Art of Systematic Prepayment Modeling

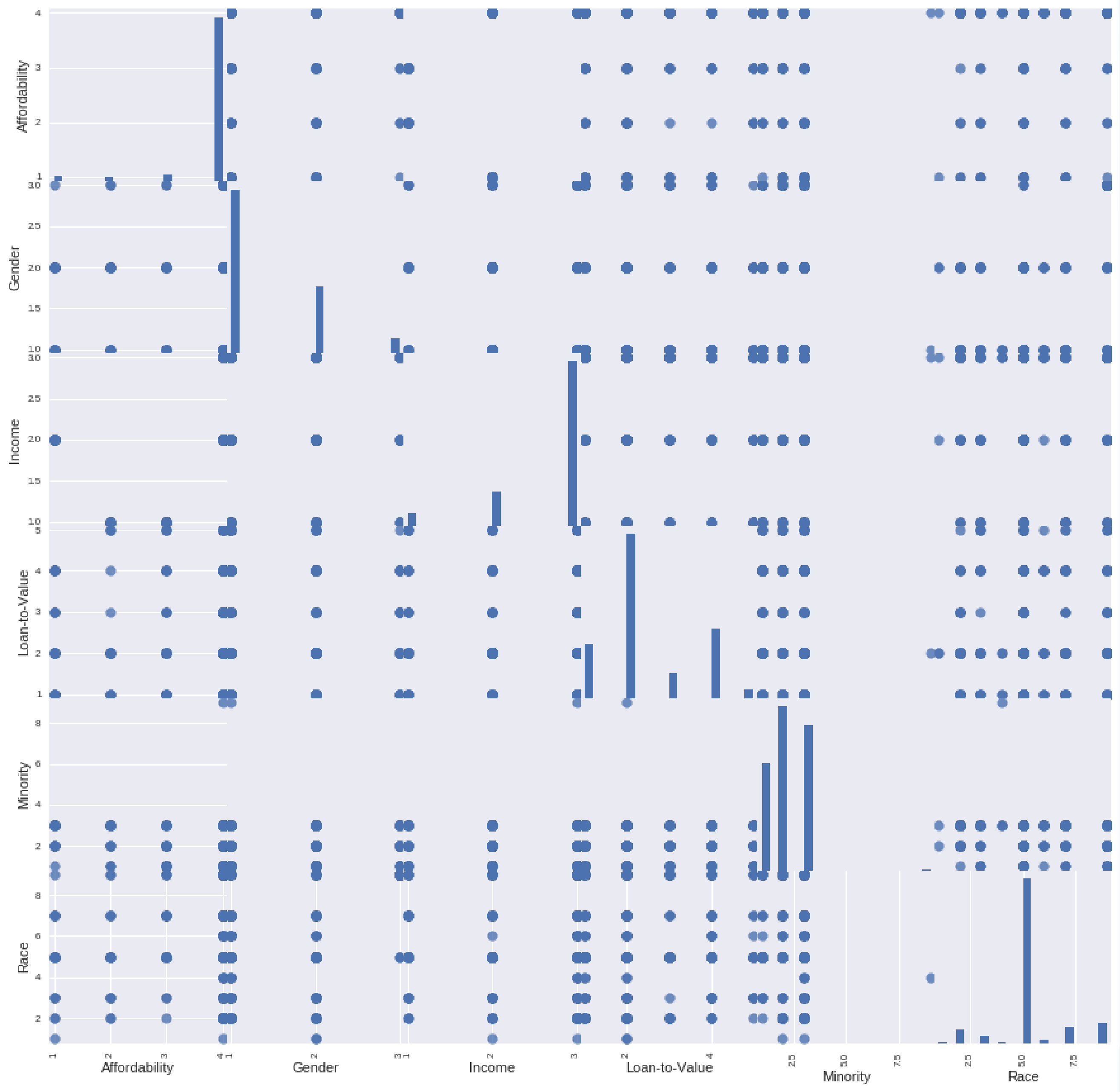
Dr. K. Sohrabi Anaraki
(Postdoc, PhD McGill U.)

