ESE 588: Pattern Recognition
Spring 2019
Faculty: Vibha Mane

Stony Brook University

Regression Module Project
Assigned Date: February 14, 2019
Last Update: February 14, 2019
Score: 21 points

1 Project Goals

Implement the following machine learning tasks, utilizing **Regression techniques**:

- Prediction
- Classification
- Feature Reduction
- Feature Independence
- Model Selection (underfitting and overfitting analysis)

2 Required Components

You may utilize Python or Matlab libraries, for the following implementations.

1. Implement prediction utilizing multiple linear regression on a data set with several features. Perform an evaluation of the residuals to check for assumptions of your model, such as linearity, noise term with zero mean and constant variance, normality and so forth.
2. Implement two-class classification utilizing logistic regression, for data sets with several features.
3. Implement feature reduction, utilizing hypothesis testing and p-values on the standard error, for each estimated parameter. Therefore, if an estimated parameter is not statistically different from zero, the associated feature maybe excluded from the model.
4. Test your algorithms and implementations on two data sets from the UCI, Kaggle or the ImageNet repository. One of these data sets can have less than 5 features. The second data set should have at least 10 features.

   (a) Some example data sets for the prediction task are:
   - Boston Housing Data to predict the median price of housing.
   - Wine Quality Data to predict the quality of wine.
   - Million Song data to predict the release year (select a sample).

   (b) Some example data sets for the classification task are:
   - Classify email as spam or not spam.
   - MNIST handwritten digit classification (you can select 2 out of the 10 digits).
   - Image object classification.
   - Medical diagnostics.

3 Creative Component

Some suggestions for creative component are:

- Ridge regression to remedy overfitting.
- Robustness to outliers.
- Chi-Squared test for independence of variables.
- Multi-class classification with logistic regression.

4 Project Report

The project **must** include a report with several sections, as suggested below.

- Introduction: Describe the project background and goals.
- Description: Detailed description of the project including any necessary equations.
- Algorithm and Implementation: Describe the algorithms, and the names of the Python/Matlab libraries utilized. You may use code snippets to describe the algorithms.
- Results and Conclusions: Analysis, plots, feature reduction/selection, interpretations and numerical errors. How do the results compare with theory? Summarize your findings.
- References and Citations: Please provide all citations, for example, the libraries utilized, books, articles and web links.

5 Project Teams

- You may work in teams of up to 2 persons.
- Each member of the team must make a **significant contribution** to the project.

6 Required Meetings

Each team is required to have 2 meetings with me. These maybe during office hours or during the posted time slots.

7 Additional Notes

- You may utilize Python or Matlab libraries for the project.
- Try to make the creative component unique, perhaps with distinctive data set. Please discuss with me.