DICE
ANALYTICS

DATA SCIENCE & MACHINE LEARNING COURSE

https://www.facebook.com/diceanalytics/
https://pk.linkedin.com/company/diceanalytics

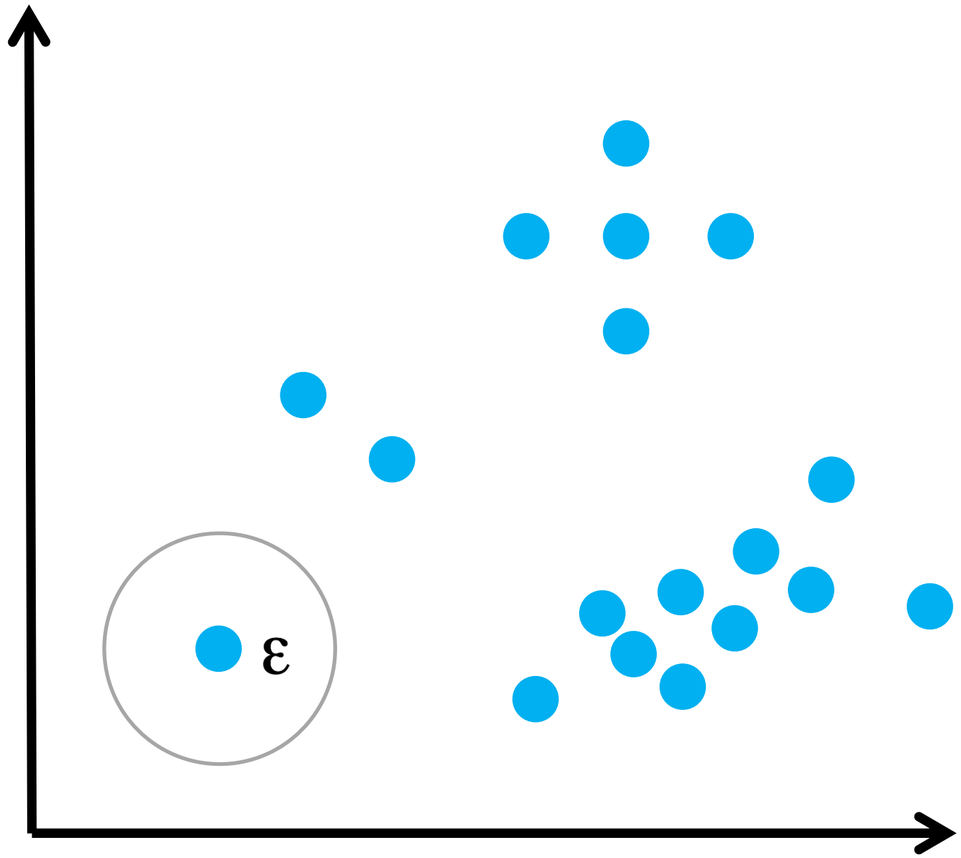
DBSCAN – Density Based Clustering

Density-Based Spatial Clustering of Applications with Noise

epsilon: 1.0
MinPoints: 4

Source: Udacity Machine Learning Nano-Degree

DBSCAN – Density Based Clustering



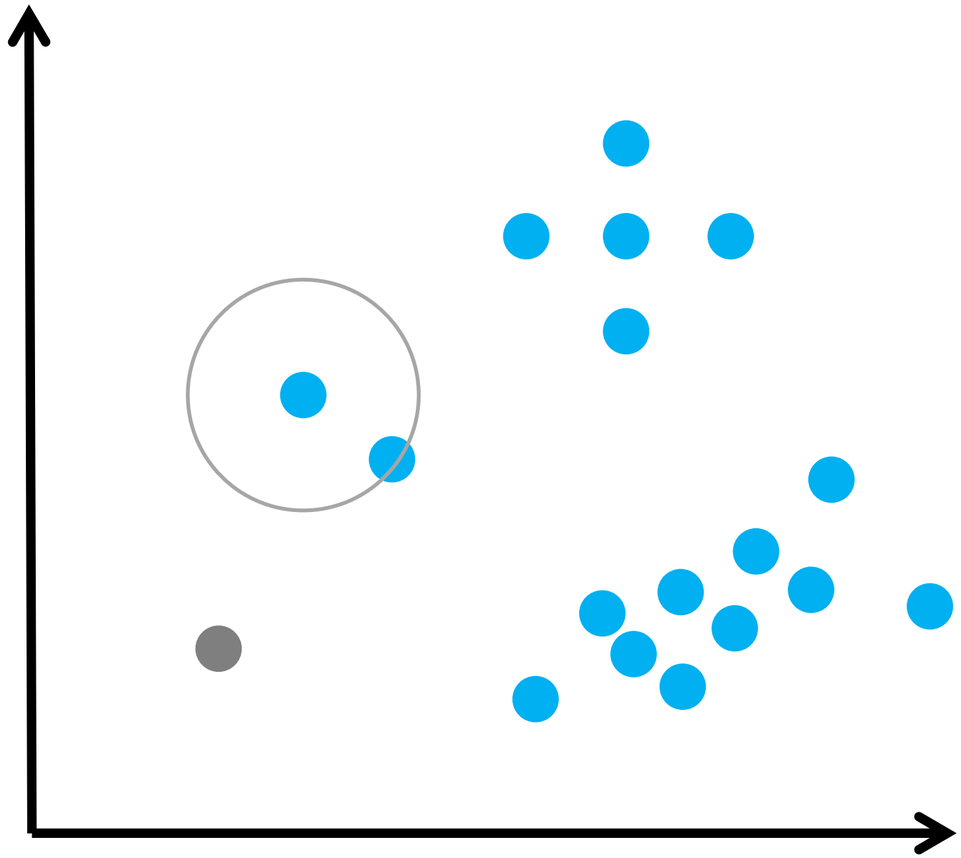
Inputs

Epsilon = 1.0

ε

Source: Udacity Machine Learning Nano-Degree

DBSCAN – Density Based Clustering



Inputs

ε

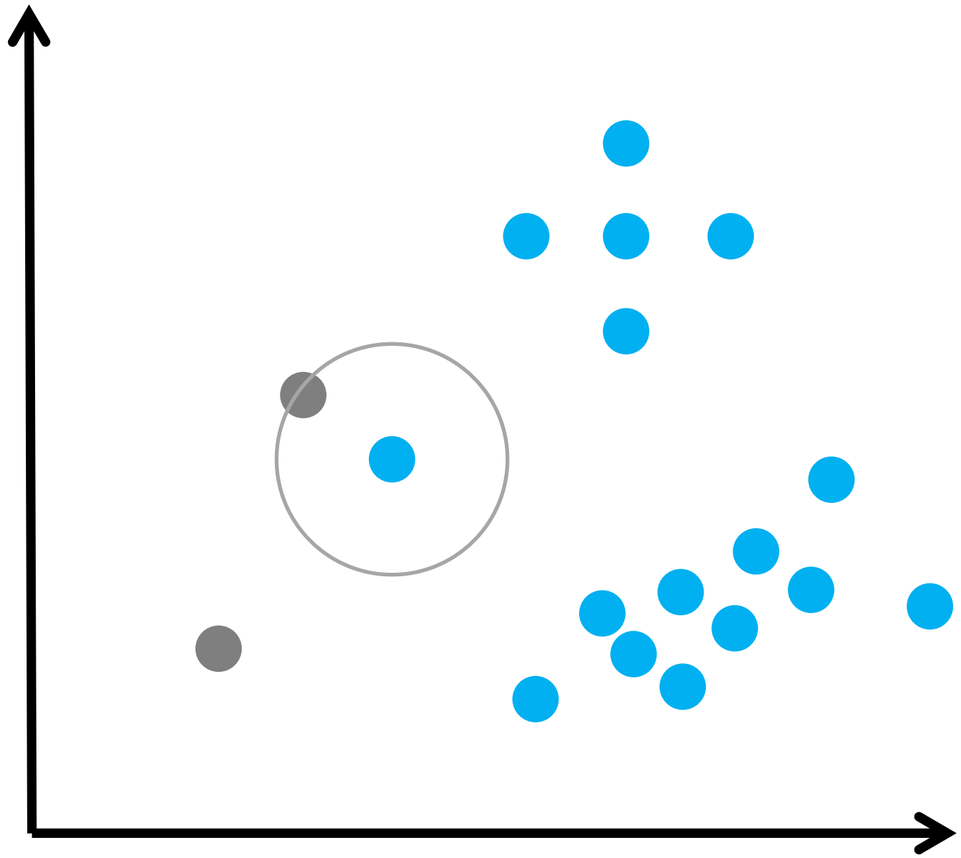
Epsilon = 1.0
Search distance around points

Min Points = 5
Minimum points required to form a density cluster

Noise Point

Source: Udacity Machine Learning Nano-Degree

DBSCAN – Density Based Clustering



Inputs

ε

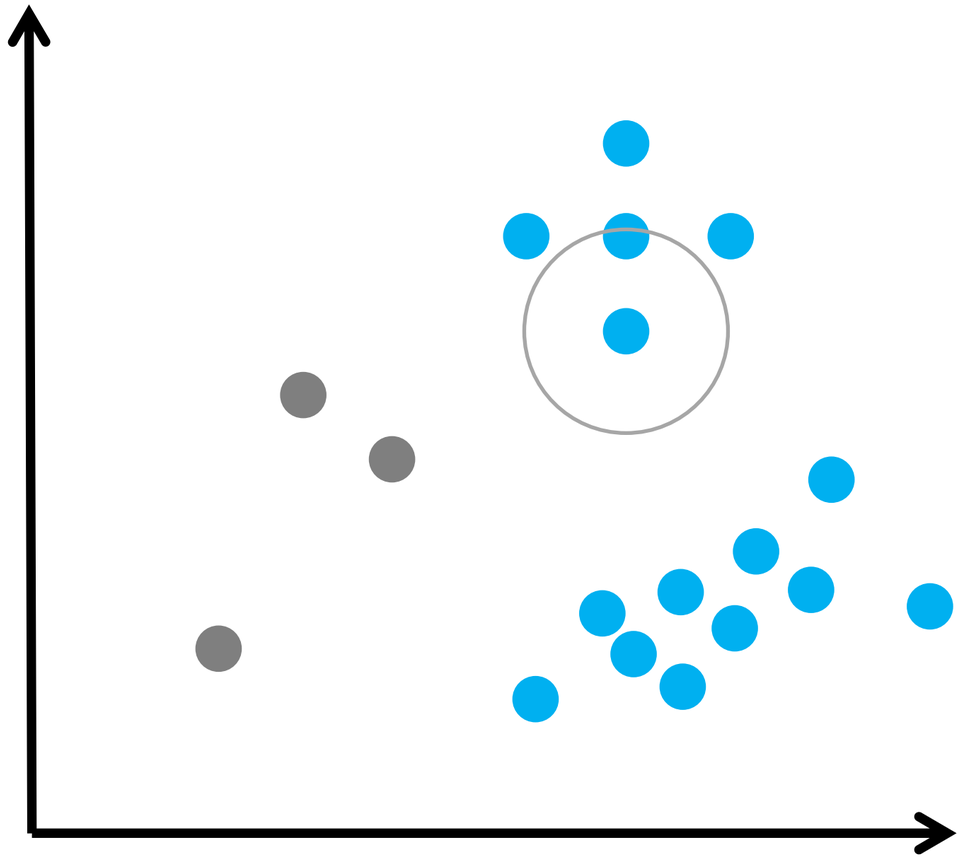
Epsilon = 1.0
Search distance around points

