Introduction to Big Data Analysis

Quiz 6 Reference Answer

1. (1) False (2) False (3) True

2. (1) Clustering is an unsupervised learning method to group the data points based on their similarities, whereas classification is a supervised learning method to classify the data point as one of given labels.

(2) Similarity: Both of them can be used for feature selection to reduce dimensionality.

Difference: PCA is to select some linear combinations of original variables while LASSO directly chooses the original variables.

3. (1) C (2) D (3) A

4. (1) 0 (2) 4 (3) 8 (4) 4