1. Use ‘kc\_house\_data.csv’ data set. Split the dataset in Training and Testing Part. Apply Simple Linear regression (sqft\_living, price) and multiple linear regression for house price prediction. Comment on accuracy of the predictions on test data set.
2. Use ‘kc\_house\_data.csv’ data set. Drop irrelevant features by comparing correlation of price with other features. Split the dataset in Training and Testing Part. Apply Simple Linear regression and multiple linear regression for house price prediction. Comment on accuracy of the predictions on test data set.
3. Use ‘kc\_house\_data.csv’ data set. Derive new features like 'house age' and 'age after renovation'. Split the dataset in Training and Testing Part. Apply multiple linear regression for house price prediction with features- bedrooms, bathrooms, sqft\_living and new derived features. Comment on accuracy of the predictions on test data set.
4. Use ‘kc\_house\_data.csv’ data set. Split the dataset in Training and Testing Part. Apply multiple linear regression and support vector regression for house price prediction. Comment on accuracy of the predictions on test data set with these two ML Algorithms.
5. Use ‘kc\_house\_data.csv’ data set. Split the dataset in Training and Testing Part. Apply multiple linear regression and polynomial regression with degree 2 for house price prediction. Comment on accuracy of the predictions on test data set with these two algorithms.
6. Use ‘Customer\_Actions.csv’ data set. Split the dataset in Training and Testing Part. Apply Logistic regression for binary classification on the label ‘purchased’. Comment on accuracy of the predictions on test data set.
7. Use ‘Customer\_Actions.csv’ data set. Split the dataset in Training and Testing Part. Apply Logistic regression and support vector machine for binary classification on the label ‘purchased’. Comment on accuracy of the predictions on test data set on these two ML classification algorithms.
8. Use ‘Customer\_Actions.csv’ data set. Split the dataset in Training and Testing Part. Apply Logistic regression and random forest for binary classification on the label ‘purchased’. Comment on accuracy of the predictions on test data set on these two ML classification algorithms.
9. Use ‘Customer\_Actions.csv’ data set. Split the dataset in Training and Testing Part. Apply Logistic regression and KNN for binary classification on the label ‘purchased’. Comment on accuracy of the predictions on test data set on these two ML classification algorithms.
10. Use ‘Customer\_Actions.csv’ data set. Split the dataset in Training and Testing Part. Apply Logistic regression and KNN for binary classification on the label ‘purchased’. Comment on accuracy of the predictions on test data set on these two ML classification algorithms.
11. Apply logistic regression to decide whether a person is interested to buy a computer or not? (Use default data set given on shared drive). Split the data for training and testing purpose. Comment on the accuracy of the decisions on test data set.
12. Apply support vector machine to decide whether a person is interested to buy a computer or not? (Use default data set given on shared drive). Split the data for training and testing purpose. Comment on the accuracy of the decisions on test data set.
13. Apply logistic regression to decide whether a person is interested to buy a computer or not? (Use default data set given on shared drive). Split the data for training and testing purpose. Comment on the accuracy of the decisions on test data set.
14. Apply logistic regression and support vector machine to decide whether a person is interested to buy a computer or not? (Use default data set given on shared drive). Split the data for training and testing purpose. Comment on the accuracy of the decisions on test data set on these two algorithms.
15. Apply linear regression and multiple linear Regression to predict CO2 emission using Cars data set. Comment on the accuracy with appropriate performance metrics.
16. Apply SVM for binary classification (Setosa vs Non-setosa) on Iris data set. Use linear kernel function. Comment on the accuracy
17. Apply SVM for binary classification (versicolor vs non-versicolor) on Iris data set. Use linear kernel function. Comment on the accuracy.
18. Apply SVM for multi classification (three classes of species) on Iris data set. Use RBF and Polynomial Kernel functions. Compare accuracy w.r.t. these two kernels.
19. Apply linear regression and Multiple linear Regression to predict CO2 emission using Cars data set. Comment on the accuracy.
20. Apply KNN to decide whether a person is interested to buy a computer or not? (Use default data set shared on shared drive). Calculate prediction accuracy with k=3 and k=5. Comment on the accuracy
21. Apply K-Means Clustering on Mall\_Cusomers.csv with respect to Age and Spending Scores for k=3 and k=5. Use Elbow method to find out optimal k value. Comment on the accuracy.
22. Use PCA for data visualization in 2-D. (IRis data set).
23. Apply PCA on MNIST data set to show the improvement in classification performance with respect to any suitable classification algorithm.
24. Use Wine data set on UCI Machine Learning Repository. Visualize the data in 2 Dimensional space. Apply feature scaling on it and again visualize the same in 2-D and comment on the results.