ML_HW1_mj2776

Mike Jaron

2/2/2017

Q1

a)

$$p(x_1...x_n|\pi_i) = \prod_{i=1}^{n}[\binom{x_i + r - 1}{x_i}\pi^{x_i}(1-\pi)^r]$$

b)

$$\sum_{i=1}^{n} \nabla_\pi \ln[\binom{x_i + r - 1}{x_i}\pi^{x}(1-\pi)^r] = 0$$

$$\sum_{i=1}^{n} \nabla_\pi \ln[\binom{x_i + r - 1}{x_i} + x_i * ln(\pi) + r * ln(1-\pi)] = 0$$

$$\sum_{i=1}^{n}(\frac{x_i}{\pi} - \frac{r}{1-\pi}) = 0$$

$$\nabla_\pi L = \frac{x}{r} + 1 + \frac{a-1}{\pi} - \frac{b-1}{1-\pi} = 0$$

$$\widehat{\pi}_{ML} = \frac{x}{r} + 1$$

c)

$$Prior = p(\pi) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\pi^{a-1}(1-\pi)^{b-1}$$

$$\widehat{\pi}_{MAP} = arg\ max_\pi \ln\ p(\pi|y, X) = arg\ max_\pi\ ln(\frac{p(y|\pi, X)p(\pi)}{p(y|X))})$$

$$\widehat{\pi}_{MAP} = arg\ max_\pi\ ln(p(y|\pi, X)) + ln(p(\pi))$$

$$\frac{b-1}{1-\pi} = \frac{r(a-1) + \pi x + \pi r}{\pi r}$$

$$r(1-a) = \pi(x - \pi x - rb - ra + 3r - \pi r)$$

$$\widehat{\pi}_{MAP} = \frac{a-1}{a+b-2}$$

d)

$$\prod_{i=1}^{n}[\binom{x_i + r - 1}{x_i}\pi^{x_i}(1-\pi)^r] * p(\pi)$$

It is now the beta posterior distribution

e)

$$

$$

Q2

a)

