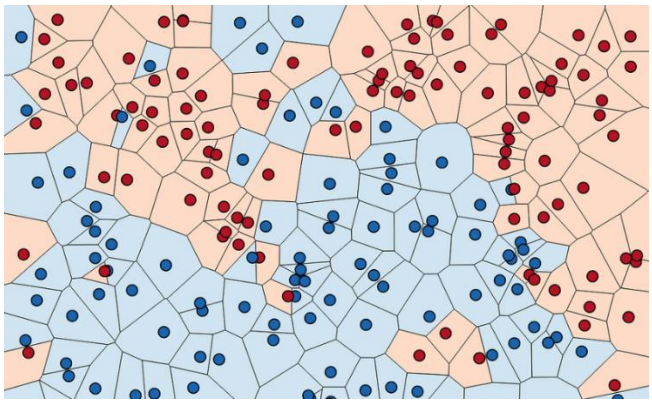
Week 6 Kahoot

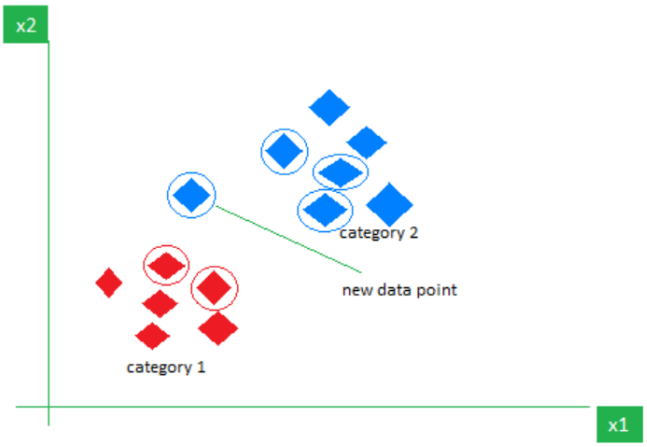
Q1



How do we determine K for K Nearest Neighbor Classifier?

- It is the value that minimizes the cost function
- **It is a hyperparameter that we tune using cross validation**
- Using elbow method in the SSE plot
- Using domain knowledge

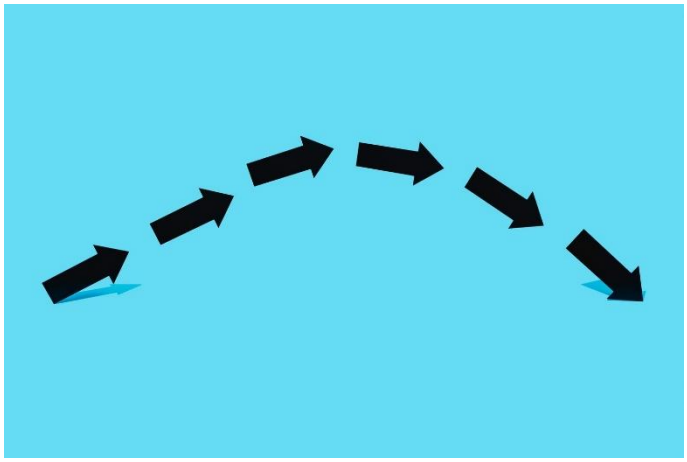
Q2



K Nearest neighbor model for K=1 has high variance.

- **True**
- False

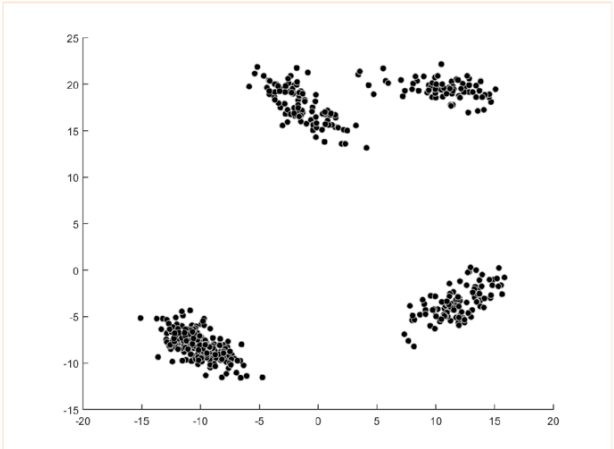
Q3



When choosing the value of K for KMeans, we should choose K that among other conditions also:

