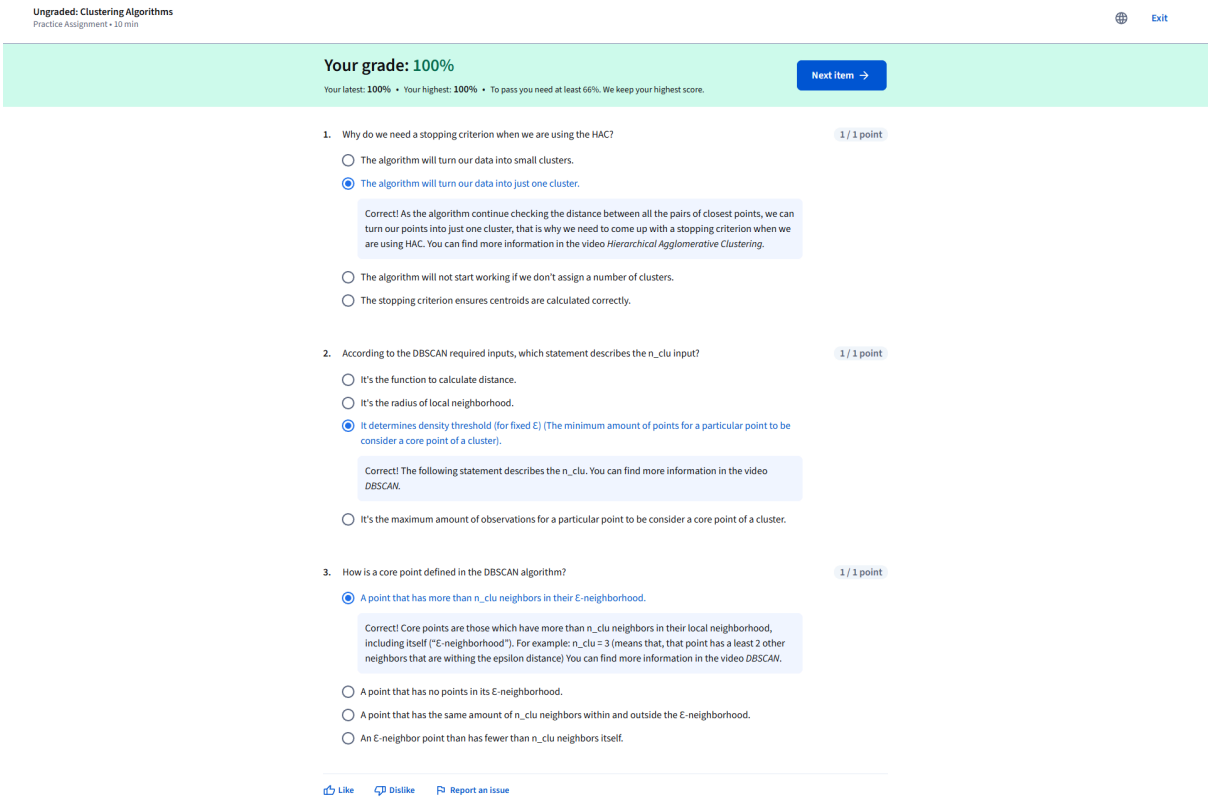
Ungraded: Clustering Algorithms



Ungraded: Clustering Algorithms
Practice Assignment • 10 min

Exit

Your grade: 100%

Your latest: **100%** • Your highest: **100%** • To pass you need at least 66%. We keep your highest score.

Next item →

1. Why do we need a stopping criterion when we are using the HAC? — 1 / 1 point

- ○ The algorithm will turn our data into small clusters.
- ◉ The algorithm will turn our data into just one cluster.

  Correct! As the algorithm continue checking the distance between all the pairs of closest points, we can turn our points into just one cluster, that is why we need to come up with a stopping criterion when we are using HAC. You can find more information in the video *Hierarchical Agglomerative Clustering*.

- ○ The algorithm will not start working if we don't assign a number of clusters.
- ○ The stopping criterion ensures centroids are calculated correctly.

2. According to the DBSCAN required inputs, which statement describes the n_clu input? — 1 / 1 point

- ○ It's the function to calculate distance.
- ○ It's the radius of local neighborhood.
- ◉ It determines density threshold (for fixed Ɛ) (The minimum amount of points for a particular point to be consider a core point of a cluster).

  Correct! The following statement describes the n_clu. You can find more information in the video *DBSCAN*.

- ○ It's the maximum amount of observations for a particular point to be consider a core point of a cluster.