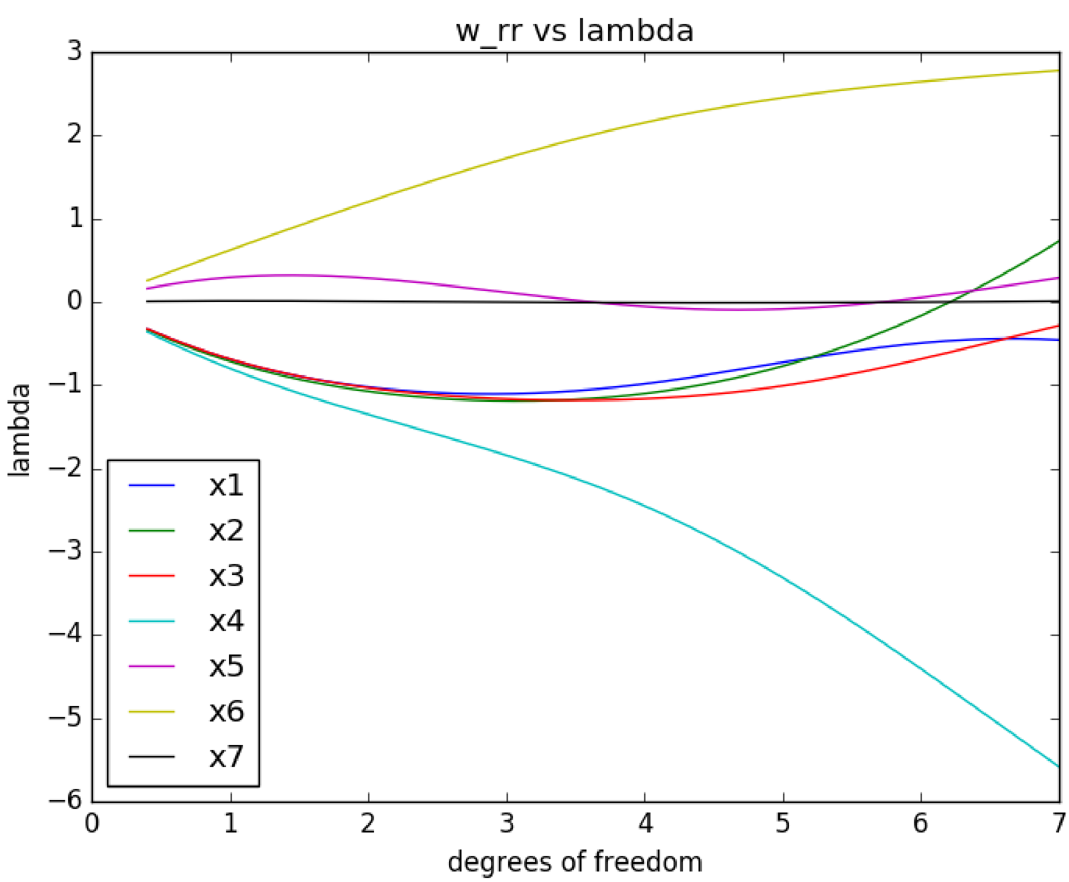

Figure 1:

b)

That the 4th adn 6th dimensions are the most important factors in predicting y.

c)

Well you want to minimize RMSE, and according to my graph that happens as lambda decreases to 0, so Ridge Regression is no better (same accuracy) than Least Squares for this.

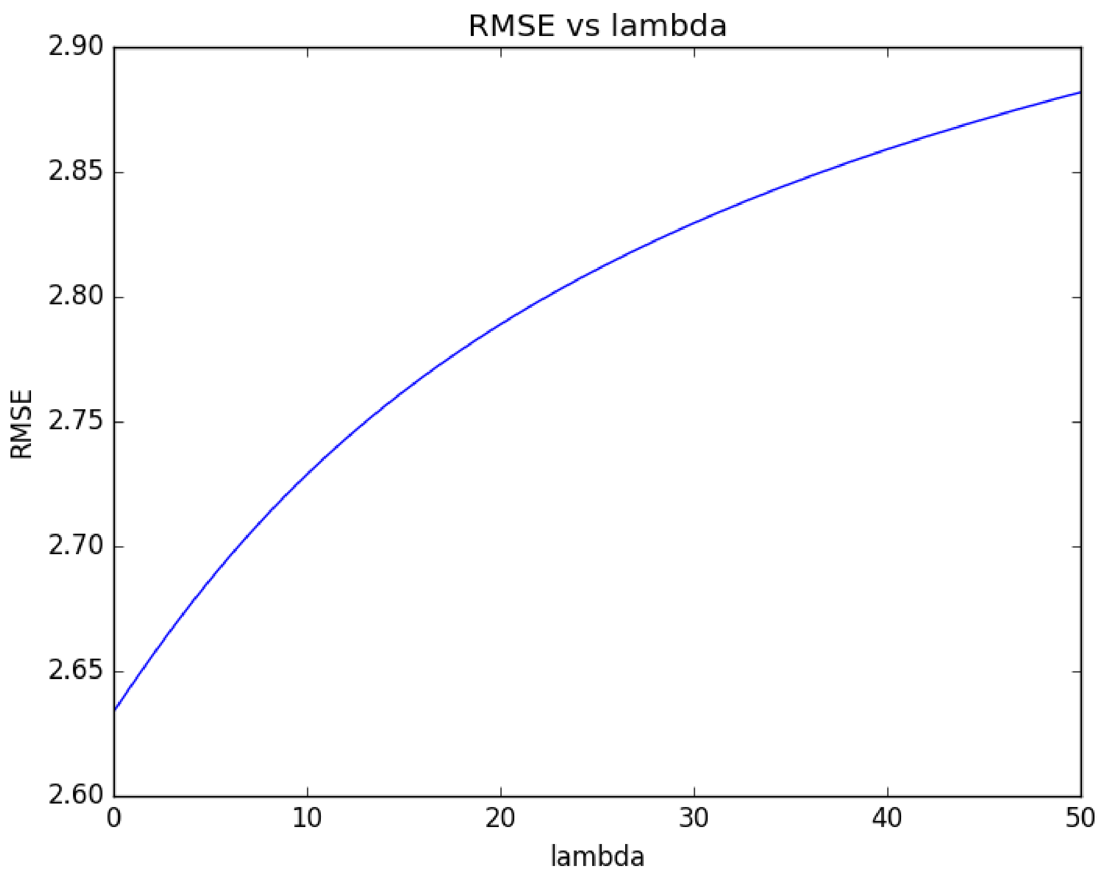

Figure 2:

d)

