Machine learning worksheet_2

1. Classification and clustering
2. Regression
3. True
4. Capping AND flooring
5. 1
6. No
7. Yes
8. All of the above
9. k-means clustering
10. Creating different models for different cluster groups
11. All of the above
12. The K-means clustering algorithm is sensitive to outliers, because a mean is easily influenced by extreme values

13. K means better beacause If we have large number of variables then, K-means would be faster than Hierarchical clustering.
14. The non-deterministic nature of K-Means is due to its random selection of data points as initial centroids. Method: We propose an improved, density based version of K-Means, which involves a novel and systematic method for selecting initial centroids.

Statistics Worksheet_2

1. Mean
2. 12
3. All of the above
4. Both of these
5. Summarizing and explaining a specific set of data
6. Data set
7. 2 or more
8. Scatterplot
9. Analysis of varience
10. z-score
11. Mean
12. 400005.2

13. Mean
14. Discriptive and infrences
15. H-L