

Rasmi Lamichhane

CSCI 3022

Jordan Boyd-Graber

Nov-06- 2016

Training

First of all, while choosing the training data I used 2016 data to test predict, but later on I realized this is not an option. Later, I decided to check my mean square error for 2012 data only. So, I decided to split the data into both train and test. In this process, I split 80%(40 States) of my data to be the train and 20%(10 states) of my data to be test data. While using 2012 data only then I realize that doing the filter was not helping my prediction then, I decided not do the filtering. During this process, I figured a way that I could add data from Republican, MOE, OBS and VALUE to make my data table to look better. This is why I decided to use more data for test data, so I made a matrix of these different data and fit it to get a better prediction.

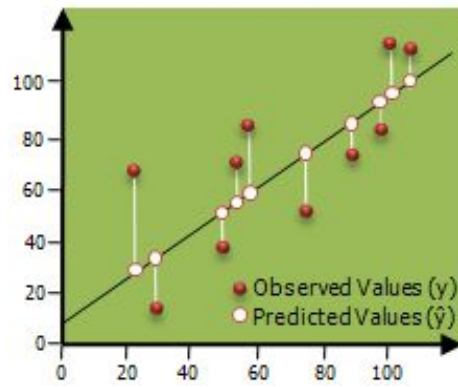
Models

Models is basically a way of going from old data to new data. It's a formula or a equation. While choosing the model, linear regression was given. I wanted to use more data and make the prediction better. For doing this I tried adding different features and during this process, I tried Different models.

a) Linear regression

Ordinary Least Squares(OLS) uses the Linear model of linear regression. According to scikit-learn “ Linear regression fits a linear model with coefficients $w = (w_1, \dots, w_p)$ to minimize the residual sum of squares between the observed responses in the dataset, and the

responses predicted by the linear approximation. Mathematically it solves a problem of the form: $\{\min w\} \{ \|Xw - y\|^2 \}^2$. This is a method for estimating the unknown parameter in linear regression model. This gives sum of the vertical distance between each data point in the set and



the corresponding point on the

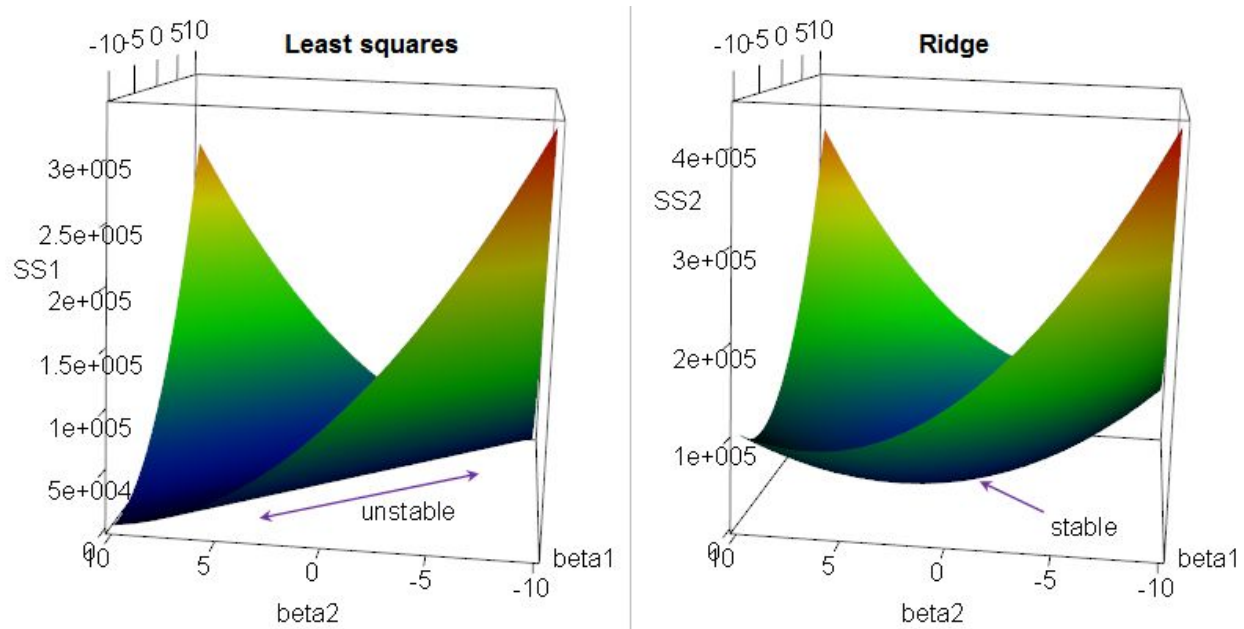
regression line.

The model fits better when the difference is smaller. Whereas if the data is scattered in different place and not close by the line it will be difficult to get the better solution.

b) Ridge regression

Ridge regression is shrinkage methods. According to sikit-learn “Ridge regression addresses some of the problems of Ordinary Least Squares by imposing a penalty on the size of coefficients. The ridge coefficients minimize a penalized residual sum of squares, $\{\min w\} \{ \|Xw - y\|^2 + \alpha \|w\|^2 \}$ Here, $\alpha = 0$ is a complexity parameter that controls the amount of shrinkage: the larger the value of α , the greater the amount of shrinkage and thus the coefficients become more robust to collinearity.” First I used this model just to test and it gave me a better mean square error. Because of the shrinkage happening the graph, it gives the better score. Ridge regression is more stable than least square because of Ridge regression

"fixes" the ridge - it adds a penalty that turns the ridge into a nice peak in likelihood space, equivalently a nice depression in the criterion we're minimizing.



The shrinking can be defined by $\lambda = \text{some value}$. After using this regression I realize having more data is helpful. this is why I added MOE, OBS, value and republican in the test data. I also tried lasso as a another method.

C) Bayesian Ridge regression

I used Bayesian Ridge regression simply thinking its a modified version of ridge regression.

According to scikit-learn “Bayesian regression techniques can be used to include regularization parameters in the estimation procedure: the regularization parameter is not set in a hard sense but tuned to the data at hand. The prior for the parameter w is given by a spherical Gaussian $p(w|\lambda) = N(w | 0, \lambda^{-1} I_p)$ The priors over α and λ are

chosen to be gamma distributions, the conjugate prior for the precision of the Gaussian.”

Bayesian regression mean square error is more than the regular Ridge regression. This is why I choose not to use this model.

D) Polynomial Regression

After doing all of this regression I came to the point of thinking that I need to do something more than just trying out models. While reading about all these different regression model I came to point of figuring out using a model in a polynomial regression. In this process, I decided to use Ridge, Bayesian Ridge in a polynomial regression. This gave me the option of using the different degree. I tried to use degree 2 to 5 and figured that 3 is the best one. This is why I decided to use a polynomial regression with Ridge. This gives the better prediction because the mean square error was less and that is because the data is scattered more likely polynomial functions.

Features

First of all, I was thinking adding different filters can be one of the feature. I reduced some mean square error by adding different filters such as Party not equal to “undecided” and “democrat”. I also used Choice not equal to “obama”, “undecided” and “other”. Somehow the filter was helping to decrease the mean square error and the variance score. later I realize different data would also make the mean square better, this is why I added feature such as MOE, rep, and OBS to data. These feature made my prediction better because it simply gives more testing data and reduce the mean square error. One of the interesting thing that I did was adding Republican data in a ASCII. Since the R is in string I change it to ASCII value of R. For adding multiple features I needed the dynamic matrix. Therefore I passes a matrix with different features.