

Assignment (Due May 15th 10AM)

Apply regression analysis or classification to a data set of your choice.

1. Collect or identify a data set (e.g., UCI Machine Learning Repository <http://archive.ics.uci.edu/ml/index.php>).
2. Implement and apply a regression or classification method you learned in class.
 - ▶ Hold out a small subset of the data to evaluate accuracy of predictions/classification.
 - ▶ Try various methods (e.g., basis functions).
 - ▶ You CANNOT just call a library function for the core analysis; you have to implement it yourself.
3. Write a report that clearly discusses the following (preferably as a Jupyter notebook).
 - ▶ Explanation of the data you used. Clearly state the objective (e.g., predict B from A, binary classification of A and B, etc.).
 - ▶ Explanation of the method you implemented and applied with math.
 - ▶ Evaluation including quantitative analysis and adequate plots.
4. Submit, via PandA, the data, code, and report. If your data is large, include a link instead of uploading.