- minimizes the inertia
- gives high value of Davies Bouldin index
- **gives high value of Silhouette coefficient**
- minimizes the validation error

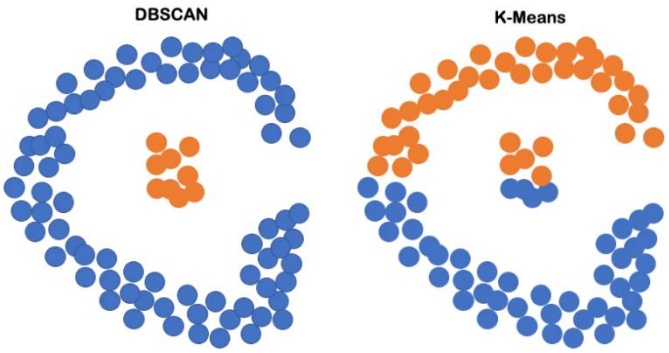
Q4



We use KMeans ++ in the context of KMeans clustering in order to

- **increase the chances of finding globally optimal assignment of points**
- ensure that all points belong to some cluster
- find the optimal number of clusters
- ensure that we find globally optimal assignment of points

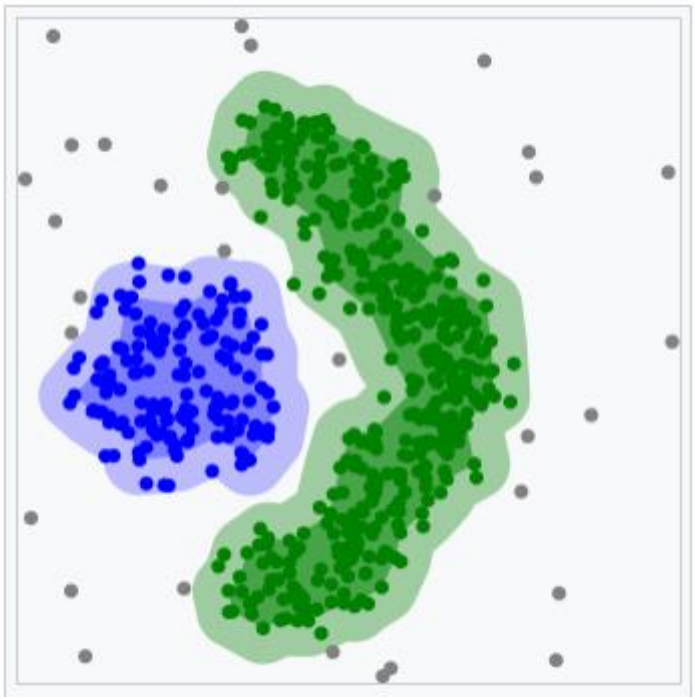
Q5



Unlike KMeans, DBSCAN clustering has no hyperparameters.

