15.072: The Analytics Edge **Fall 2021**
Problem Set 1: Regularization and Resampling

Out: September 20; Due: October 05, 11:59 pm.

Milagro: Predicting Store Profitability at a Fast-Casual Restaurant Chain

Acknowledgments: This problem is based on a business case written by Prof. Robert Freund and Phil Zakahi (©Robert Freund). The questions have been adapted for the purpose of this assignment.

Introduction

Since graduating from business school three years ago, Kathleen Ortiz has been a consultant at Berkeley Vista Associates (BVA), a boutique consulting firm based in San Francisco, California, that specializes in work for multi-store retail companies. Kathleen is considered a rising star by the partners at BVA. Indeed, many of her clients have raved about her commitment to their success, as well as her uncanny ability to challenge them to think differently about their companies' strategies. Kathleen joined BVA upon graduation from business school at MIT's Sloan School of Management. She assumed she would spend two years with the firm before moving into an industry position. However, Kathleen has enjoyed being a consultant with BVA, and she is now happily setting her sights on becoming a partner.

Stefan Bartell, a partner at BVA, is the other Sloan graduate at the firm. Stefan has come to rely on Kathleen for many of his client engagements. He is one of two partners responsible for work with Milagro, a fast-casual restaurant chain based in the Southwest United States.

Milagro

Siblings and first-time restaurateurs Julia and Noah Silva founded Milagro in 1995 in Tucson, Arizona. The restaurant was originally intended as a casual sit-down Mexican food restaurant, typical of the family-owned Mexican restaurants throughout the Southwest. However, the Silvas quickly discovered they were getting more business from lunchtime takeout than from evening dining. Julia Silva wisely re-engineered Milagro accordingly. She gave up their expensive lease and moved from a stand-alone restaurant to a storefront in a strip-mall across the street with about half the space. They laid off their wait staff and decided to focus exclusively on lunch-time counter service. They pitched themselves as a sit-down quality Mexican restaurant at Taco Bell speed. Profits soared, and they quickly expanded to breakfast and dinner-time takeout as well. After just over a year in the new format, Julia realized Milagro had tapped into a huge unmet demand in the market. She convinced Noah that they needed other locations in the city. She launched four more Milagros in and around Tucson. In 1998 Milagro was named the Number One up-and-coming business by Tucson Business Weekly. On January 1, 1999, Julia became the (sole) CEO of Milagro.

Under Julia's leadership Milagro experienced tremendous growth. By the end of 2003 they had 25 stores in 8 states. Milagro's location strategy was to focus on small-footprint stores positioned at busy intersections. Most Milagro stores remained in strips malls, but it did have success with stand-alone stores featuring drive-through operations. In the early 2000s Milagro was perfectly positioned to ride the rise in popularity of fast-casual restaurants. Consumers were fleeing both low-quality fast-food restaurants, and the time and expense involved in a trip to a sit-down casual restaurant; instead they turned to restaurants like Panera Bread and Milagro. Fueled by an injection of outside capital, Julia drove for increased penetration in the Southwest. By 2008, Milagro had 153 stores, by 2015 it had 459.

Milagro and Berkeley Vista Associates

It was over this period of rapid growth that BVA began its relationship with Milagro and Julia Silva. In 2005 Julia brought in BVA to help roadmap new stores for the next five years. Julia initially opposed BVA's 2006 recommendation to favor store leases over owning real estate, but she apprehensively adopted this strategy nevertheless. It turned out that this decision likely saved the company when the bottom fell

out of the real estate market in 2008. In 2009, BVA helped Milagro cut costs and renegotiate leases to weather the trough created by the recession and the financial crisis. Finally, in 2012, BVA helped Milagro engineer a new marketing strategy emphasizing food quality and health—a growing concern given long-term impressions of Mexican food as not health-oriented.

Milagro has proven to be a critical relationship for BVA as well. Milagro's success has been a constant point of pride for BVA, and Julia has not been shy about pitching the value of BVA's work to other CEOs in her network. Furthermore, the work Julia Silva gave BVA in 2009 helped BVA avoid layoffs in a very challenging year.