Min Points = 5
Minimum points required to form a density cluster

Noise Point

Source: Udacity Machine Learning Nano-Degree

DBSCAN – Density Based Clustering



Inputs

ε

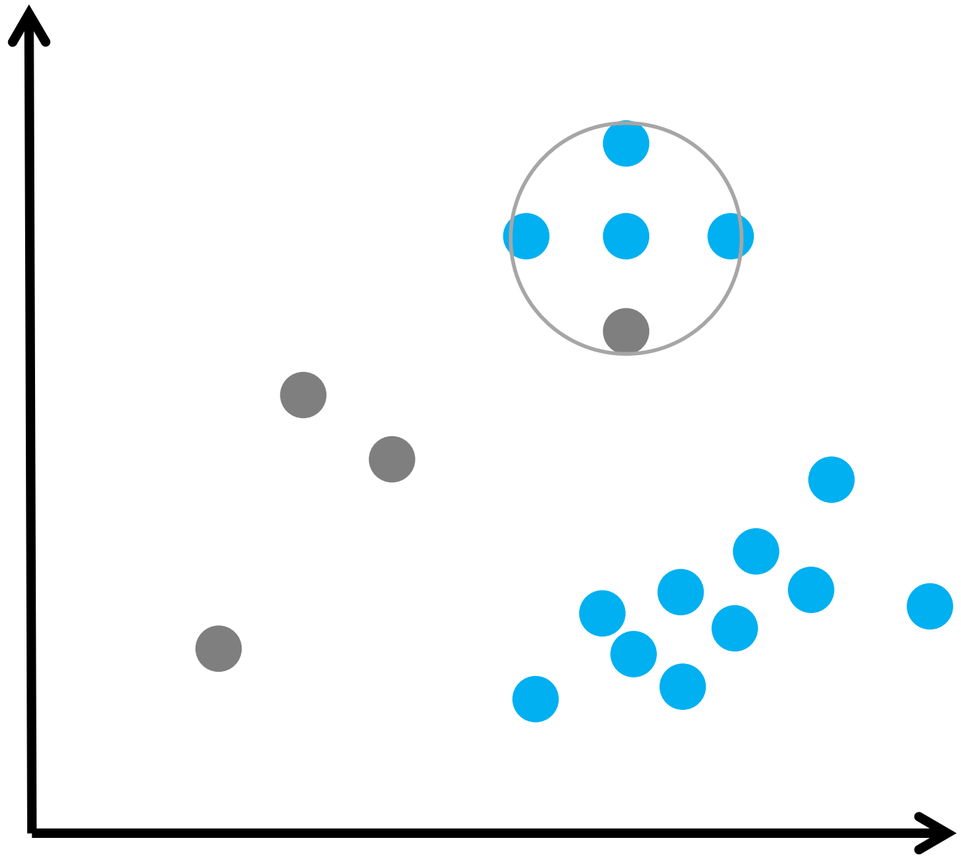
Epsilon = 1.0
Search distance around points

Min Points = 5
Minimum points required to form a density cluster

Noise Point

Source: Udacity Machine Learning Nano-Degree

DBSCAN – Density Based Clustering



Inputs

ε

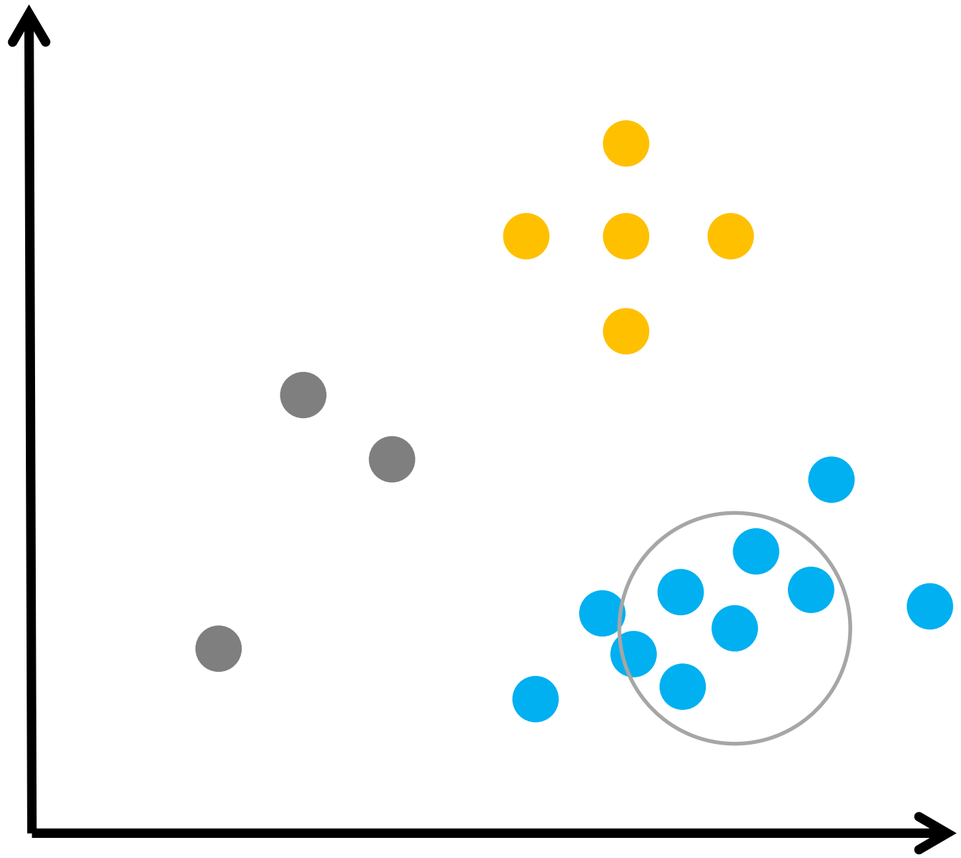
Epsilon = 1.0
Search distance around points

