

14. (a) Yes

15. (a) Yes

16. (a) True

17. (c) $AB \rightarrow ABC \rightarrow ABCD$

18. (d) 7

19. (a) AB and EF

20. (b) False

21. (a) Points where all the coordinates are the same except for one

22. (c) Cosine Distance

23. (b) 1

24. (a)

25. False

26. (a) Normal