Decision to Exit

After 20 years of growing Milagro, Julia Silva was ready for a new challenge. In early 2015, Julia decided to shop the company around to potential private equity (PE) buyers. Her outside investors were ready to exit, and she decided it was an opportune time to sell her (and her brother's) stake as well.

Julia asked BVA to represent Milagro in the strategic end of the due diligence process. As PE firms explored buying Milagro, BVA would prepare analyses and reports showing how Milagro is positioned for future growth and profitability. BVA would also make itself available for strategic and financial analyses.

To aid Julia as she started discussions with firms, BVA conducted several key pitch documents in early 2015. This included a general report on Milagro's strategic position, a financial analysis projecting the company's profitability for several years into the future, and a potential growth assessment. Underlying these documents were a variety of quantitative and qualitative analyses. One of these analyses was a regression model built by Kathleen Ortiz and her team, designed to predict the future profitability of the 48 Milagro stores currently under construction (and which have obviously not yet opened).

BVA's Regression Model of Projected New Store Profitability

Given Milagro's high growth business model, Milagro's operational success depends critically on how they choose locations for new stores. At first, such store location decisions were made by intuition combined with real estate site analysis. Over the last several years, BVA—and in particular Kathleen Ortiz and her team—turned to analytics to select the most profitable sites for new Milagro stores.

In preparing analyses and forecasts for the Milagro sale pitch documents, Kathleen realized she could adapt this regression approach to create a solid forecast for the profitability of the 48 Milagro stores under construction. She pitched the idea to Stefan Bartell, who enthusiastically supported the idea. In April, she started work on this analysis.

Kathleen's team assembled the basic data for their regression modeling work. They split the data on the 459 stores into a training set of 374 stores and a testing set of 85 stores. Table 1 presents the full list of variables and their descriptions. the team developed a linear regression model to forecast annual profits of the stores under construction. After a bit of analysis, the team decided on the following model:

$$\begin{aligned}\texttt{Annual Profitability} = &83597.09 + 0.0028073 \times \texttt{agg.inc} + 383.36 \times \texttt{sqft} + \\ &346821.13 \times \texttt{col.grad} + 218265.764 \times \texttt{com60}\end{aligned}$$

After studying the regression output, Kathleen was reasonably pleased with the model. The R^2 value was 0.786, which was very good. Also, all of the coefficients were significantly different from zero. They also ran the model on the 85 stores in the test data, i.e., on stores 375 through 459. They found that the out-of-sample R^2 was 0.719, which again was good.

Using this regression, predicting the total profitability of the 48 stores under construction was then a simple plug-and-compute exercise. According to the above model, the expected profitability of the 48 stores under construction is $40.02 million. Stefan Bartell communicated the $40.02 million figure to Julia Silva. In addition, this figure was incorporated into a variety of the pitch materials that Julia was shopping around to potential buyers.

Toward a Better Regression Model

With the goal of building a better model of profitability, Kathleen instructed her team to seek ways to improve the model by collecting additional relevant data. One of Kathleen's ideas for improving the accuracy of the model was to collect data on the competitive environment surrounding each store, such as the number of competitor stores nearby. Another was to look at prevailing labor costs. Yet another idea was to incorporate information about whether or not stores were freestanding (which meant they had a drive-thru) or not, and how far they are from the nearest other Milagro store. Kathleen gave her team two weeks to gather the new data—which has now been completed.

The amended data includes six additional variables, shown in Table 2. Due to other pressing project work, Kathleen was not able to incorporate this new data into the model and analysis.

For reference, the 374 stores in the training data (with the added variables) are available in the `train_data.csv` file. The 85 stores in the testing data (again with the added variables) are available in the `test_data.csv` file. Finally, the 48 stores under construction, along with Kathleen's original profitability predictions, are contained in the `site_const_data.csv` file.