Min Points = 5
Minimum points required to form a density cluster

Noise Point

Source: Udacity Machine Learning Nano-Degree

DBSCAN – Density Based Clustering



Inputs

ε

Epsilon = 1.0
Search distance around points

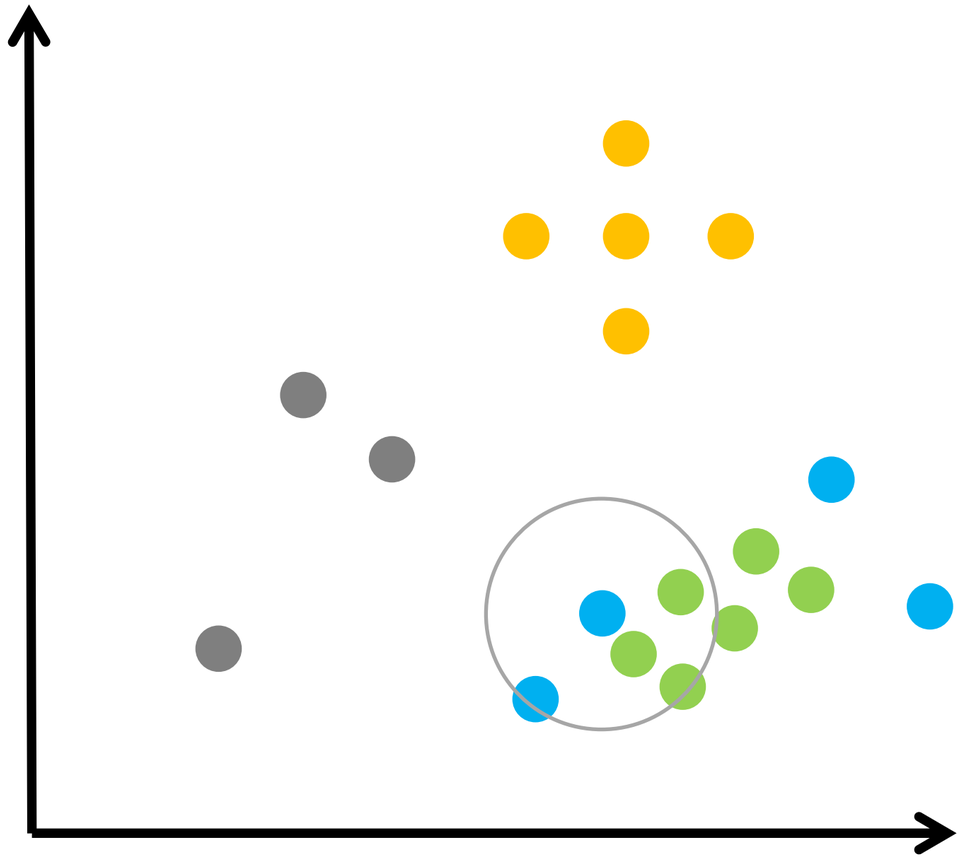
Min Points = 5
Minimum points required to form a density cluster

