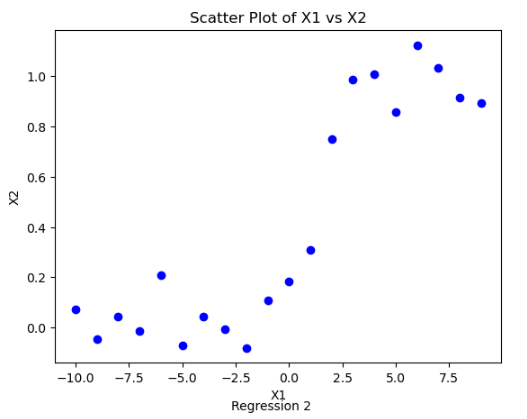
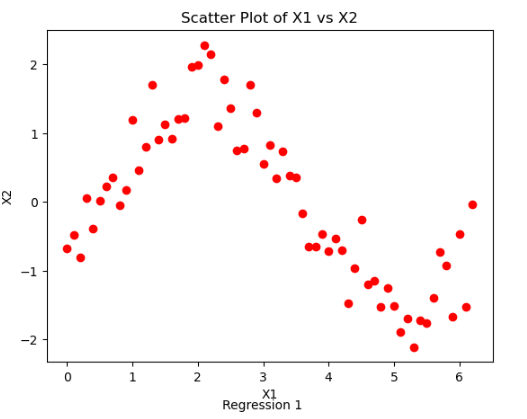
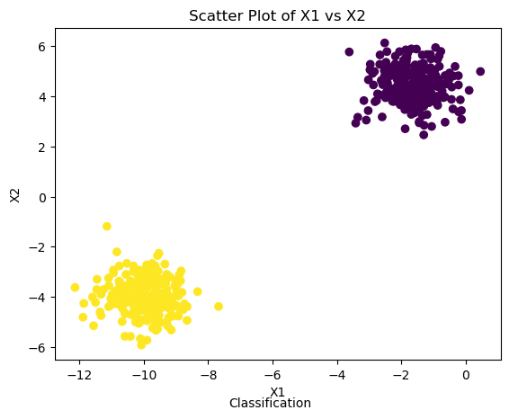
Data Exploration