Milagro Enters Negotiations with Harriman Capital

Following several months of active shopping, Milagro formally entered negotiations with Harriman Capital in May 2015. Harriman, a Denver-based private equity (PE) firm, owned a variety of companies focusing on consumer goods and services.

Negotiations proceeded smoothly for several months. Harriman reviewed a variety of financial data provided by Milagro, conducted store visits, completed a survey of Milagro customers, and performed a variety of other analyses. They also built financial models to estimate their expected returns on an investment in Milagro. BVA's strategic and growth forecasts were included in these models, as were Harriman's own projections. Harriman seemed particularly interested in the potential for new Milagro stores—including the 48 stores currently under construction, as well as a potential expansion to the east coast.

An Urgent Request from Harriman and Milagro

In a set of emails (attached), Harriman Capital and Julia Silva have asked Stefan Bartell (who has turned this over to Kathleen Ortiz) to do an on-record analysis of the profit predictions for the 48 stores under construction as part of the due diligence process. It turns out that the future profit predictions of the 48 stores under construction have become a point of focus and potential contention in the negotiations between Milagro and Harriman Capital. Julia needs BVA to provide Harriman with back-up and an explanation for the $40 million forecast of the profitability of the 48 stores under construction. The analysis should also include some kind of supplementary material that explains how the model was validated.

In the context of this case you are to play the role of Kathleen Ortiz. You will update the linear regression model, incorporating the new features with some of the old features. The resulting model should be easily explainable and make good sense to parties at Harriman—the fewer variables are used, the better! Then you should use the model to predict the profitably of the 48 stores under construction. You will prepare a memo for Harriman Capital to communicate and support your findings.

Questions

1. Kathleen Ortiz rushed to develop an initial regression model to refresh her memory on the problem and to look ahead into what her results will look like. She typed the following `R` commands:

```
train = read.csv(''train_data.csv'')
test = read.csv(''test_data.csv'')
sites = rbind(train,test)
sites.const = read.csv(''site_const_data.csv'')
model.1 = lm(annual.profit~ . - store.number, data=sites)
summary(model.1)
newpred.1 = predict(model.1,newdata=sites.const)
value <- sum(newpred.1)
R2 <- summary(model.a)$r.squared
```

   Replicate Kathleen's analysis and investigate her initial results. Discuss the strengths and weaknesses of her initial approach. **[10 pts]**

2. Kathleen realizes that the dataset comprises many variables. As an exploratory step, compute the correlation matrix between all the numerical predictors—you can visualize it using the `ggcorrplot()` function in the `ggcorrplot` package in `R`. Report three pairs of variables with strong correlation and provide an intuitive explanation for it. **[10 pts]**

3. You now aim to fit and evaluate three models:
   - Kathleen's original model;
   - the full regression model, using all variables in the expanded dataset; and
   - a model of your choice, which you build manually to improve prediction accuracy (as compared to the original model) and to keep the valuation of the new stores above or around $40 million.

   For each of these models, report the in-sample R^2 and out-of-sample R^2, and compute the projected valuation of the 48 new stores. **[10 pts]**

4. To guide model selection, perform forward stepwise subset selection using cross-validation based on the Mean Squared Error metric. Once you have selected a model, report its coefficients, its in-sample R^2, its out-of-sample R^2 and the projected valuation of the 48 new stores. **[25 pts]**

   Hint: Remember that the Mean Squared Error (MSE) is defined as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2,$$

   where n denotes the number of observations, y_i denotes the outcome of observation $i = 1, \cdots, n$, and $\hat{y}_i$ denotes the prediction for observation $i = 1, \cdots, n$.

   Hint: Unfortunately, there is no built-in function to perform cross-validation along with subset selection. You can proceed as follows:

   a. First, define the `predict.regsubsets()` function. It is needed to make a prediction based on a model obtained through subset selection (with the `regsubsets()` function).

   b. Create 10 folds manually. You may find the following `R` command useful:
   `folds <- sample(1:10,nrow(train),replace=TRUE)`

   c. Loop over the 10 folds. For each one, fit the stepwise selection model on the appropriate training data and, for every number of predictors, calculate the MSE on the appropriate data.

   d. Average the MSE over all the folds, for every number of predictors.

   e. Select your "final" model and proceed.

