NAME: Wei Dai   UNI: wd2281

COMS W4721

Problem 1.

(a) because we already known that the sequence is i.i.d

$$\to L(\pi, r) = \prod_{i=1}^{N} P(x_i | \pi, r) = \prod_{i=1}^{N} \binom{x_i+r-1}{x_i} \pi^{\sum_{i=1}^{N} x_i} (1-\pi)^{N\cdot r}$$

(b) $\ell(\pi, r) = \ln L(\pi, r) = \sum_{i=1}^{N} \ln \binom{x_i+r-1}{x_i} + \sum_{i=1}^{N} x_i \ln \pi + N\cdot r \ln(1-\pi)$

$$\frac{\partial \ell(\pi, r)}{\partial \pi} = \frac{\sum_{i=1}^{N} x_i}{\pi} + (-1)\frac{N\cdot r}{1-\pi} = 0$$

$$\sum_{i=1}^{N} x_i (1-\pi) = N r \cdot \pi$$

$$(N\cdot r + \sum_{i=1}^{N} x_i)\pi = \sum_{i=1}^{N} x_i$$

$$\to \pi = \frac{\sum_{i=1}^{N} x_i}{\sum_{i=1}^{N} x_i + N\cdot r} = \frac{\bar{x}}{\bar{x}+r} \quad (\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i)$$

$$\to \hat{\pi}_{ML} = \frac{\bar{x}}{\bar{x}+r}$$

(c). $\hat{\pi}_{MAP} = \underset{\pi}{\arg\max}\ P(\pi | x) = \underset{\pi}{\arg\max} \frac{P(x|\pi)P(\pi)}{P(x)} = \underset{\pi}{\arg\max}\ P(x|\pi)P(\pi)$ ($P(x)$ don't have π)

$$\to \hat{\pi}_{map} = \underset{\pi}{\arg\max} \prod_{i=1}^{N} P(x_i|\pi)\, p(\pi)$$

$$= \underset{\pi}{\arg\max} \sum_{i=1}^{N}\left(\ln P(x_i|\pi)\right) + \ln p(\pi)$$

$$\to \frac{\partial}{\partial \pi} \sum_{i=1}^{N}\left(\ln P(x_i|\pi)\right) + \ln p(\pi) = 0$$

$$\frac{\partial}{\partial \pi}\left(\sum_{i=1}^{N} \ln \binom{x_i+r-1}{x_i} + \sum_{i=1}^{N} x_i \cdot \ln \pi + N\cdot r \ln(1-\pi) + \ln \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \pi^{a-1}(1-\pi)^{b-1}\right) = 0$$

$$\frac{\sum_{i=1}^{N} x_i}{\pi} - N\cdot r \frac{1}{1-\pi} + (a-1)\frac{1}{\pi} - (b-1)\frac{1}{1-\pi} = 0$$

$$\frac{1}{\pi}\left(\sum_{i=1}^{N} x_i + a - 1\right) = \frac{1}{1-\pi}(Nr + b - 1)$$

$$\to \hat{\pi}_{MAP} = \frac{\sum_{i=1}^{N} x_i + a - 1}{\sum_{i=1}^{N} x_i + Nr + a + b - 2} = \frac{\bar{x} + \frac{a-1}{N}}{\bar{x} + r + \frac{a+b-2}{N}}$$

(d) $$P(\pi | x) = \frac{\prod_{i=1}^{N} P(x_i | \pi)\, P(\pi)}{\int_0^1 \prod_{i=1}^{N} P(x_i | \pi)\, P(\pi) d\pi}$$

$$P(\pi | x) \propto \prod_{i=1}^{N} \left[\binom{x_i + r - 1}{x_i} \pi^{x_i} (1-\pi)^r \right] \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \pi^{a-1} (1-\pi)^{b-1}$$

