Predicting the Demand for Taxis in Chicago

Susmita Datta

Project Luther at Metis
April 28, 2017

Predicting Ridership

Investigation:

The influence of time and weather on taxi ridership.

Objective:

Establish a relationship between time, weather and ridership demand.

Predictive Model:

Target: the number of taxi rides

Predictors: weather condition and seasonality

Data Acquisition

- **Weather Data**

Source: Weather Underground

Period: January 2014 – March 2017

Example of predictor variables:

Temperature

Wind speed

Humidity

Conditions such as rain, heavy rain, thunder storm, snow, etc.

Data Acquisition

- **Taxi Data**

Source: City of Chicago

Period: January 2014 – March 2017

Target variable:

Taxi trips

Example of predictor variables:

Time of the day

day or night, noon or mid-night

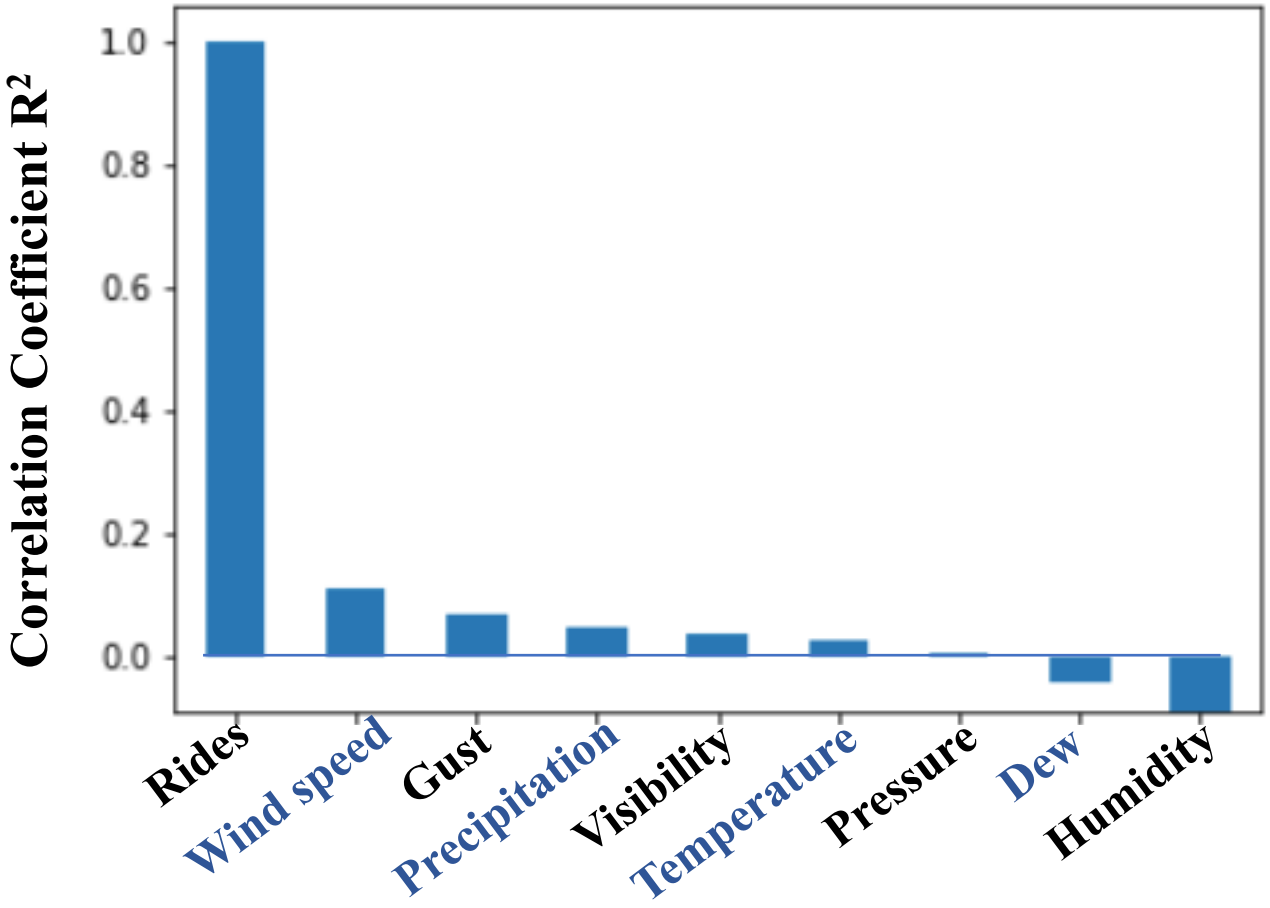
Day of the week:

weekday or weekend

Predictive Modeling

Correlation between number of rides and weather variables

Example: January 2014



Predictive Modeling

Weak linear relationship.

