

Machine Learning

Aaron

Ex1

Bonus Target And Features

- Regressing Bonus Against LTI
- Salary Vs. LTI For Predicting Bonus
- Outliers Break Regressions

Ex2

- Slope and Score of Regression with Outliers
- Slope and Score after Cleaning
- Enron Outliers
- Any More Outliers?
- Identifying Two More Outliers. Remove These Outliers?

Unsupervised Learning



<http://www.naftaliharris.com/blog/visualizing-k-means-clustering/>

Ex3

- What features will your clustering algorithm use?
(*k_means/k_means_cluster.py*)
- Deploying Clustering
- Clustering With 3 Features
(*'salary', 'exercised_stock_options', 'total_payments'*)
- Stock Option Range
- Salary Range

Ex4

- What Kind Of Scaling
- Computing Rescaled Features (*a "salary" value of \$200,000 and an "exercised_stock_options" value of \$1 million*)
- Computing Rescaled Features
- When To Deploy Feature Scaling

Ex5

- Warming Up With ParseOutText()
(*run "tool/parse_out_email_text.py"*)
- Deploying Stemming
- Clean Away "Signature Words"
- TfIdf It
- Accessing TfIdf Features