5. Perform model selection using Lasso. Again, use cross-validation based on the Mean Squared Error metric but, to show robustness, repeat the cross-validation procedures 100 times. Specify how you select your model and report the corresponding value of λ. Report your model's coefficients, its in-sample R^2, its out-of-sample R^2 and the projected valuation of the 48 new stores. **[20 pts]**

   Hint: Pay attention to the dataset on new stores; there are a couple of difficulties arising from the particular form of the data. All your model matrices should have the same set of columns.

6. Outline your key findings in a memo addressed to Harriman, which will also be made available to Julia Silva. The memo should be concise (at most one page). In particular, your memo should provide an updated estimate of the annual profitability of the 48 new stores; describe your analytics approach in easy-to-understand terms; explain in intuitive terms the rationale behind the estimate; and identify the managerial implications of your findings. **[25 pts]**

Table 1: Original variables in the datasets.

Variable	Description
`store.number`	Store identifier number.
`annual.profit`	Net income of the store for one year (in $) (dependent variable).
`state`	US State where store is located.
`sqft`	Size of the restaurant (in square feet).
`intersect`	1 if the restaurant is located at the intersection of two streets, 0 otherwise.
`pop`	Number of people in the census tract.
`agemed`	Median age of the census tract (in years).
`non.us.citizen`	Percent of the census tract who are not US citizens.
`agg.inc`	Total annual income of the census tract, from the Census Bureau (in $).
`med.inc`	Median household income of the census tract, from the Census Bureau (in $).
`noHS`	Percent of people in the census tract who have not completed high-school.
`HS`	Percent of people in the census tract who have completed high school, but not college.
`some.col`	Percent of people in the census tract who have some college, but not a 4 year degree.
`col.grad`	Percent of people in the census tract who have a 4 year, but not a post-grad degree.
`post.grad`	Percent of people in the census tract who have a post-graduate degree.
`com0`	Percent of people in the census tract who commute for less than 15 minutes to work.
`com15`	Percent of people in the census tract who commute for 15–30 minutes to work.
`com30`	Percent of people in the census tract who commute for 30–60 minutes to work.
`com60`	Percent of people in the census tract who commute for over 60 minutes to work.
`drive`	Percent of people in the census tract who drive to work.
`public`	Percent of people in the census tract who take public transit to work.
`walk`	Percent of people in the census tract who walk to work.
`home`	Percent of people in the census tract who work at home.
`other`	Percent of people in the census tract who take any other method to work.

Table 2: New variables in the datasets.

Variable	Description
`lci`	Measure of retail store labor costs provided by a third party vendor.
`nearcomp`	Number of competing fast-casual stores in the census tract.
`nearmil`	Distance to the nearest Milagro (in miles).
`freestand`	1 if the restaurant has its own freestanding building, 0 otherwise.
`gini`	Measure of income inequality in the census tract, provided by the Census Bureau.
`housemed`	Median monthly household housing spend (mortgage or rent) (in $).

Fwd: Profitability of stores under construction

Stefan Bartell [Stefan.Bartell@BVA.com] Sent: Monday, July 27, 2015 7:21 AM
To: Kathleen Ortiz [Kathleen.Ortiz@BVA.com]

Hey Kathleen,
Great work on the TAIBO data-center technology presentation Friday!

I just got this request from Julia Silva at Milagro. Please take care of it. I'm trying to get out the door for vacation with the family.

It's been a long few weeks. I think we are almost to the finish line on this Milagro sale. These store valuations are one of the last sticking points.

I won't have reliable internet access this week. You can reach me until I board the plane at 11:30. Please take a look this morning and let me know if you have any questions. In the meantime, I'll email Rob at Harriman Capital and let him know to expect a memo by the end of the week.