- True
- **False**

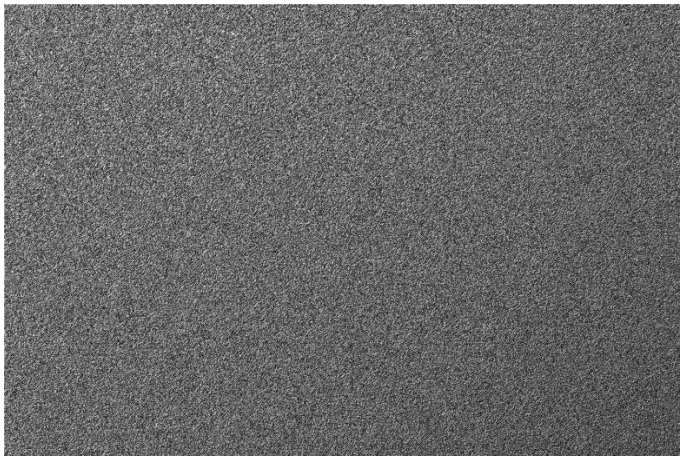
Q6



How does sklearn DBSCAN treat points that are not sufficiently dense to belong to a cluster?

- It assigns them to the nearest cluster.
- It increases the cluster size to include them.
- It creates new clusters for each of such points.
- **It treats them as noise, labeling them as cluster -1**

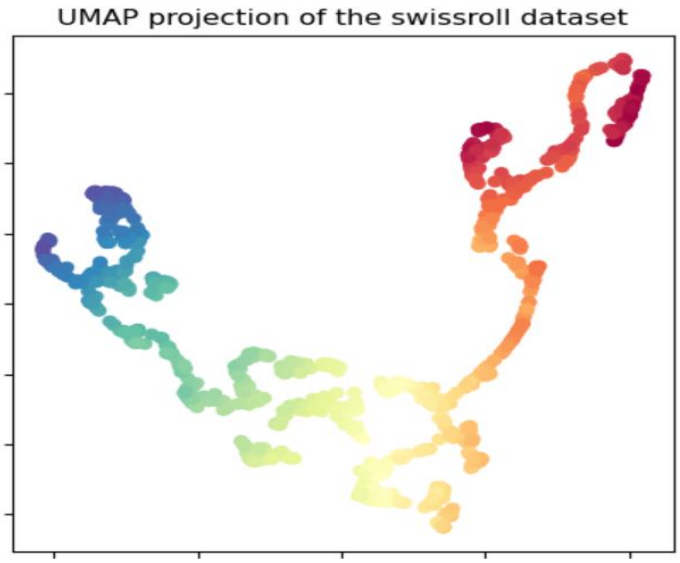
Q7



We performed DBSCAN and saw that all the points were labeled as noise. Next, we should try to:

- **increase epsilon**
- increase min_sample (min number of neighbors)
- specify a larger number of clusters
- run again with different initialization

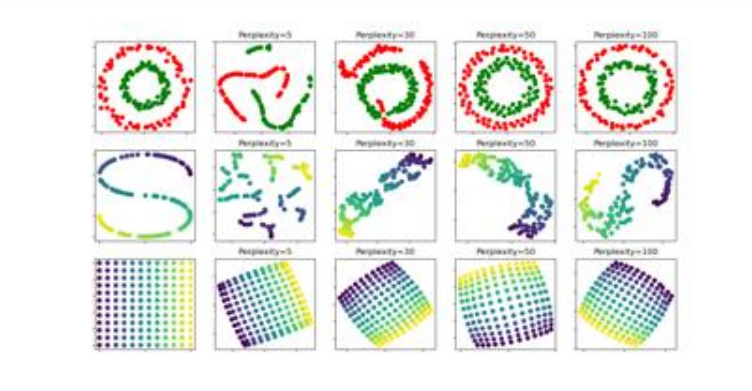
Q8



In 2-dimensional UMAP visualization, what does each point represent?

- original datapoint plotted using 2 first features
- original data point plotted using 2 most important features
- **Original data point projected in 2d space**
- The distance between 2 points in the original space

Q9



How do we obtain lower dimensional representation of the data using sklearn TSNE?

- **fit_transform(X)**
- transform(X)
- fit_transform(X,y)
- transform(X,y)

Q10

What I think about this course after half of the semester:

- I should have dropped the course, it takes too much time
- I thought I was going to learn more AI stuff
- I thought we were going to focus less on code
- **I am learning some interesting things**