DATA-101 IN-CLASS ASSIGNMENT OCTOBER 12, 2021

1. (3 points) Is linear regression a supervised learning method?
   (a) True
   (b) False
   (c) It depends
   (d) None of the above
2. (3 points) Which of the following methods do we use to find the best fit line for a dataset?
   (a) Maximum likelihood estimation
   (b) Least squares method
   (c) Logarithm loss estimation
   (d) None of the above
3. (3 points) Which of the following is **TRUE** about the residuals?
   (a) smaller is better
   (b) bigger is better
   (c) It depends on the problem
   (d) None of the above
4. Consider the `Automobile_data.csv` datafile. The Automobile dataset has a different characteristic of an auto such as body-style, wheel-base, engine-type, price, mileage, horsepower and many more. **In Python**, answer the following:
   (a) (3 points) Using the pandas library, read the csv datafile and create a data-frame called `autos`
   (b) (4 points) Create a scatter-plot between `horsepower` and `price`. Comment on the plot.
   (c) (5 points) Build a linear model, in which `price` is the target variable and `horsepower` is the input variable. Interpret the slope of the line.
5. Consider the `Automobile_data.csv` datafile. The Automobile dataset has a different characteristic of an auto such as body-style, wheel-base, engine-type, price, mileage, horsepower and many more. **In R**, answer the following:
   (a) (3 points) Using the pandas library, read the csv datafile and create a data-frame called `autos`
   (b) (4 points) Create a scatter-plot between `wheel.base` and `length`. Comment on the plot.
   (c) (5 points) Build a linear model, in which `length` is the target variable and `wheel.base` is the input variable. Interpret the slope of the line.