Dr. Tony Diana
DATA 602 Introduction to Machine Learning
Homework | Week 4

---

1. **Simple Regression Using statsmodel**

- Import statsmodel, pandas, and numpy
- Import Boston datasets from scikit-learn datasets and load it
- Describe the dataset
- Define the data/predictors as the pre-set feature names
- Put the target (housing value -- MEDV) in another DataFrame
- Print first five rows
- Define 'RM' as X or the feature and 'MEDV' as y or the label
- Build the model and define 'predictions'
- Print out the statistics

2. **Multiple Regression Using statsmodel**

- Use 'RM' and 'LSTAT' as features and keep 'MEDV' as the target - Define the model, define 'predictions', and print the summary -What does a coefficient of determination of R^2= .948 mean?
- Is there any autocorrelation? If not, why?
- Are the data coming from a normal distribution? Why?

3. **Multiple Regression Using scikitlearn**

- Define the data/predictors as 'RM' and 'LSTAT'
- Put the target (housing value -- MEDV) in another DataFrame
- Define the linear model as lm and fit -
Print the predictions
- Print the coefficient of determination
- Print intercept and coefficients