STAT 510 Final Project

Final Description

- **Main Objective**

  Find the best (linear) regression model.

- **Objective: I would like to see**
  1. Your capability of understanding all the topics we learned.
  2. Your optimal decision for choosing the hyperparameters.
  3. Your capability of code implementation.

- **Team members**
  1. Every team has 2 students (only one team will be 3 students).
  2. Choose a team member by 10/18.

- **Timeline**
  1. Send your data set (link is okay) and brief description by 10/25.
     i. It should have more than 20 input features.
  2. Submit your final report by 11/29.**12/3**.
  3. Presentation (last week).

- **Detailed Description for Final Report**
  1. Length: up to 10 pages. **(not a strict rule)**
  2. Contents:
     i. Motivation and Data summary (1-2 pages)
     ii. Descriptive Statistics (1 page)
     iii. Result.
     iv. Discussion (analysis)
  3. Result part:
     i. A default linear model (without any hyperparameter tuning).
     ii. A linear regression with selected features.
     iii. Modified regression (polynomial, Ridge, Lasso, …)
     iv. You don't need to show the process of hyperparameter tuning.
     v. Comparison table.
  4. Discussion and analysis about your models.
     i. **Show your extra effort to make the best model compared to a default model**
        - **Feature selection**
        - **Outlier/influential points**
        - **Multicollinearity**
        - **Regularization**

- **More changes if needed (feature transformation, standardization, normalization).**

ii. **To this end, you may split your data into training/test data.**

iii. Why a certain model is better than other models (with respect to what?)?

iv. What is the meaning of your result?