Analysis of Peer-Lending Through The Lending Club

Carl A. B. Pearson

November 17, 2013

Introduction

The Lending Club[1] is a service which connects borrowers and lenders. It advertises itself as a way to short-circuit banking institutions, thereby reducing transaction costs, and sharing the resulting gains back to the participants. This analysis provides a fitting of their mediated interest rates for small loans.

1 Methods

For specific code, see github.com/pearsonca/data-analysis-assn-1 for the source `.rnw` file.

1. obtained data set from https://spark-public.s3.amazonaws.com/dataanalysis/loansData.csv
2. transform interest rate into numeric data,
3. transform Debt-To-Income ratio into numeric data,
4. create a new numeric field FICO.num from the FICO[2] range from the middle of the range,
5. scale Amount Requested to thousand dollars and FICO to hundreds of FICO,
6. assign 0 to resolve missing values in Monthly Income, Open Credit Lines, Revolving Credit Balance, and Inquires in the Last 6 Months fields,
7. review box plot of interest rate by the FICO range factors, which indicates a fairly clear "elbow"
8. iterating over the various levels of FICO score, review scatter/box plots of the interest rate as a function of other parameters.

2 Results

For a broad range of scores, the amount of money requested appears to have noticeable linear relation with the interest rate. The loan duration also appears to distinguish groups. Using R and its `lm(formula, ...)` fitting function, we can

```
> dt <- clean(read.csv("./loansData.csv"))
> # with clean(...) operations in Methods
> model<-with(dt,{lm(Interest.Rate~Amount.Requested+FICO.num+Loan.Length)})
```

which yields a model (with coefficients rounded to the significant figures of the input)

$$\text{Int.Rate} = 73 + 0.14 * \text{Amt.Req} + -8.8 * \text{FICO} + 3.3 * (\text{Loan.Len} == 60)$$

with fit statistics as summarized by R: $r^2 = 0.74$ and coefficient significances: We can evaluate these fits visually by considering them as image plots (see Fig 1 plots). The images for the data display consistent color transitions, rather than jumping between interest rates, indicating a smooth surface across the two parameters.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	72.6319	0.8548	84.97	0.0000
Amount.Requested	0.1384	0.0060	23.20	0.0000
FICO.num	-8.7590	0.1210	-72.37	0.0000
Loan.Length60 months	3.2927	0.1121	29.37	0.0000

3 Conclusion

The primary drivers of loan interest are (1) the expected risk associated with loan (FICO), (2) the amount of capital at risk (Amount Requested), and (3) the maximum time that capital is at risk (Loan Length). Most of the variation in the interest rates can be explained by these parameters alone.

4 References

References

[1] Corporation L (2013). The lending club. URL `https://www.lendingclub.com/`.

[2] Anonymous (2013). Fico score. URL `http://en.wikipedia.org/wiki/FICO_score#FICO_score`.

```
quartz_off_screen
                2
```

```
quartz_off_screen
                2
```