Introduction

- For many financial institutes mortgages are considered important part of revenues.
- With any possible revenue comes its corresponding risk.
- Some researchers have scrached the surface but the problem is complex due to many different parameters correlating with each other.
- One direct factor of measuring risk is this regard is called Prepayment. That's when the borrower of the mortgage leaves the contract.
- Depending on the time of Prepayment, a financial institute can profit or lose. We use the public data of Fannie Mae listed in the Federal Housing Finance Agency for our purpose.

Complex relation Between:

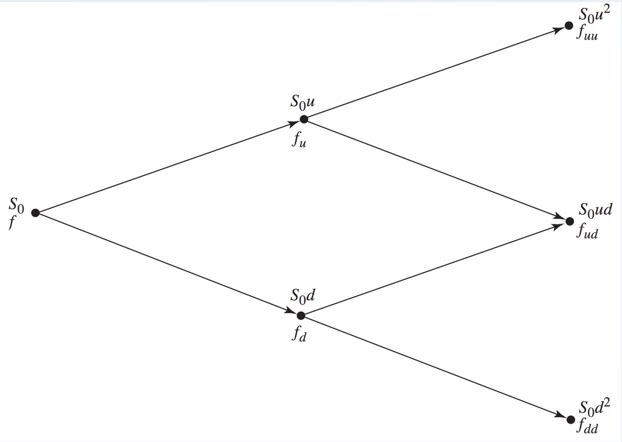
- Affordability
- Gender
- Income
- Loan to Value
- Maturity
- Race
- …



Our Road Map

- Our goal is to propose a deep learning method for the prediction of prepayment.
- One way to shed light on this, is to determine the profile of the mortgage borrowers and classify them.
- Classification can result in the predictability of the mortgage termination or refinance.
- The first step is always to understand the data and make it as intuitive as possible.
- Eventually this will help financial institutes to trade fixed-income products efficiently.
- We start with a simple example

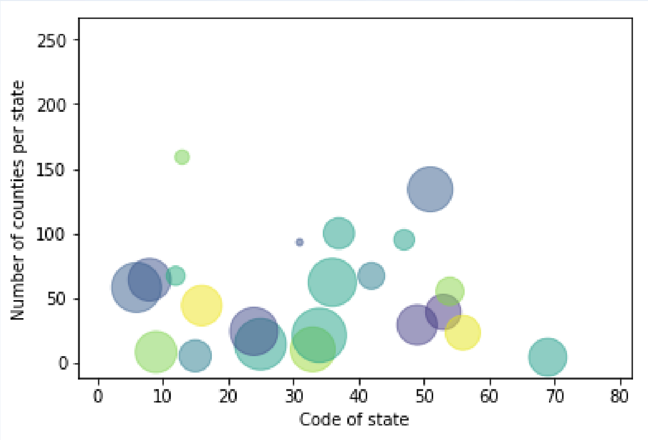
Geometric Random Walk

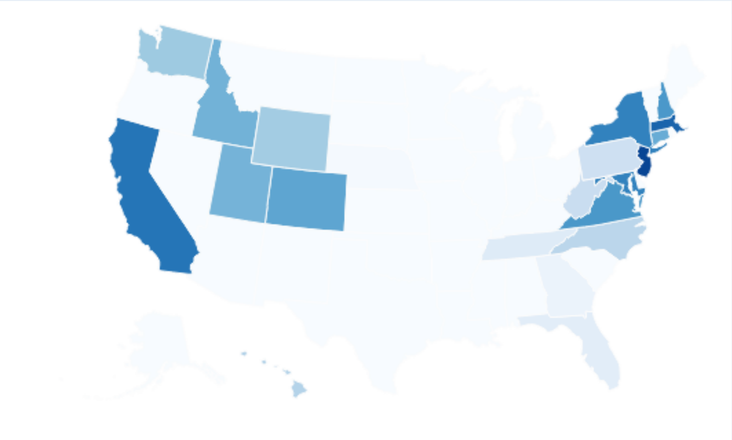


$$V_s = \frac{1}{1+r_s}\left[\frac{1}{2}\left(V_0 + V_{su}\right) + \frac{1}{2}\left(V_0 + V_{sd}\right)\right]$$

- If the interest rate has a random walk behaviour, we can use Backward induction to find the value of each node.
- We can generalize the non−callable bond expression to include coupon payment and then look at resulting values to dissolve the mortgage. This will lead to Prepayment.
- In general the problem is more complex due to extra parameters.

