COMS4721_hw1

jc4609

Problem 1:

(a) $P(x_1, x_2, \cdots x_n \mid \pi, r) = \prod_{i=1}^{N}\left[\binom{x_i+r-1}{x_i}\pi^{x_i}(1-\pi)^r\right]$

$= \prod_{i=1}^{N}\binom{x_i+r-1}{x_i}\pi^{\sum_i^N x_i}(1-\pi)^{Nr}$

(b) $\frac{\partial P(x_1, \cdots x_n \mid \pi, r)}{\partial \pi} = 0$

$-Nr(1-\pi)^{Nr-1} \cdot \hat{\pi}_{ML}^{\sum x_i} + (\sum x_i)\hat{\pi}_{ML}^{\sum x_i - 1} \cdot (1-\hat{\pi}_{ML})^{Nr} = 0$

$\hat{\pi}_{ML} = \frac{\sum x_i}{\sum x_i + Nr}$

(c) $P(\pi \mid x_1, \cdots x_n) \propto \pi^{\sum x_i}(1-\pi)^{Nr} \cdot \pi^{a-1}(1-\pi)^{b-1}$

$\propto \pi^{\sum x_i + a - 1}(1-\pi)^{Nr+b-1}$

So $P(\pi \mid x_1, \cdots x_n) = Beta(\sum x_i + a, Nr + b)$

$\nabla_\pi Beta(\sum x_i + a, Nr + b) = 0 \Rightarrow$

$\hat{\pi}_{MAP} = \frac{\sum x_i + a - 1}{\sum x_i + a + Nr + b}$

(d) as shown in (c):

$P(\pi \mid x_1, \cdots x_n)$ is a Beta distribution.

$\sim Beta(\sum x_i + a, Nr + b)$.

(e).

(e) the mean is $\frac{\sum x_i + a}{\sum x_i + a + Nr + b}$

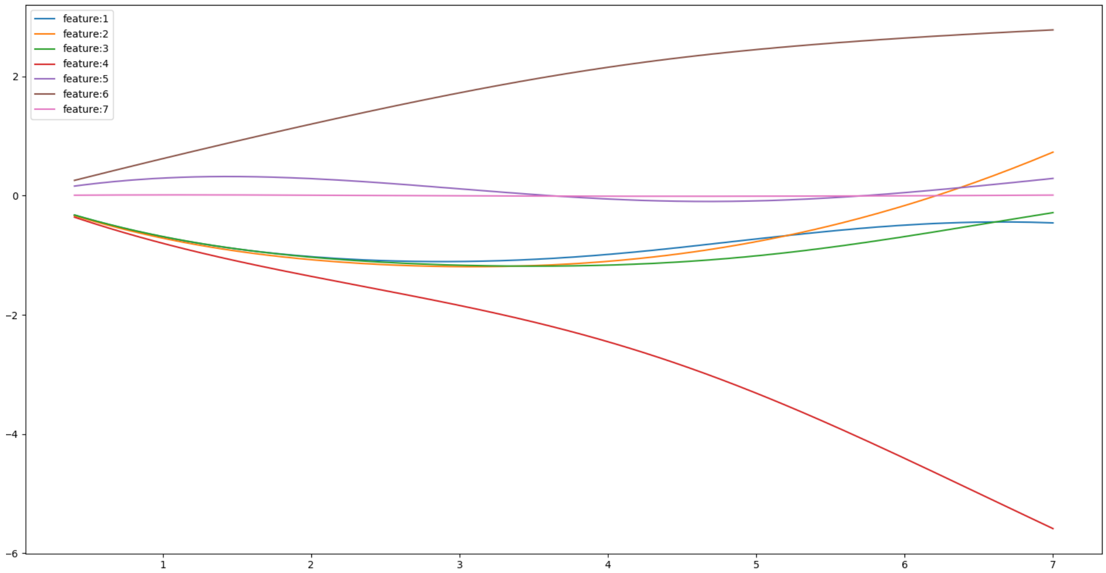
the variance is $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$, $\alpha = \sum x_i + a$, $\beta = Nr + b$

