Linear regression

Hong Zhang

Brown University

May 21, 2016

We consider the mean of real estate price in two areas in providence, Blackstone area and College area.

Basic linear regression

$$Y = \beta_0 + X\beta,$$

where $\beta = (\beta_1, \ldots, \beta_d)$. We write the formula above in a compact form

$$Y = X_{new}\beta_{new},$$

i.e., $X_{new} = (1, X)$ and $\beta_{new} = (\beta_0, \beta)$. From the least square, we have

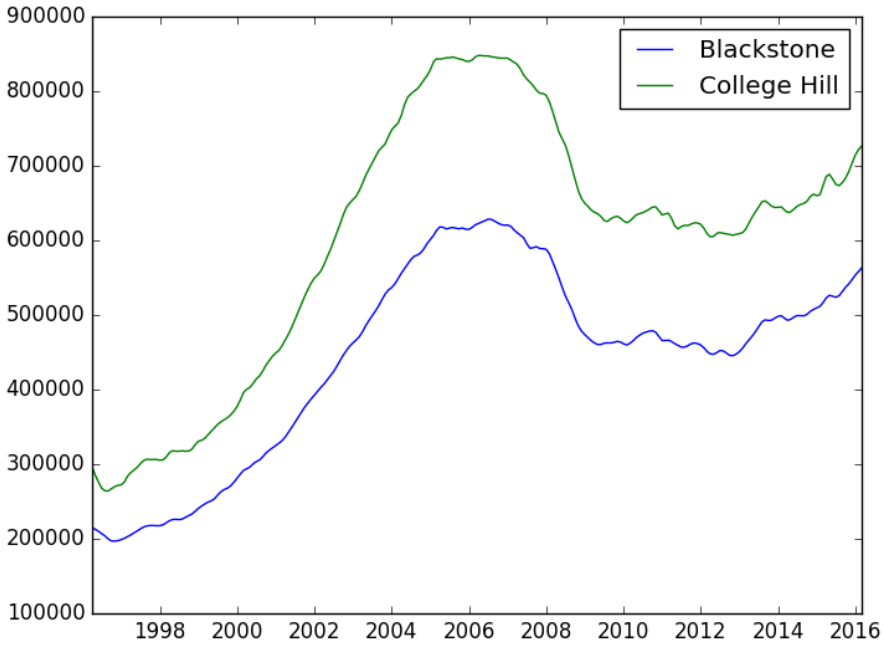
$$\beta_{new} = (X_{new}^t X_{new})^{-1} X_{new}^t Y.$$

In our case, we use the ridge regularization in case that $X_{new}^t X_{new}$ is not invertible. In other words,

$$\beta_{new} = (X_{new}^t X_{new} + \lambda I)^{-1} X_{new}^t Y,$$

where $\lambda \geq 0$ is a parameter.

The following figure is the plot of the mean of the price of real estate in the two regions with respect to time.



Our results show that the r square estimator of the five fold linear regression is 0.9902. The r square estimator from sklearn package is 0.9924. The average prediction error is 19235 dollars.