Thanks, Stefan

From: Julia Silva [Julia@Milagro.com]
Sent: Monday, July 27, 2015 4:35 AM
To: Stefan Bartell [Stefan.Bartell@BVA.com] Subject: Profitability of stores under construction

Stefan,

We need you to put together a memo for Harriman Capital on the profitability of the 48 stores under construction.

They're trying to play hardball with the valuation at the last minute, but this dollar figure is the only way we are going to get our current investors to stay with the deal.

A memo that shows new store profitability of at least $40 million along with explanation should be sufficient. The real key is just that it's coming from you guys, not us.

Julia

Re: Profitability of stores under construction

Kathleen Ortiz [Kathleen.Ortiz@BVA.com] Sent: Monday, July 27, 2015 8:08 AM
To: Stefan Bartell [Stefan.Bartell@BVA.com]

Stefan,

Thanks, I'm happy to own this. I've been focused on other things for the last week, so I'm happy to get back to the Milagro work, and to integrate the new data into the model and analysis. Here is what I'm planning to do:

1. 1) Basic exploratory analysis with the new data we've put together.
2. 2) Incorporate the new data to expand and hopefully further improve on the current regression model. I suppose there's a chance the numbers will come down, but presumably not by much.
3. 3) Predict annual profitability for the stores under construction.
4. 4) Write a short memo outlining our case for why profitability is likely to exceed $40M, and explain the rationale for the (updated) model and analysis and why it makes the most sense. I can also attach the model output and predictions in an appendix.

Sound good?
Enjoy your vacation. Kathleen

Memo on new stores

Stefan Bartell [Stefan.Bartell@BVA.com] Sent: Monday, July 27, 2015 8:14 AM
To: Rob@HarrimanCapital.com
Cc: Kathleen Ortiz [Kathleen.Ortiz@BVA.com]

Rob,

Julia Silva just told me you want to see an analysis from us on the projected profitability of the 48 Milagro stores under construction. We are running our regression models, and will put together a memo for. You will get it by Friday, it will be sent by Kathleen from my team (cc'd here).

Stefan

Re: Profitability of stores under construction

Stefan Bartell [Stefan.Bartell@BVA.com] Sent: Monday, July 27, 2015 8:22 AM
To: Kathleen Ortiz [Kathleen.Ortiz@BVA.com]

Kathleen,
This looks great. A couple of notes:

1. 1) Make the takeaways from this memo clear. The projected profitability should be the lead in the memo. Keep it concise with bullets and summaries.
2. 2) Have an appendix/supplement in the back reviewing how we validated the model (OSR^2 and the like) for their analysts to review and be impressed with!
3. 3) Send the memo directly Rob at Harriman by Friday.

Thanks again,

Stefan

Re: Memo on new stores

Rob [Rob@HarrimanCapital.com]
Sent: Monday, July 27, 2015 8:45 AM
To: Stefan Bartell [Stefan.Bartell@BVA.com] Cc: Kathleen Ortiz [Kathleen.Ortiz@BVA.com]

Stefan and Kathleen,

Thank you, please send ASAP. We just want to make sure Silva's projections hold up to our verification standard.

Also, please be sure to include explanation/backup so that we can evaluate and verify the work. Rob

Re: Profitability of stores under construction

Kathleen Ortiz [Kathleen.Ortiz@BVA.com] Sent: Monday, July 27, 2015 8:55 AM
To: Stefan Bartell [Stefan.Bartell@BVA.com]

Thanks, I'm on it. Kathleen

Automatic Reply: Profitability of stores under construction

Stefan Bartell [Stefan.Bartell@BVA.com] Sent: Monday, July 27, 2015 8:56 AM
To: Kathleen Ortiz [Kathleen.Ortiz@BVA.com]

I will be out of the office from Monday, July 27th through Friday, July 31st. I will have only limited access to email during this time. If you need immediate assistance, please contact Wes Sanders (wsanders@BVA.com).

Stefan