Cluster-1

Noise Point

Source: Udacity Machine Learning Nano-Degree

DBSCAN – Density Based Clustering



Inputs

ε

Epsilon = 1.0
Search distance around points

Min Points = 5
Minimum points required to form a density cluster

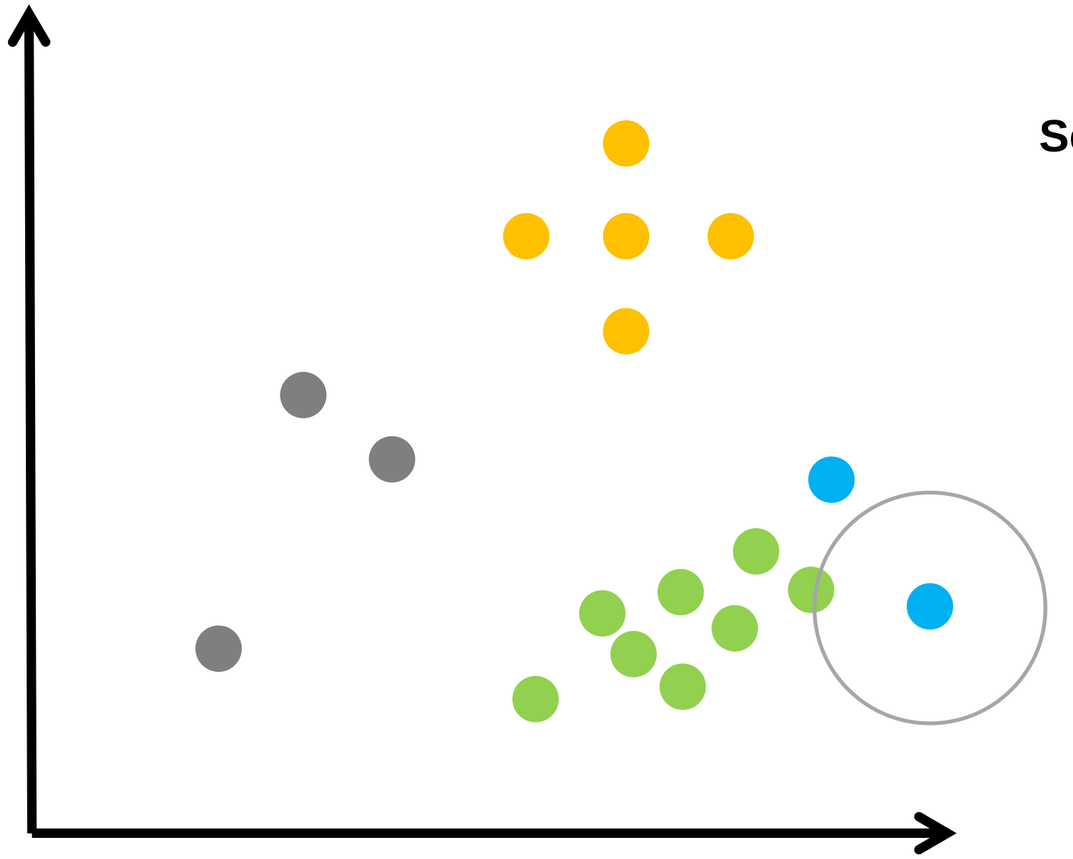
Cluster-2

Cluster-1

Noise Point

Source: Udacity Machine Learning Nano-Degree

DBSCAN – Density Based Clustering



Inputs

ε

Epsilon = 1.0
Search distance around points

Min Points = 5
Minimum points required to form a density cluster

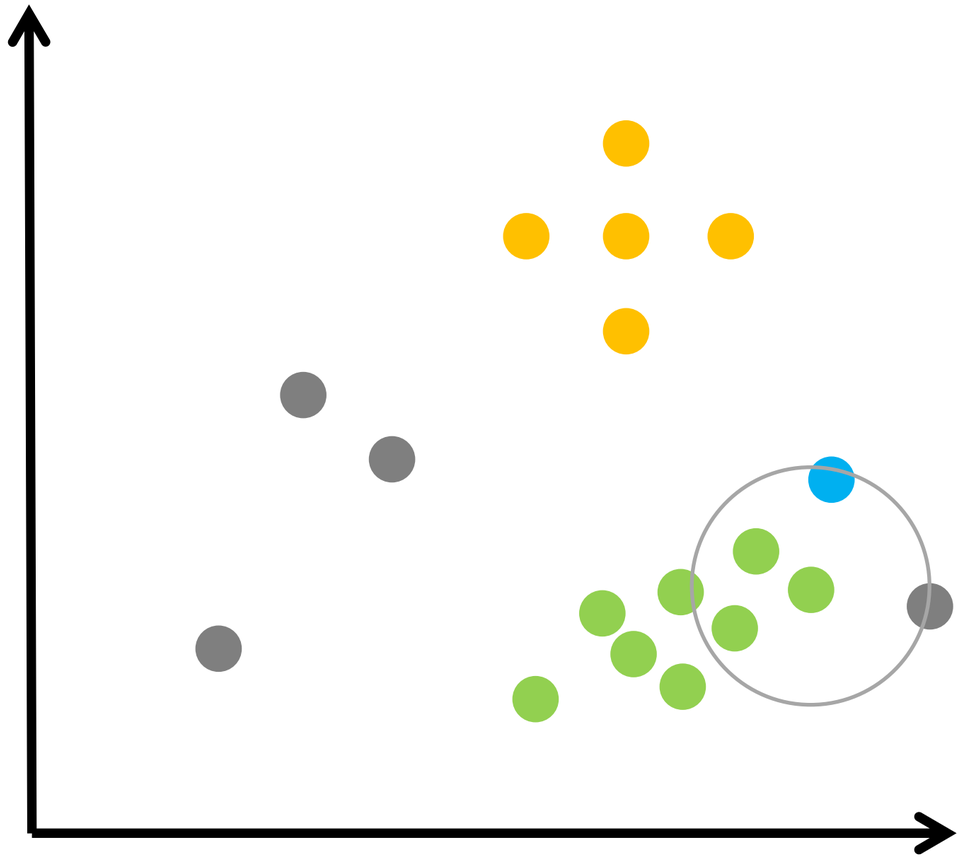
Cluster-2

Cluster-1

Noise Point

Source: Udacity Machine Learning Nano-Degree

DBSCAN – Density Based Clustering



Inputs

ε

Epsilon = 1.0
Search distance around points

Min Points = 5
Minimum points required to form a density cluster

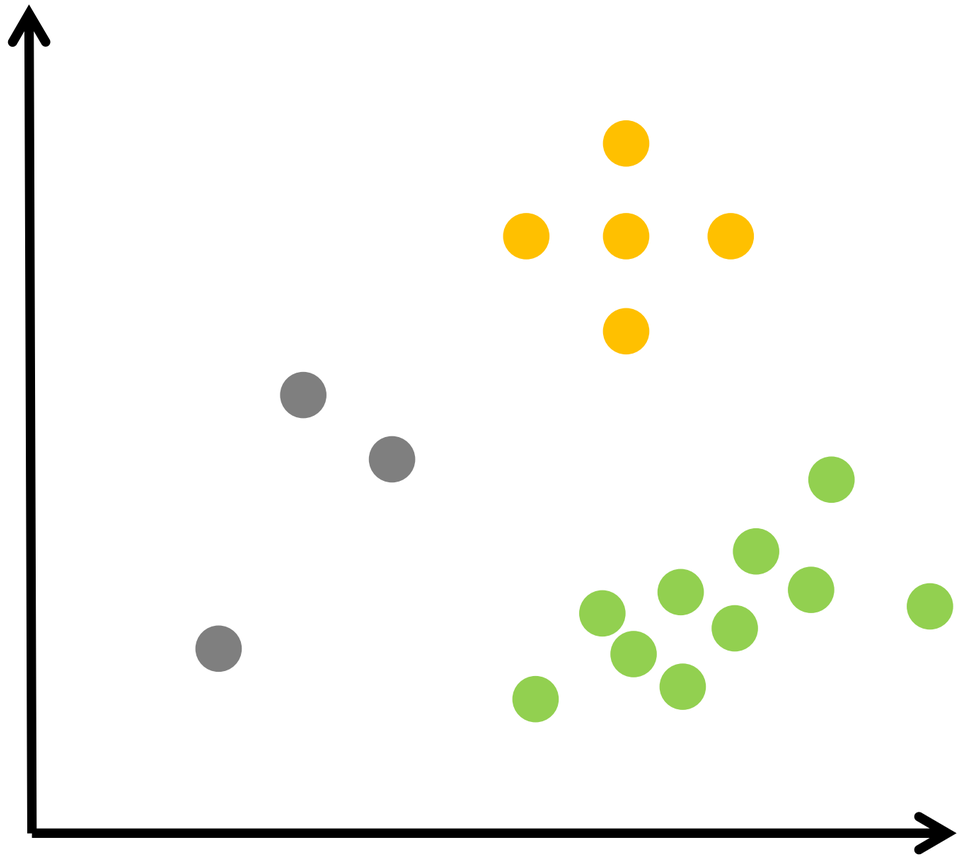
Cluster-2

Cluster-1

Noise Point

Source: Udacity Machine Learning Nano-Degree