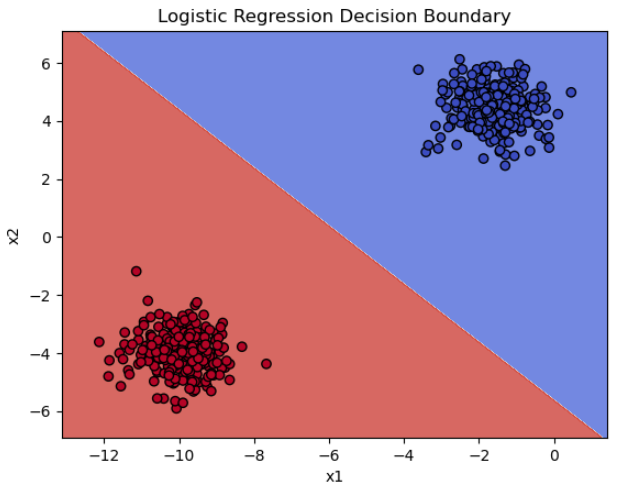


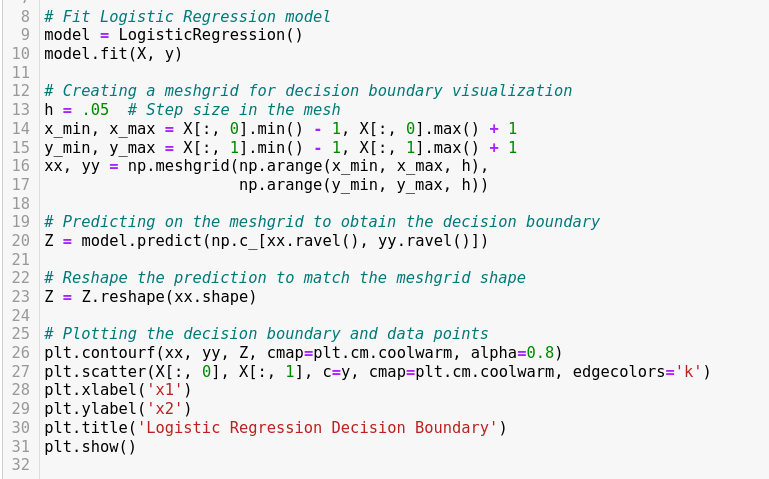




The above representation of data illustrates that, Classification and Regression 1 dataset are less noisy, although there are some outliers. On the other hand, Regression dataset is more noisier.

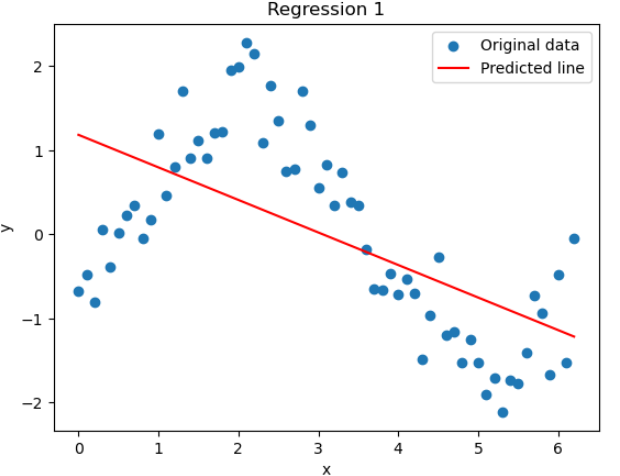
Classification data

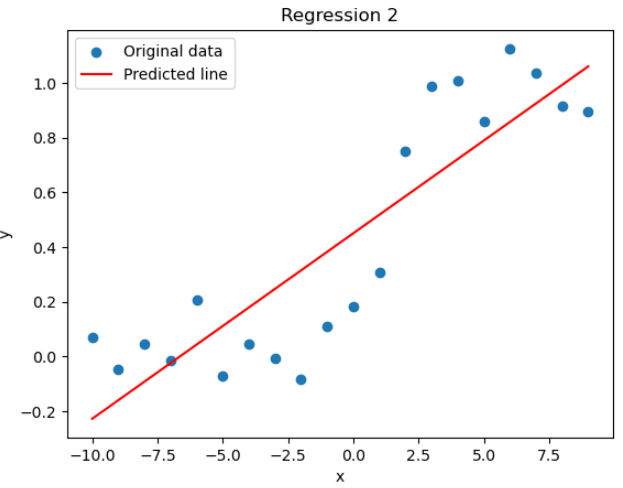


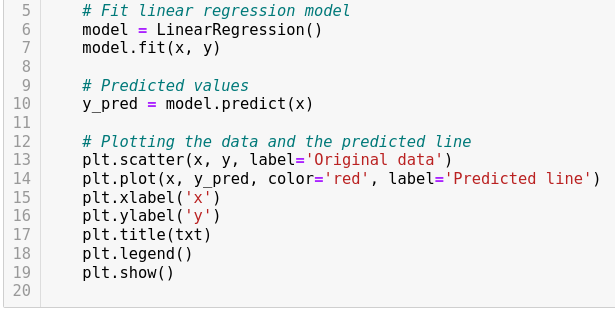


The classification.csv had binary label as 0s and 1s. For such a data I decided to use binary classification algorithm as it models the probability of binary outcome. It also makes clear about how each features contributes to the classification. At first, the data was fitted to a LogisticRegression model and after creating a mesh grid, a decision boundry was plotted to distinguish different regions across the features space.

Plot for regression data







The regression\_1.csv and regression\_2.csv both has only two discrete data x1 and x2. For such a data I decided to use Linear Regression. The data is also quite noisy mostly in case of Regression 2. However, the data plot clearly shows a linear trend reflecting a linear realtionship between the variables although the line is not fitted perfecly.