3. How is a core point defined in the DBSCAN algorithm? — 1 / 1 point

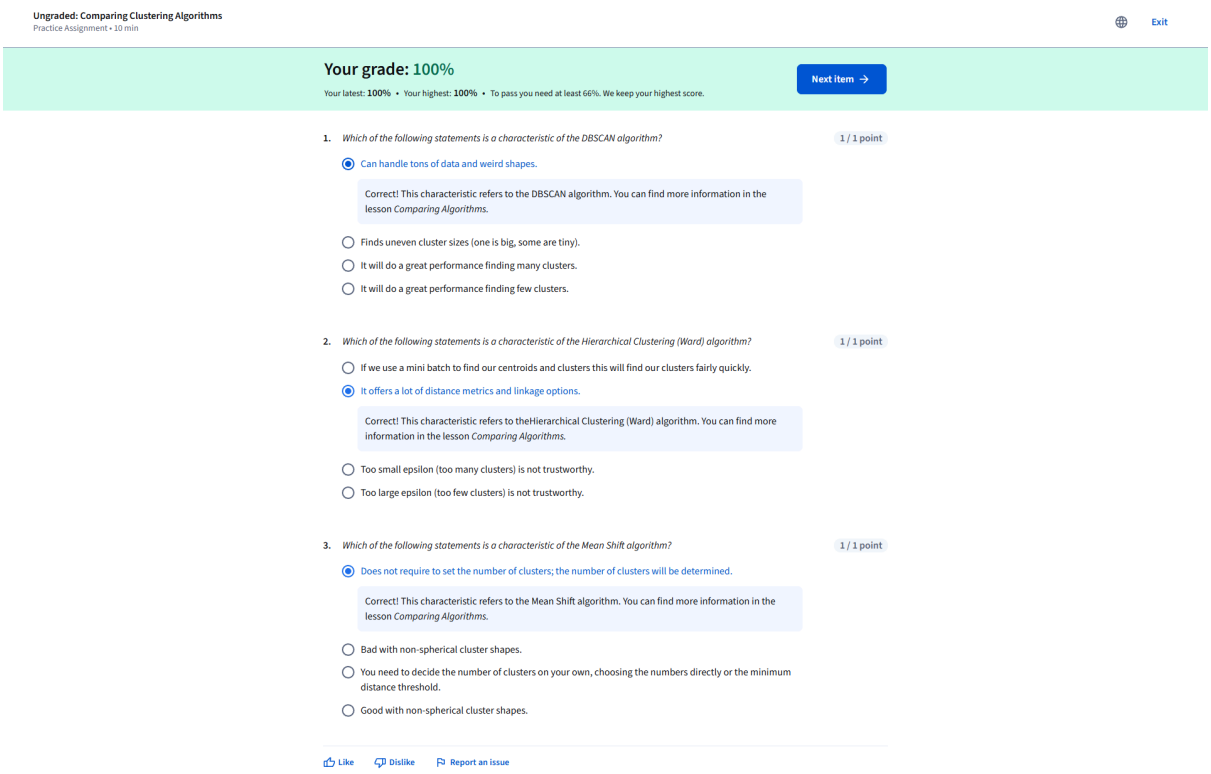
- ◉ A point that has more than n_clu neighbors in their Ɛ-neighborhood.

  Correct! Core points are those which have more than n_clu neighbors in their local neighborhood, including itself ("Ɛ-neighborhood"). For example: n_clu = 3 (means that, that point has a least 2 other neighbors that are withing the epsilon distance) You can find more information in the video *DBSCAN*.

- ○ A point that has no points in its Ɛ-neighborhood.
- ○ A point that has the same amount of n_clu neighbors within and outside the Ɛ-neighborhood.
- ○ An Ɛ-neighbor point than has fewer than n_clu neighbors itself.

Like Dislike Report an issue

Ungraded: Comparing Clustering Algorithms



Graded: Module 3 Quiz

Your grade: 100%

Your latest: **100%** • Your highest: **100%** • To pass you need at least 80%. We keep your highest score.

Next item →

1. When using DBSCAN, how does the algorithm determine that a cluster is complete and is time to move to a different point of the data set and potentially start a new cluster? 1 / 1 point

- ○ When the algorithm requires you to change the input.
- ○ When the algorithm forms a new cluster using the outliers.
- ◉ When no point is left unvisited by the chain reaction.
- ○ When the solution converges to a single cluster.

Correct
Correct! We keep going until we find the entire cluster, and no point is left unvisited by this chain reaction. If we have no neighbors left, randomly try a new unvisited point to potentially start a new cluster. You can find more information in the lesson *DBSCAN Part 2*.

2. Which of the following statements correctly defines the strengths of the DBSCAN algorithm? 1 / 1 point

- ◉ No need to specify the number of clusters (cf. K-means), allows for noise, and can handle arbitrary-shaped clusters.
- ○ Do well with different density, works with just one parameter, the n_clu defines itself.
- ○ The algorithm will find the outliers first, draw regular shapes, works faster than other algorithms.
- ○ The algorithm is computationally intensive, it is sensitive to outliers, and it requires few hyperparameters to be tuned.

Correct
Correct! These 3 characteristics describe the strengths of the algorithm. You can find more information in the lesson *DBSCAN Part 2*.

3. Which of the following statements correctly defines the weaknesses of the DBSCAN algorithm? 1 / 1 point

- ○ The clusters it find might not be trustworthy, it needs noisy data to work, and it can't handle subgroups.
- ◉ It needs two parameters as input, finding appropriate values of Ɛ and n_clu can be difficult, and it does not do well with clusters of different density.
- ○ The algorithm will find the outliers first, it draws regular shapes, and it works faster than other algorithms.
- ○ The algorithm is computationally intensive, it is sensitive to outliers, and it requires too many hyperparameters to be tuned.

Correct
Correct! These 3 characteristics describe the weaknesses of the algorithm. You can find more information in the lesson *DBSCAN Part 2*.

4. (True/false) Does complete linkage refers to the maximum pairwise distance between clusters? 1 / 1 point

- ◉ True
- ○ False

Correct
Correct! By using the complete linkage measuring method we take the maximum distance value to decide which one is the smallest and then we can boost the hierarchy. You can find more information in the lesson *Hierarchical Agglomerative Clustering Part 2*.

5. Which of the following measure methods computes the inertia and pick the pair that is going to ultimately minimize the inertia value? 1 / 1 point

- ○ Single linkage
- ○ Average linkage
- ◉ Ward linkage
- ○ Complete linkage

Correct
Correct! The merge of this measure method is based on inertia. You can find more information in the lesson *Hierarchical Agglomerative Clustering Part 2*.