

CSSE 386 – Data Mining with Programming
Rose-Hulman Institute of Technology

Worksheet 12

Name (Print): _____ Section: _____

1. True/False: PCA is a supervised learning method.
2. Dimensionality reduction is a technique used to _____ the number of features in a dataset while _____ as much of the important information as possible
3. Name 2 issues with multi-dimensional data
 - 1)
 - 2)
4. PCA is particularly useful when the features in the dataset are _____.

5. Complete the table

Characteristics	Feature Selection	Feature Extraction
Original features		
Interpretability		
Use Case		
Goals		

6. PCA is particularly useful when the features in the dataset are _____.
7. _____ is used for quantitative data
 _____ is used for qualitative data with 2 variables
 _____ is used for qualitative data with > 2 variables
8. What is the rule of thumb for selecting the number of principal components to retain in PCA?
9. True/False: The first principal component explains the maximum variance in the dataset.