Again I would want to minimize the RMSE, and the graph shows me that the 2nd and 3rd order have about the same RMSE at their lowest points. The ideal lambda changes for this as it is greater than 0 (roughly ~30), so Ridge Regression is more accurate in this case.

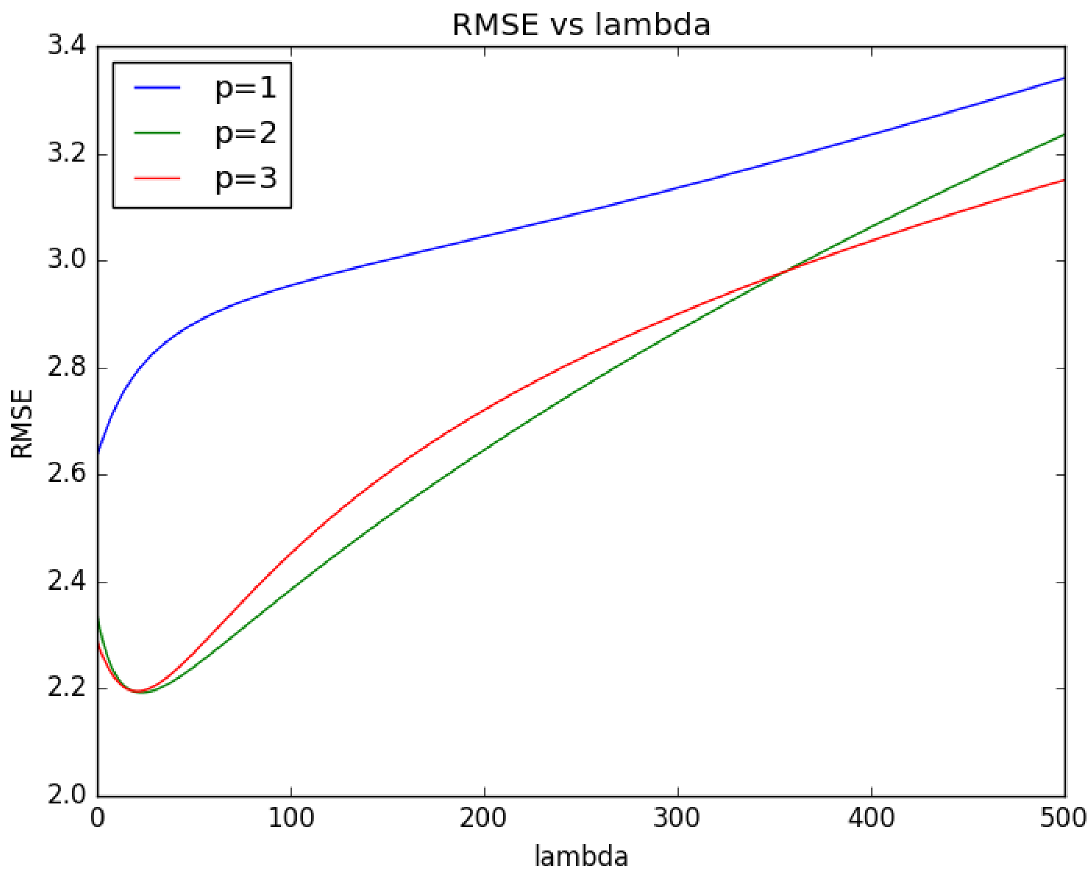
RMSE vs lambda
3.4
3.2
3.0
2.8
2.6
2.4
2.2
2.0
RMSE
0
100
200
300
400
500
lambda
p=1
p=2
p=3

Figure 3: