Randomized Regression Project Outline

Roles

- Data Generation
- Counter Researcher
- Python Translator
- Unsupervised Learning
- Project Manager

Data Generation +
Counter Researcher

New Flavors of
Random Lasso

Robust Python
Translation

Robust Python
Translation

? Unsupervised
Leanring ?

Data Generator

Goals:

- Find real data sets for the group to test our randomized regression algorithms.
- Generate simulated data by creating a robust function that takes inputs {n, p, s}.

  {n=#samples, p=#features, s=#significant}.

- Implement automated testing of Random Lasso using simulated data generation function. What happens as n, p, and s increase?

Benefits

- We know that random lasso tends to preform better than other flavors of regression analysis when:

p >> n features >> samples

p >> s features >>> significant features

- We want to be able to definitively state for what datasets random lasso is more optimal.

Suggested Skills

- Strong background in statistics.
- Has used random numbers to generate data before.
- Some programming experience.
- Does not have to implement random lasso, only has to implement automated testing of random lasso.

Counter Researcher

Goals:

- Find state-of-the-art algorithms in literature that surpass our best random lasso algorithms in accuracy metrics.
- Decide on best accuracy metrics to used based on what existing literature, e.g. RMSE, F1, etc.
- Organize results in tables and charts that others can easily understand.

Benefits

- Determine's if our algorithm establishes a new norm for state-of-the-art.
- Group capitalize on any competitive algorithms that we find by: (1) implementing a randomized flavor, and (2) testing if that randomized flavor performances better.
- In other words, we can absorb any other algorithm we find as a phase in random lasso.
- Wow factor of results!!!

Suggested Skills

- Has completed some kind of literature review.
- Has compiled and used tools created by other researchers before – from a variety of languages.
- Interest in exploring the field of regression analysis.
- In some cases, the Counter Researcher may have to implement these competitive algorithms on their own in order to test their performance.

Python Translator

- Translate existing random lasso and high lasso algorithms from R → Python.
- Accuracy test the results of their implementation.
- Compare to existing random lasso algorithm.

Suggested Skills

- Coding and debugging.
- Some knowledge of functional programming.

Unsupervised Learner

- Independent research.
- Implement a random-lasso-like algorithm for unsupervised learning. Use the traits of random lasso in existing unsupervised learning algorithms.
- Heavily in collaboration with project manager. Project manager may assist with implementation, but independent researcher is to formulate algorithm.

Benefits

- This is required for the project (TBD).
- It is quite possible that these core traits of the random lasso could improve existing unsupervised learning algorithms.
- Could result in a novel algorithm.
- It is okay if results are poor, but implementation of an algorithm is required. If results are poor, explain why random lasso does not work for supervised learning.

Suggested Skills

- Comfortable with independent research.

- Already has some knowledge in unsupervised learning algorithms.

Project Requirements

- **Supervised**: The computer is presented with inputs (independent variables "x") and associated labels indicating the class of the observation (dependent variable "y"). The computer attempts to learn the rule that maps inputs to each class. New data is classified based on the rule learned by the computer.
    - [✓] Regression Analysis is supervised.
- **Unsupervised**: The computer is presented only with inputs (independent variables "x"). The computer attempts to classify based on similarity/dissimilarity
    - [x] Does not exist.

More Info of Supervised v. Unsupervised

- https://stats.stackexchange.com/questions/61390/is-there-unsupervised-regression

Project Manager

- Organize the logistics of the project, e.g. group meetings, timelines.
- Responsible for analysis and most of writing.
- Be available at anytime to provide guidance on random lasso.
- Provide accuracy functions for group, e.g. RMSE, F1. Provide random lasso function. Git repo available soon.
- Any areas of the project that need require extra effort will be picked up by the project manager, i.e. Unsupervised Learning.
- Ensure project fully meets course guidelines.

“Type a quote here.”

Is it common for the Supervised Learning and Unsupervised Learning portions of the project be disjoint?

–Johnny Appleseed

Let's have scrums two times a week!

When are you all available?