Demystifying the Workings of Lending Club

Project Overview

Data Pre-processing
- Classification
- Regression
- Principal Component Analysis (PCA)
- Data Visualization
- Predicting Outcome
- Feature Significance
- Predict Interest Rate
- Approval Standards
- Clustering (T-SNE)

Results

Classification
- Predict if loan application is approved
- 98% F-measure in 2015
- Increasing ability of the model to make better predictions given more data
- Hypothesize that the dip in '08-'09 performance is due to limited data and absence of significant attributes

Regression and PCA
- Predict interest rate and tackle noise by dimensionality reduction
- Notice that there is a noticeable 'kink' in the data after 3 principal components
- Loan Grade is highly predictive of interest rate
- PCA ($K = 3$) increases RMSE $\sim$25%
- RMSE increases steadily over the years
- Hypothesize that this is due to increasing underlying model complexity

Data Visualization
- Apply clustering to find structure related to Loan Purpose and States
- 22 dimensions visualized using T-SNE
- Clear clusters in the high dimensional space suggest definite sparse structure where similar loan purpose are found together
- Results may be used to generate artificial examples, specially for years with limited data available

Analysis

Predicting Outcome
- Lending Club has gradually relaxed its loan approval standards
- Of the approved applications, notice growing riskiness through increases in debt-to-income ratio and loan amount
- Hypothesize this relaxation was to prepare for their IPO in 2014

Feature Significance
- Certain features are constantly predictive of whether a loan is approved or denied
- Educational loans are likely to be denied
- Credit card consolidation and debt consolidation loans are likely to be approved
- These features are significant at $p = 0.01$
- Put 'credit card consolidation' as the loan purpose to game the system and have better chance of loan getting approved.

Pujun Bhatnagar, Nick Chow, Max Lai