DNSC 6279
Fall 2018
Section 06 Quiz

Name:
Email:

1. (2 pts.) Write the centroid for the following rows of data in the same ***k***-means cluster.

Age	Income	DTI	Interest Rate
30	100,000	5	10
32	110,000	6	12
34	105,000	4	11

(32, 10500, 5, 11) – ½ pt. for each

2. (2 pts.) Fill in the blanks. _____ **(Hierarchical / Partitive)** clustering is computationally expensive but can create very accurate clusters. _____ **(Hierarchical / Partitive)** clustering is highly scalable but creates only simple clusters.

3. (1 pt.) Of the choices below, which is probably the most accurate but slowest way to determine a good estimate for ***k*** in ***k***-means clustering?

a. Aligned box criterion
b. Cubic clustering criterion
c. Gap statistic
d. Silhouette statistic

4. (1 Pt.) True or False: Squared error from cluster centroids nearly always increases when adding more clusters into an analysis.

False

5. (1 Pt.) True or False: Standardization of numeric inputs is required for valid cluster analysis results.

True

6. (3 pts.) Write a brief profile (or description) of the two clusters below. Given only the information below, which cluster (1 or 2) is more likely to respond to marketing for expensive products? Cluster 2 (1 pt.)

Age	Income	DTI	Interest Rate	Cluster
30	100,000	5	10	1
32	110,000	6	12	1
34	105,000	4	11	1

Younger, lower income, more debt, higher interest rates on debt. (1 pt.)

Age	Income	DTI	Interest Rate	
50	150,000	2	8	2
55	160,000	0	8	2
49	140,000	4	9	2

Older, higher income, less debt, lower interest rates on debt. (1 pt.)