

Quiz 6

1. (1 Pt.) State **three** assumptions of k -means clustering.

Disjoint/separate clusters

Spherical clusters

Roughly similarly sized clusters

2. (1 Pt.) State **three** data preparation steps that should be considered before conducting a clustering analysis.

Standardization, removing/replacing outliers, binning of rare categorical input levels, appropriate numeric encoding of categorical inputs, combining or removing highly correlated features

3. (1 Pt.) **True or False:** Squared error from cluster centroids nearly always decreases when adding more clusters into an analysis.

True

4. (1 Pt.) State **four** mathematical or statistical techniques for determining the number of clusters in a data set.

ABC

CCC

Gap Statistic

Silhouette

5. (1 Pt.) State **two** drawbacks of hierarchical clustering.

Time/Scalability/Computational Complexity

Determining the number of clusters