Non-linear relationship between predictors and targets.

Expand feature space

Consider non-linear features in addition to linear features:

$$X \rightarrow [\, x,\ x^{0.5},\ x^{2},\ x^{3} \,]$$

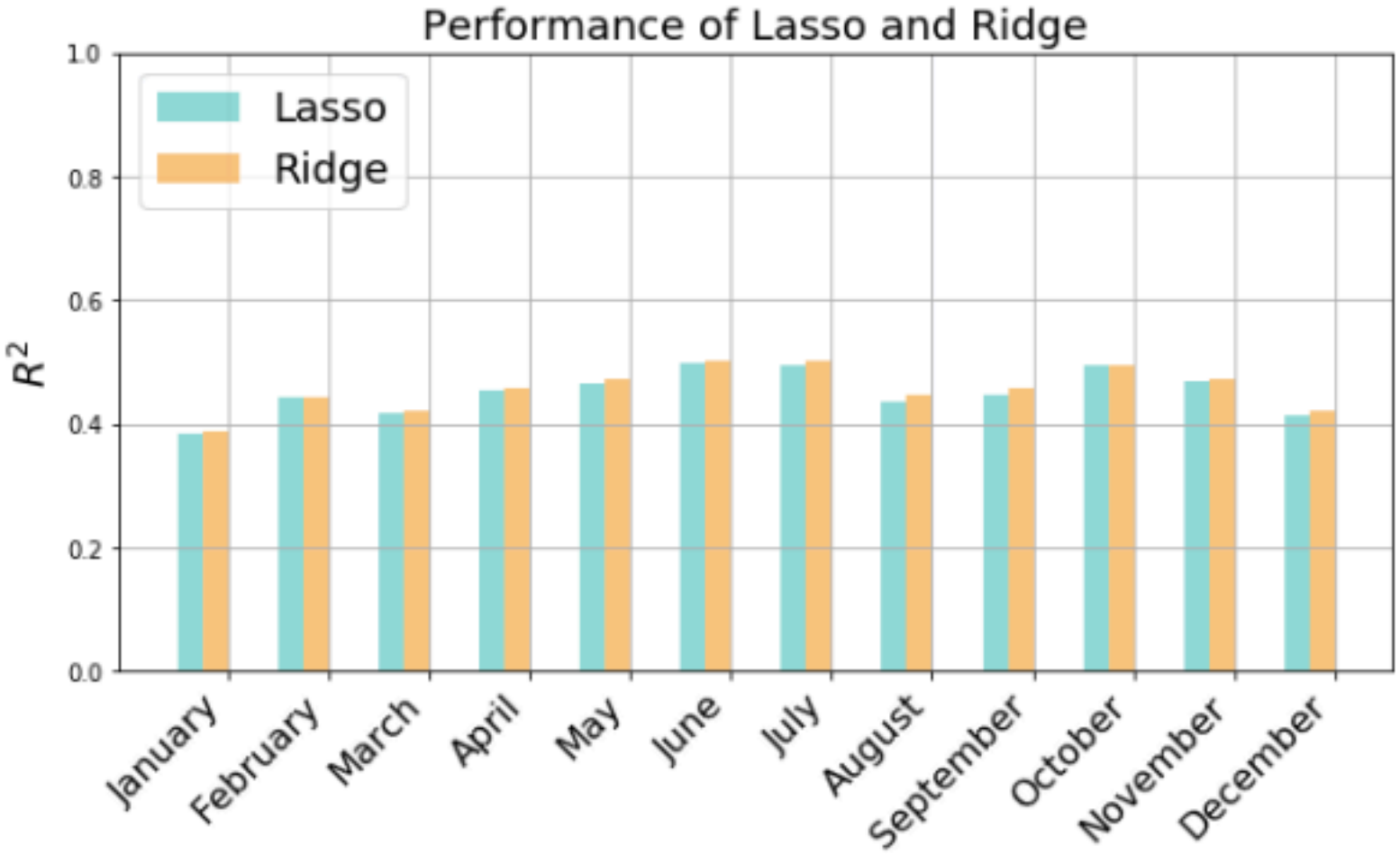
Predictive Modeling

Linear Regression with Regularization

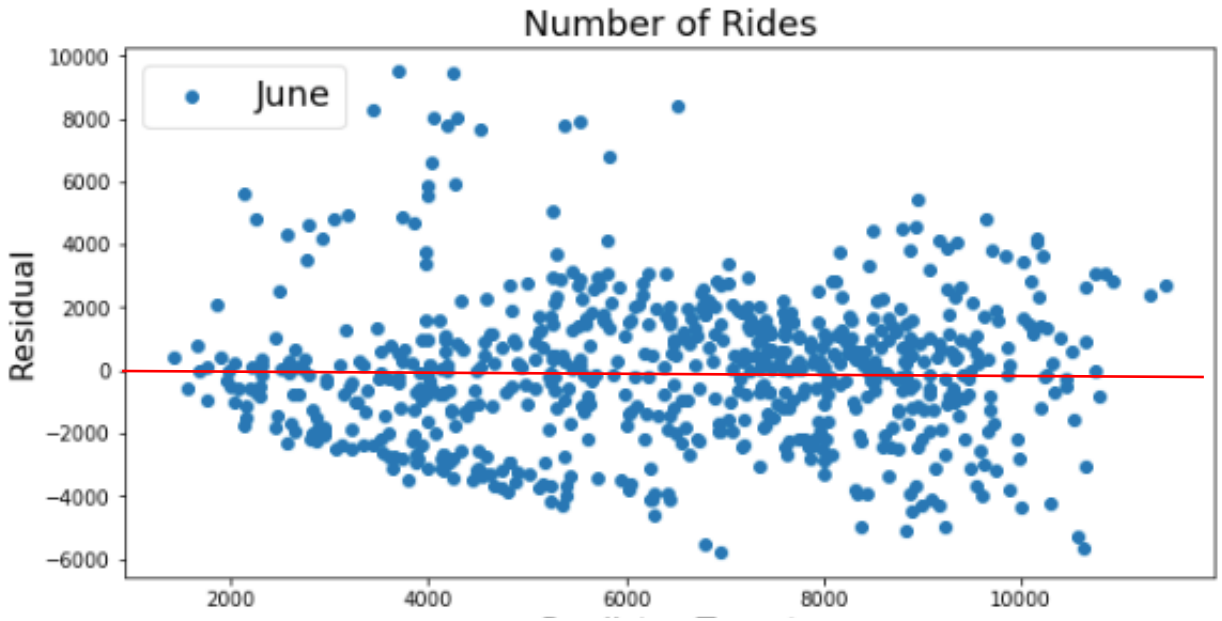
- Lasso Regression (L_1-minimization)
- Ridge Regression (L_2-minimization)