DBSCAN – Density Based Clustering



Inputs

ε

Epsilon = 1.0
Search distance around points

Min Points = 5
Minimum points required to form a density cluster

Cluster-2

Cluster-1

Noise Point

Source: Udacity Machine Learning Nano-Degree

DBSCAN Implementation

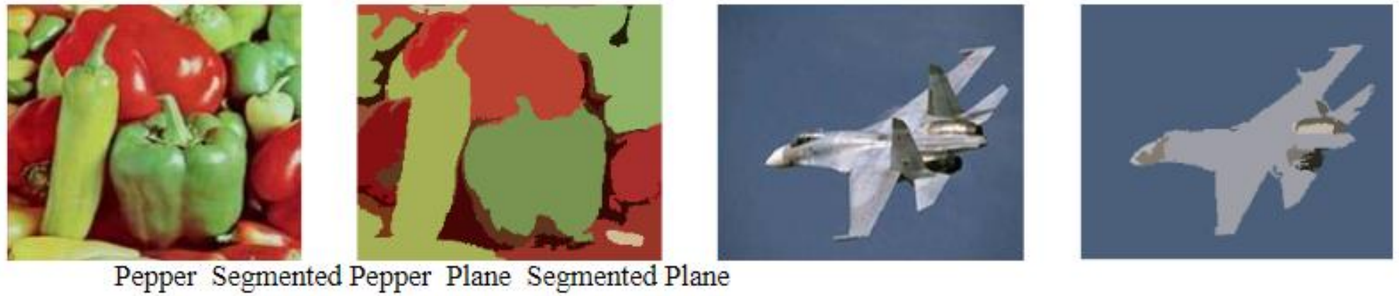
class `sklearn.cluster.` **DBSCAN** (*eps=0.5, min_samples=5, metric='euclidean', metric_params=None, algorithm='auto', leaf_size=30, p=None, n_jobs=None*) [source]

```
>>> from sklearn.cluster import DBSCAN
>>> import numpy as np
>>> X = np.array([[1, 2], [2, 2], [2, 3],
...               [8, 7], [8, 8], [25, 80]])
>>> clustering = DBSCAN(eps=3, min_samples=2).fit(X)
>>> clustering.labels_
array([ 0,  0,  0,  1,  1, -1])
>>> clustering
DBSCAN(algorithm='auto', eps=3, leaf_size=30, metric='euclidean',
    metric_params=None, min_samples=2, n_jobs=None, p=None)
```

