1. (3 points) A regressor with smaller RMSE on the testing set is preferred.
   (a) True
   (b) False
   (c) It depends
   (d) All of the above
   (e) None of the above
2. (3 points) Which of the following models is preferred?
   - Model 1 has a MAE of 10.7 on the testing dataset
   - Model 2 has a MAE of 12.8 on the testing dataset
   - Model 3 has a MAE of 8.5 on the testing dataset

   (a) Model 1
   (b) Model 2
   (c) Model 3
   (d) Models 1 and 2
   (e) Models 1 and 3
   (f) Models 2 and 3
   (g) All of the them

Consider the `autos.csv` datafile. Each row represents a car, each column contains information such as horsepower, number of cylinders, etc. The goal is to predict the miles per gallon, `mpg`, using the other car attributes.

3. **In R**, answer the following:
   (a) (3 points) Using the `read.csv` function, read the csv file and create a data-frame called `autos`.
   (b) (5 points) Using the z-score standardization formula, put `cylinders`, `displacement`, `horsepower`, `weight`, and `acceleration` on the same scale.
   (c) (4 points) Split the data into `train` (80%) and `test` (20%)
   (d) (6 points) Using the `train` dataset and the `knnreg` function from `caret` package, build a 5-nearest neighbors regression model called `knn_md`, in which `cylinders`, `displacement`, `horsepower`, `weight`, and `acceleration` are the input variables, and `mpg` is the target variable. Using the `knn_md` model, predict the `mpg` in the `test` dataset. Compare the predictions and actuals using RMSE and MAE.
   (e) (6 points) Using the `train` dataset and the `randomForest` function from `randomForest` package, build a random forest regression model called `RF_md`, in which `cylinders`, `displacement`, `horsepower`, `weight`, and `acceleration` are the input variables, and `mpg` is the target variable. Using the `RF_md` model, predict the `mpg` in the `test` dataset. Compare the predictions and actuals using RMSE and MAE. Make sure you use `ntree = 500`.
   (f) (6 points) Using the `train` dataset and the `gbm` function from `gbm` packages, build a gradient boosting regression model called `gbm_md`, in which `cylinders`, `displacement`, `horsepower`, `weight`, and `acceleration` are the input variables, and `mpg` is the target variable. Using the `gbm_md` model, predict the `mpg` in the `test` dataset. Compare the predictions and actuals using RMSE and MAE. Make sure you use `n.trees = 500`, `interaction.depth = 4` and `distribution = 'gaussian'`.
   (g) (3 points) Considering RMSE and MAE, what model would you select to make predictions? 5-nearest neighbors? random forest? or gradient boosting? Explain.