Regularizer prevents us from over representation of features.

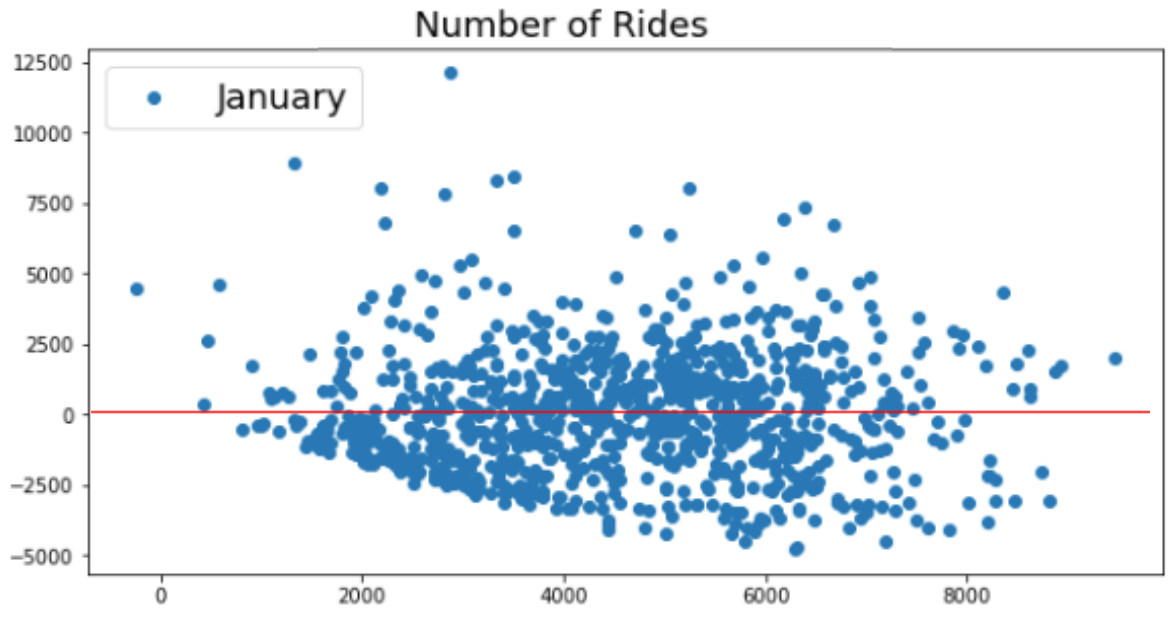
Lasso and Ridge Regression Results



Ridge Regression



Predicted Target



Predicted Target

Remarks

- Predictors capture ~50% of the variation in taxi ridership.
- Grouping data by weather pattern rather than by month only may capture the ridership better.
- Get more data.