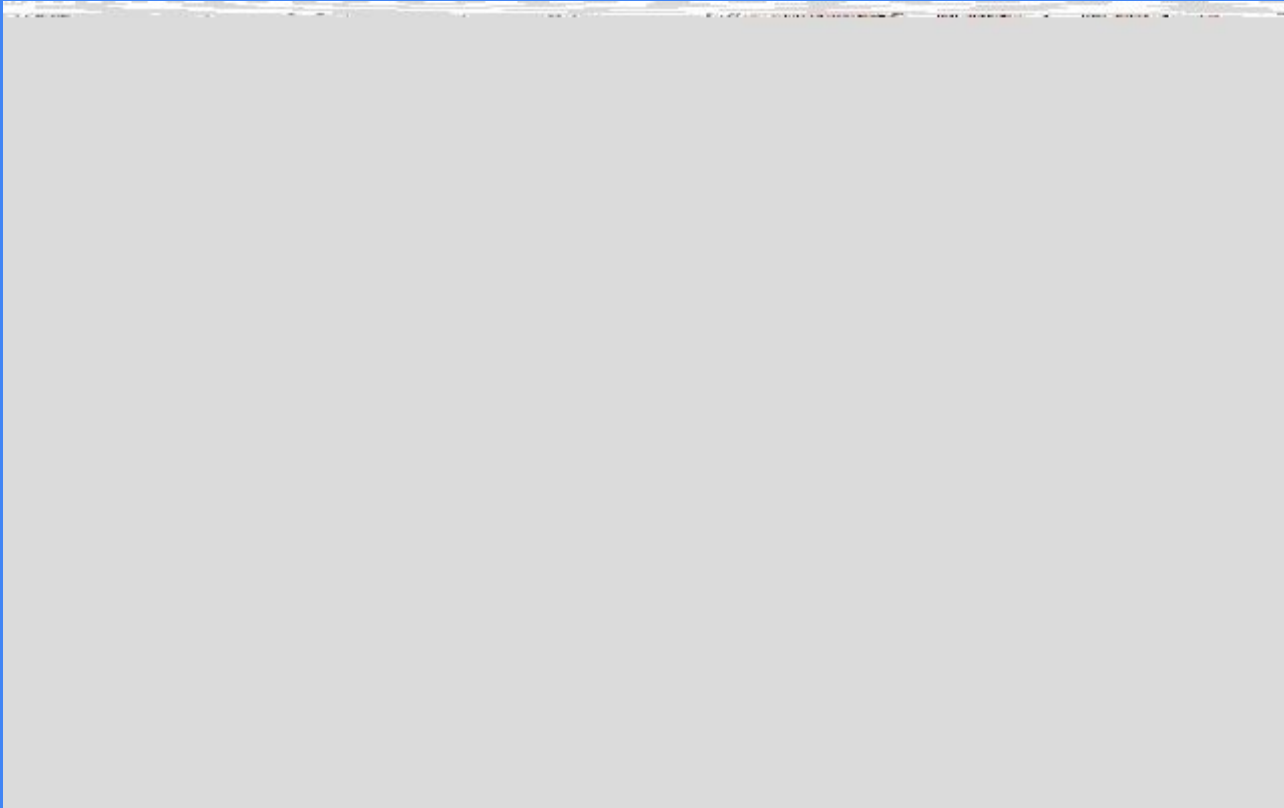


DBSCAN

Charlie Rehder, Luis Rivera, Parth Patel, Marjea Mckoy



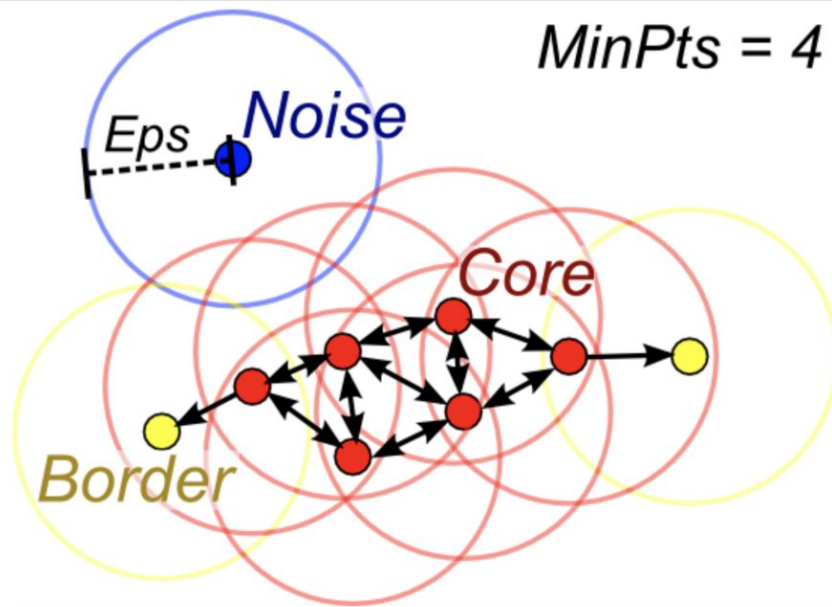
Advantages and Disadvantages

Advantages	Disadvantages
<ul style="list-style-type: none">- Does a great job separating clusters of high and low density	<ul style="list-style-type: none">- It struggles with clusters of similar density.
<ul style="list-style-type: none">- Identifies outliers and noise while clustering	<ul style="list-style-type: none">- Suffers with high dimensionality
<ul style="list-style-type: none">- Does not require a specification on the number of clusters	<ul style="list-style-type: none">- Very sensitive with EPS and minimum points that you set (this heavily influences clustering)
<ul style="list-style-type: none">- Can work with nested wrapping and arbitrarily shaped clusters	<ul style="list-style-type: none">- It fails in identifying clusters if the density varies and if the dataset is too sparse

Hyperparameters

- **Epsilon (ϵ)** - radius for all points
- **Minimum Points** - minimum number of data points to define a cluster

DBSCAN is highly sensitive to the values of these parameters.



Data Processing Steps

- Generally, when clustering, standardization is helpful, but there are times when it is not.
- Missing values must be removed or imputed; DBSCAN cannot handle missing values.
- DBSCAN is robust against outliers and noise; not necessary to handle these in preprocessing.
- All data must be numeric; however, avoid dummy variables as adding extra, sparse features can lead to poor performance.
- Suffers from the [Curse of Dimensionality](#).