Assignment 2b (logistic regression and Multilayer perceptron) [50 points]

You will use the same Auto.csv dataset from Week 3. In this assignment, you will train some models and compare their accuracies. **[each item is worth 5 points]**

1. Before the 'MPG' column is categorized to 'Good' and 'Bad', compute the pairwise correlation between the columns in the data. This will give you an idea of how the 'MPG' correlate with the features and how the features are correlated to each other.
2. Split the Auto.csv to training and testing sets. Perform Scaling.
3. Using the training set, train and Neural Network Model (Multilayer perceptron) with two hidden layers. Number of nodes in the first, and second hidden layers should be 100, and 50 respectively. Use early stopping while training. Predict the accuracy on the testing set.
4. Using the same training set, train a logistic regression model without regularization. Predict the accuracy on the testing set. The coefficients of the logistic regression are expected to indicate how the response variable (MPG) depend on the features. Do the coefficients in the model make sense considering what you observe from (1)?
5. How do the accuracies on the testing set compare from (3) and (4) compare? You will attempt to use only two features to train a logistic regression which can outperform model in (3). Plot pairwise scatter plots of the features (use different colors to differentiate the two classes "Good" and "Bad" MPG). The plots should give you an idea of what are the best two features in the data set which make the two class labels linearly separable. Use the two features and train a logistic regression model without regularization. Check if the accuracy on the testing set is higher that the models in (3) and (4). If it is not visually clear, you may attempt different pairwise combinations of few selected features to train logistic regression models.