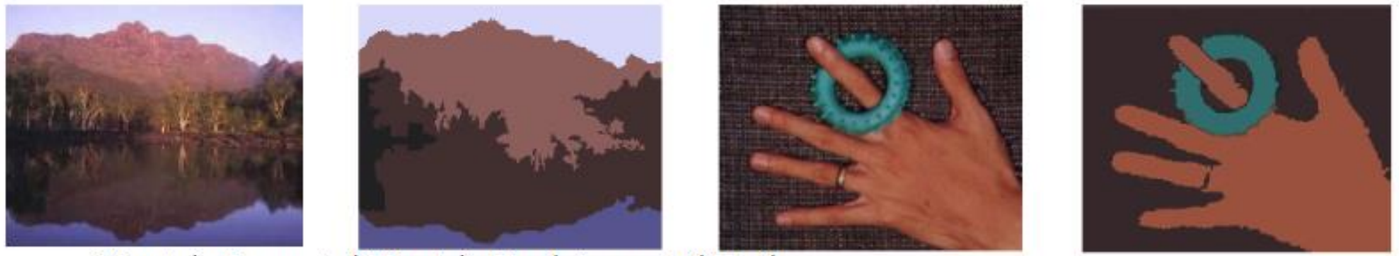
https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html

DBSCAN Applications

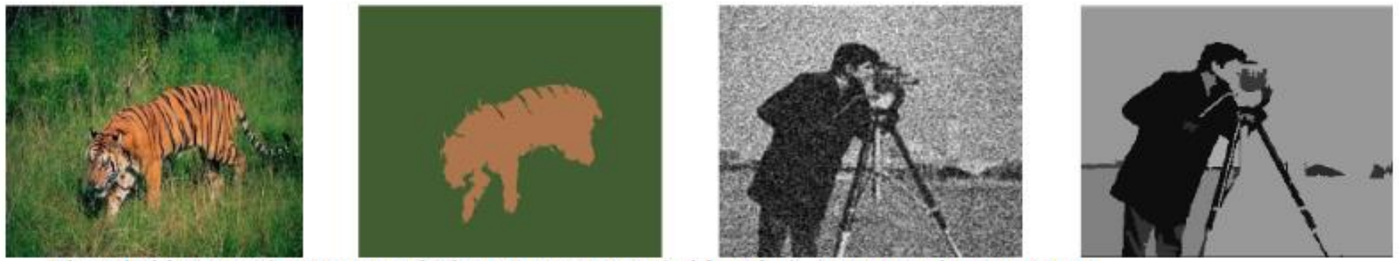
Color image segmentation using density-based clustering



Pepper Segmented Pepper Plane Segmented Plane



Mountain Segmented Mountain Hand Segmented Hand



Tiger (with texture) Segmented Tiger Cameraman (with noise) Segmented Cameraman

https://www.researchgate.net/publication/4028066_Color_image_segmentation_using_density-based_clustering

DBSCAN Applications

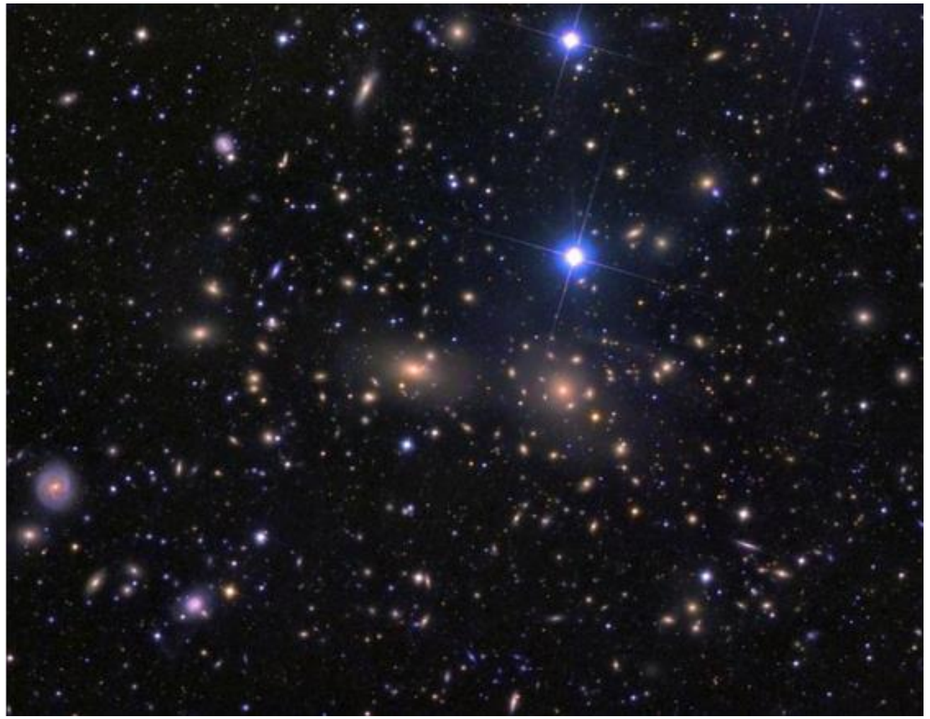
Density Based Clustering to Oil Spill Detection on Satellite Images

https://shodhganga.inflibnet.ac.in/bitstream/10603/25515/11/11_chapter%205.pdf

DICE ANALYTICS

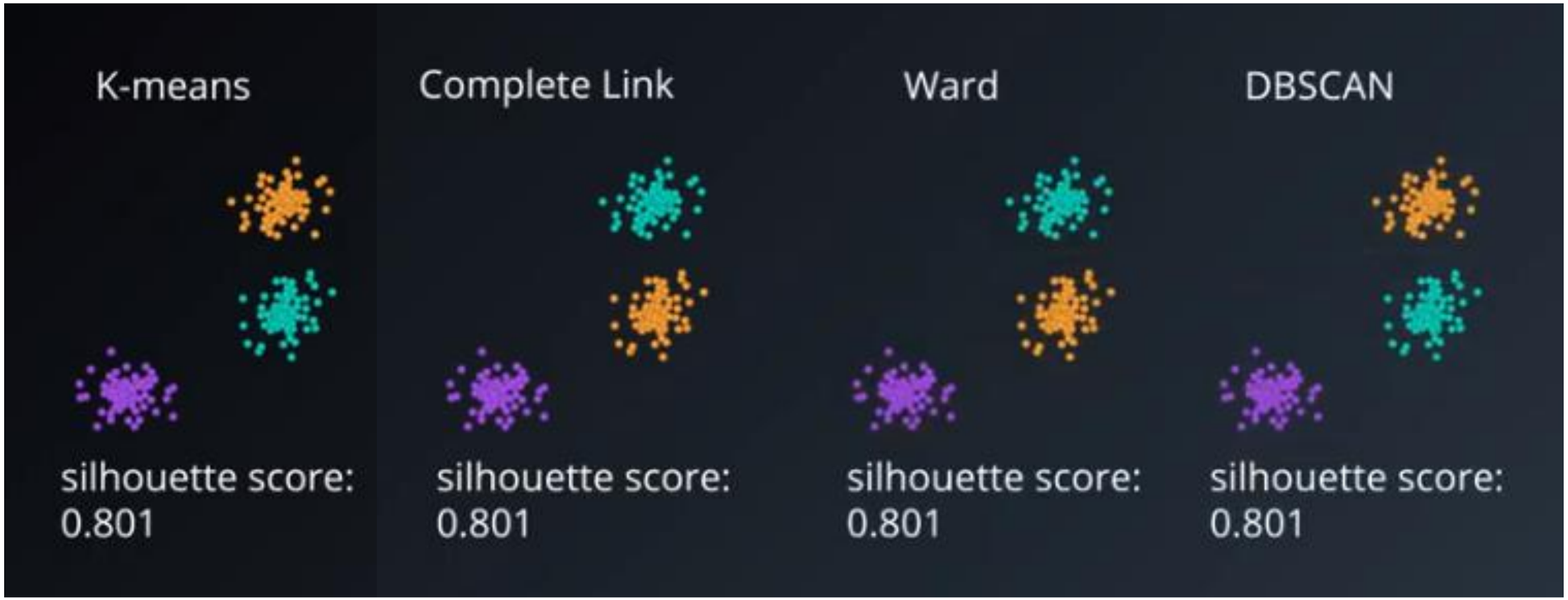
DBSCAN Applications

Evolution of Star Formation of Dwarf Galaxies within Extragalactic Cluster Substructures



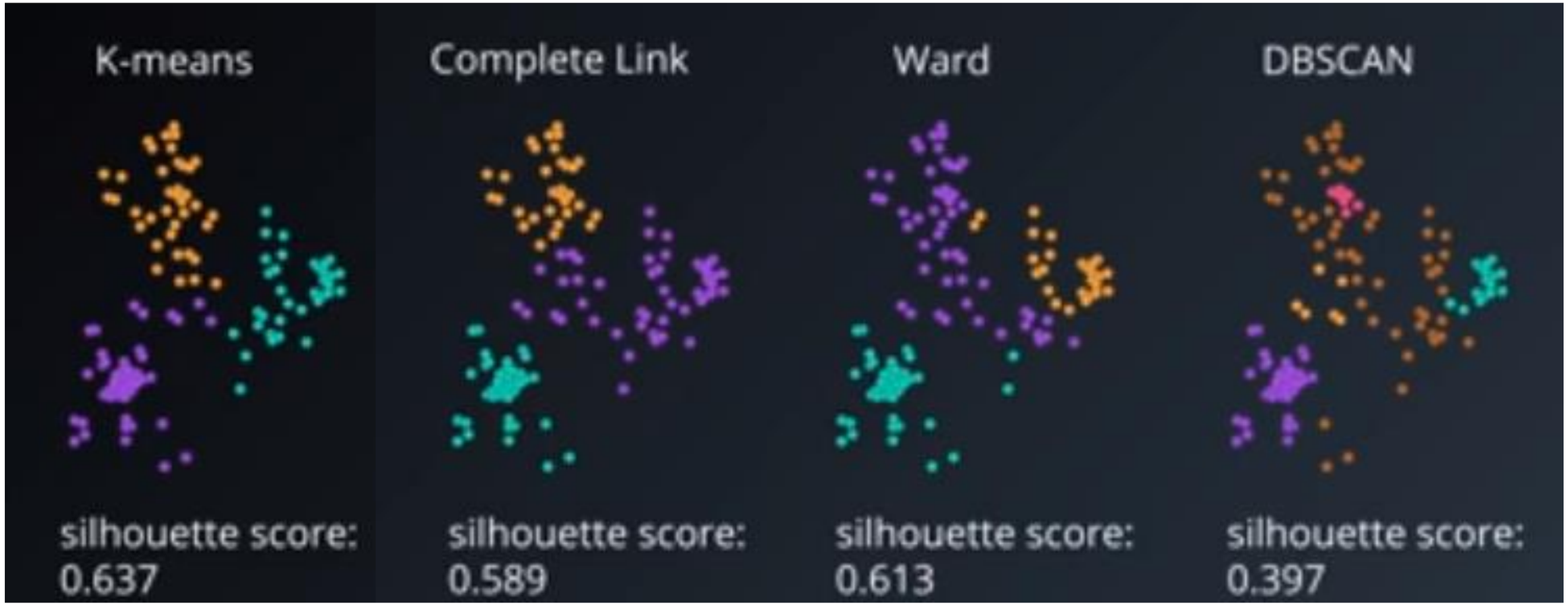
https://www.haystack.mit.edu/edu/reu/2016/files/2016_Archer_Presentation.pdf