$\hat{\pi}_{MAP}$ is really close to the mean of distribution

And when α and β are all zero. the mean is equals to $\hat{\pi}_{ML}$

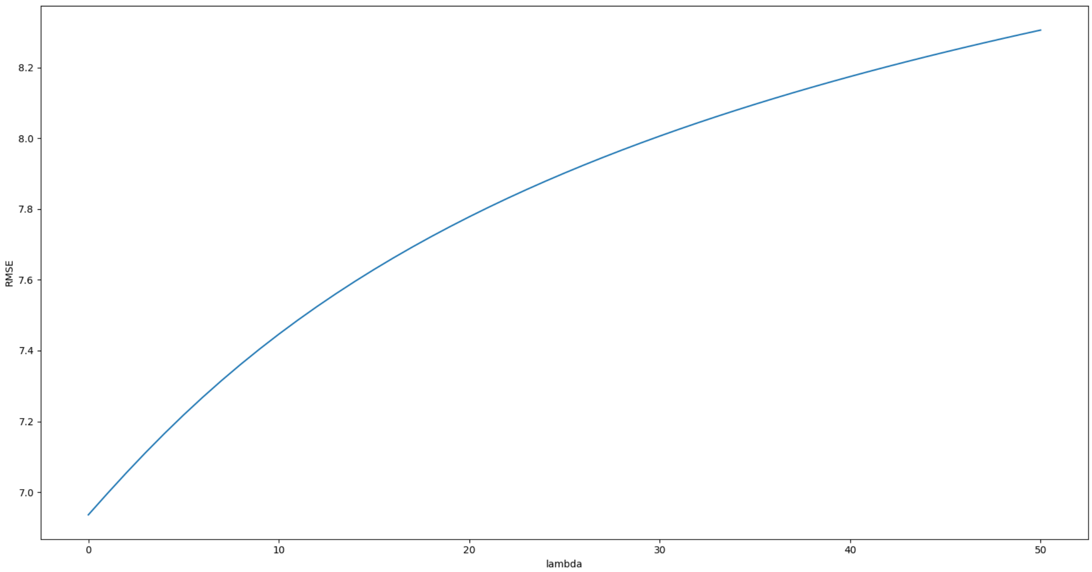
Problem2:

Part 1.(a)



(b)It seems the forth and the sixth dimension have greater influence on "y" than other dimensions. When λ is small , this two dimensions are much more influential than others. And it becomes less obvious when λ increase.

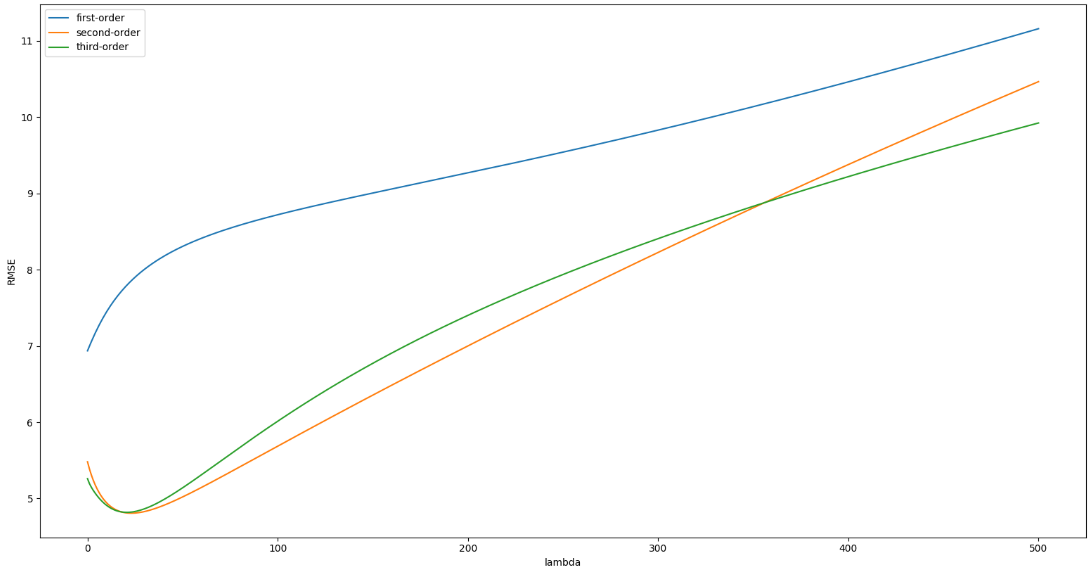
(c).



I would choose λ =0, in other words I would choose least square regression.

Part 2.

(d)



I think p = 2 is better.Because the RMSE of second order and third order is very similar, so I prefer to predict the result with less features since it can save calculation resource.

The primary goal of using λ is preventing overfitting.In the case of first order regression, there is no issue about overfitting. So when λ increase the RMSE just increace. And In the case of second and third order regression, the overfitting may affect our prediction. Thus we introduce a proper λ (about 25) to handle this ,and it works good. However , if λ is too big it will still increase the RMSE.