Complexity of Data



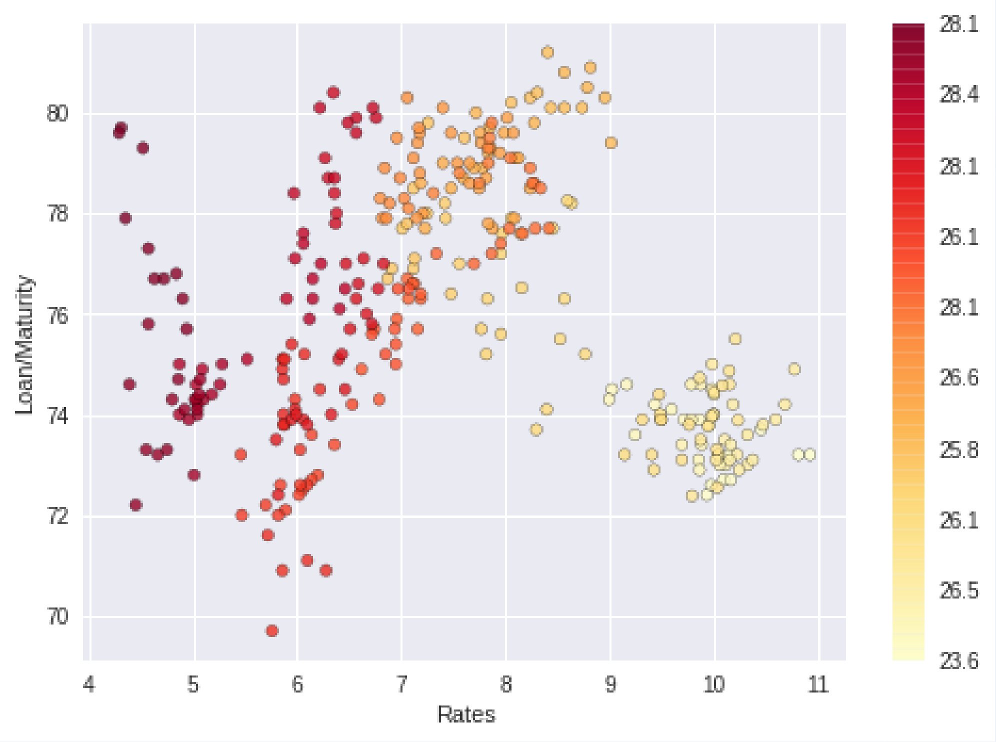


- Let's start with a bottom-up approach. Each state has a various number of counties. Using the 2017 data from Fannie Mae and Freddie Mac of Single-Family Mortgages, we can see an interesting concentration between the number of counties and standard deviation of the Maximum Loan Limit Mortgages.
- On the right plot, record of these deviations are sorted and represented by color. California and New Jersey have the highest variance while mid-east states have the lowest. Fun fact: A relatively good overlap with Republican State Houses.

Patterns in the Data

- Within this complex data, we can look for patterns. Using the 2016 data of the Federal Housing Finance Agency for the Single–Family Mortgages, we can plot Loan to Price ratio vs. Interest Rates while representing Maturity of the Mortgage by the red color. It's clear that 4 different clusters appear.

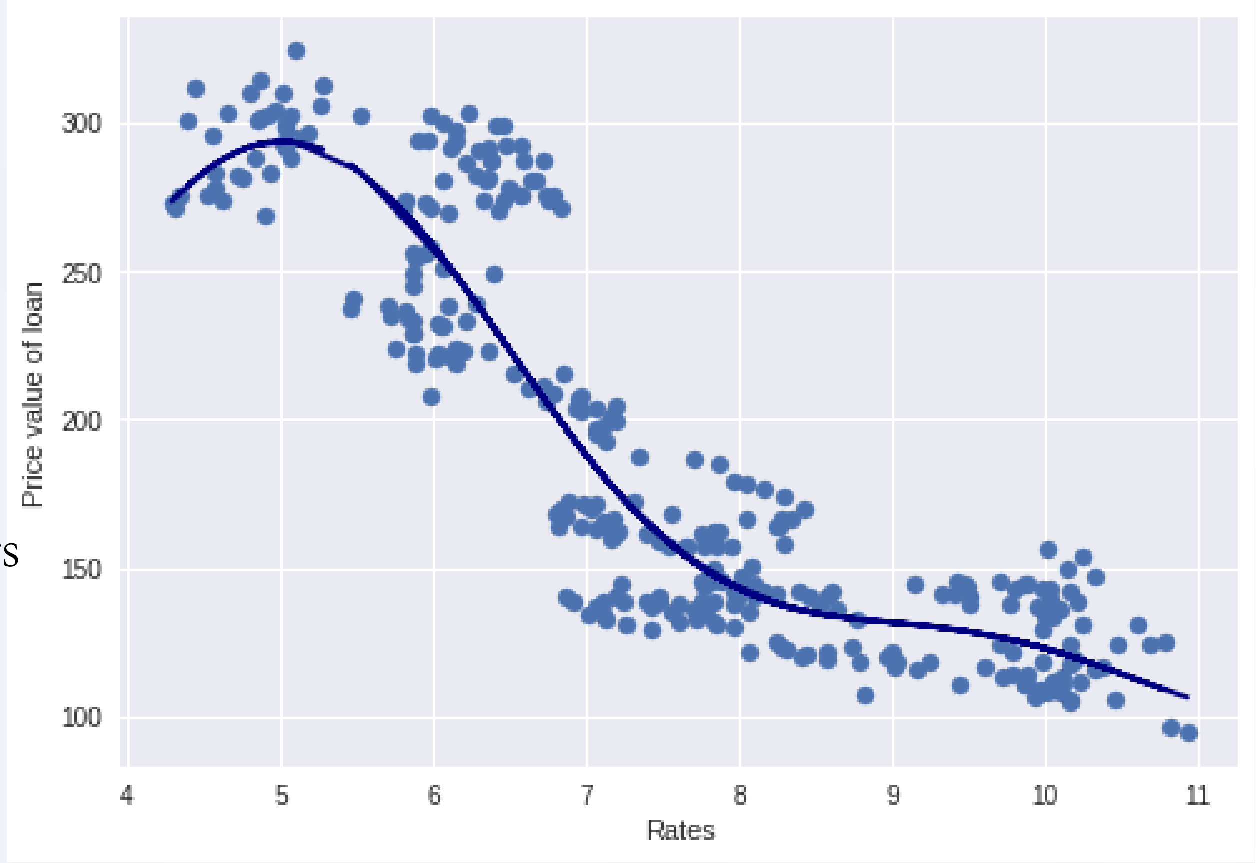
Finding patterns and relations are part of understanding the data and will help us in designing our Deep Learning model and specifically Classification Regression modeling.



Predictability

- We can dig a bit deeper in the previous data set by plotting Purchase Price vs. Interest Rate. Clearly the relation is not linear. We use Support Vector Regression for the non-linear fit below.

This is a very well known plot for experts in Prepayment Modeling. As the Interest Rate goes down, Prepayment goes up. But there is a limit to it and eventually all mortgage borrowers will Prepay thus explaining the top negative convexity.



Conclusion

- We have analysed public Home Mortgage Data from Federal Housing Finance Agency acquired by Fannie Mae and Freddie Mac for a period of 2008–2017 and cherry picked some of our results.
- While Prepayment is a complicated problem, our analyses uncover promising relations that will be helpful in Deep Learning modeling.
- We could reproduce one of the most famous plots in Prepayment Modeling utilizing Supervised Machine Learning and historical data.
- Our goal is to extend these analyses to Unsupervised Machine Learning for example through Tensor Flow in order to do the Systematic Behavioral Modeling. Our results will then have a predictive power.