Comparing Clustering Algos



Source: Udacity Machine Learning Nano-Degree

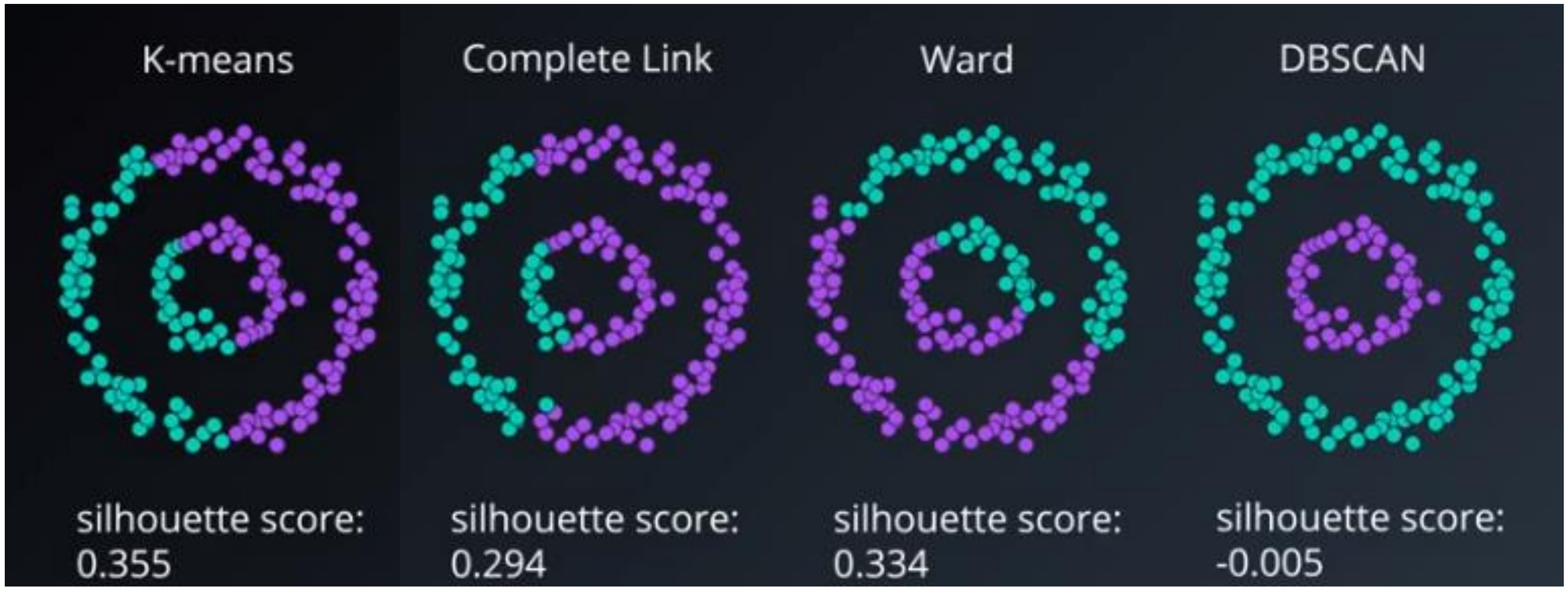
Comparing Clustering Algos



Source: Udacity Machine Learning Nano-Degree

Comparing Clustering Algos

DBCV

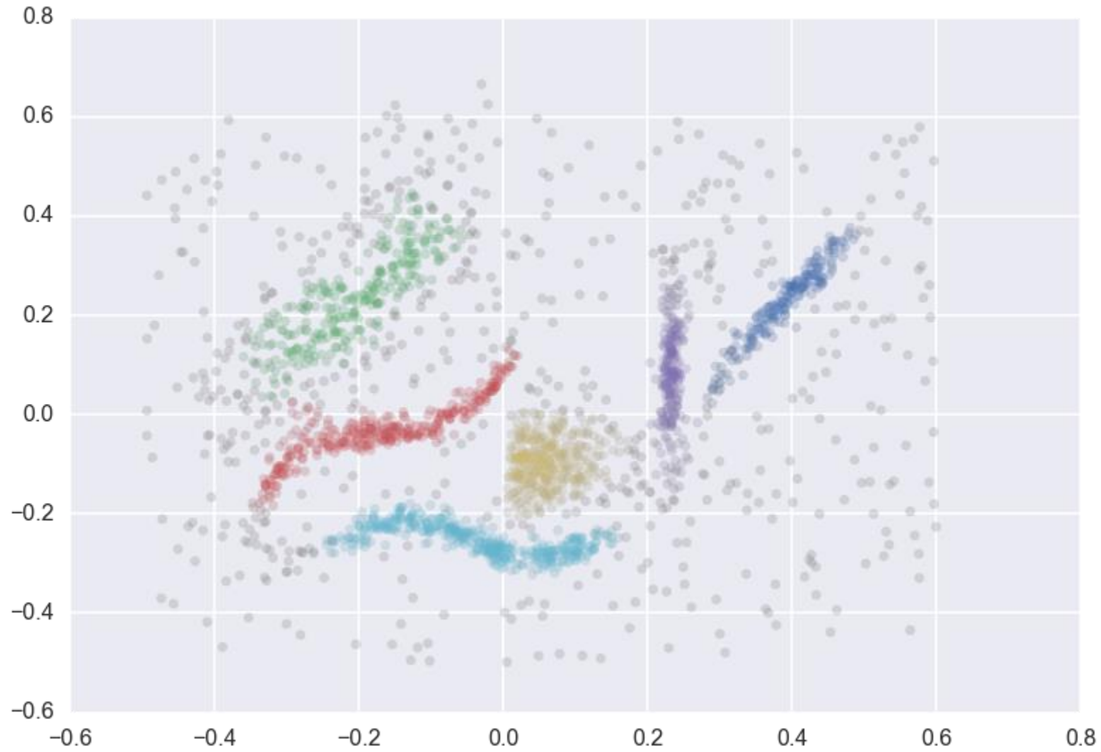


Source: Udacity Machine Learning Nano-Degree

DBCV – Density Based Cluster Validation

DBCV can validate clustering assignments on non-globular, arbitrarily shaped clusters. In essence, DBCV computes two values:

- The density **within** a cluster
- The density **between** clusters



https://epubs.siam.org/doi/pdf/10.1137/1.9781611973440.96