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

∴ → $P(\pi | x) \sim Beta\left(\sum_{i=1}^{N} x_i + a, Nr + b\right)$

(e). $$E(\pi | x) = \frac{\sum_{i=1}^{N} x_i + a}{\sum_{i=1}^{N} x_i + Nr + a + b}$$

$$Var(\pi | x) = \frac{\left(\sum_{i=1}^{N} x_i + a\right)(Nr + b)}{\left(\sum_{i=1}^{N} x_i + Nr + a + b\right)^2 \left(\sum_{i=1}^{N} x_i + Nr + a + b + 1\right)}$$

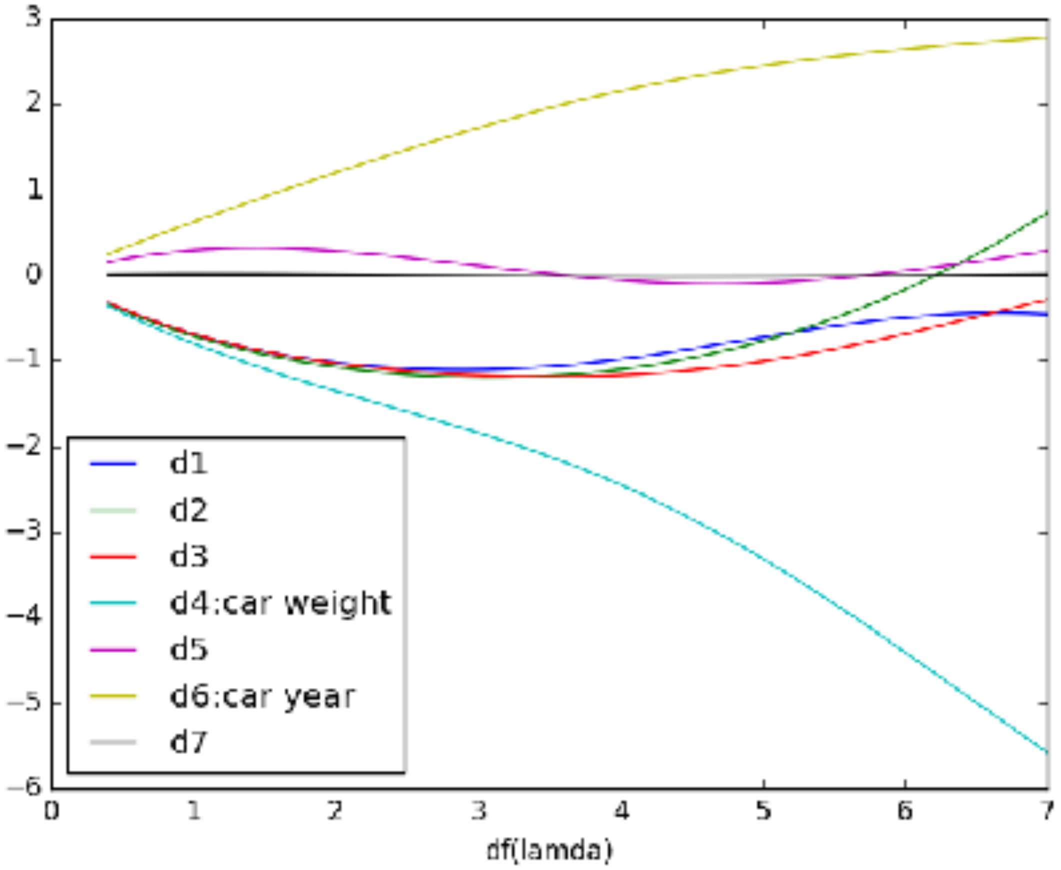
Δ(if $x \sim Beta(\alpha, \beta)$, $E(x) = \frac{\alpha}{\alpha+\beta}$, $Var(x) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$)

∴ $\hat{\pi}_{ML}$ is the mean of $Beta\left(\sum_{i=1}^{N} x_i, Nr\right)$ $(a' = 0, b' = 0)$

$\hat{\pi}_{MAP}$ is the mean of $Beta\left(\sum_{i=1}^{N} x_i + a - 1, Nr + b - 1\right)$ $(a' = a - 1, b' = b - 1)$

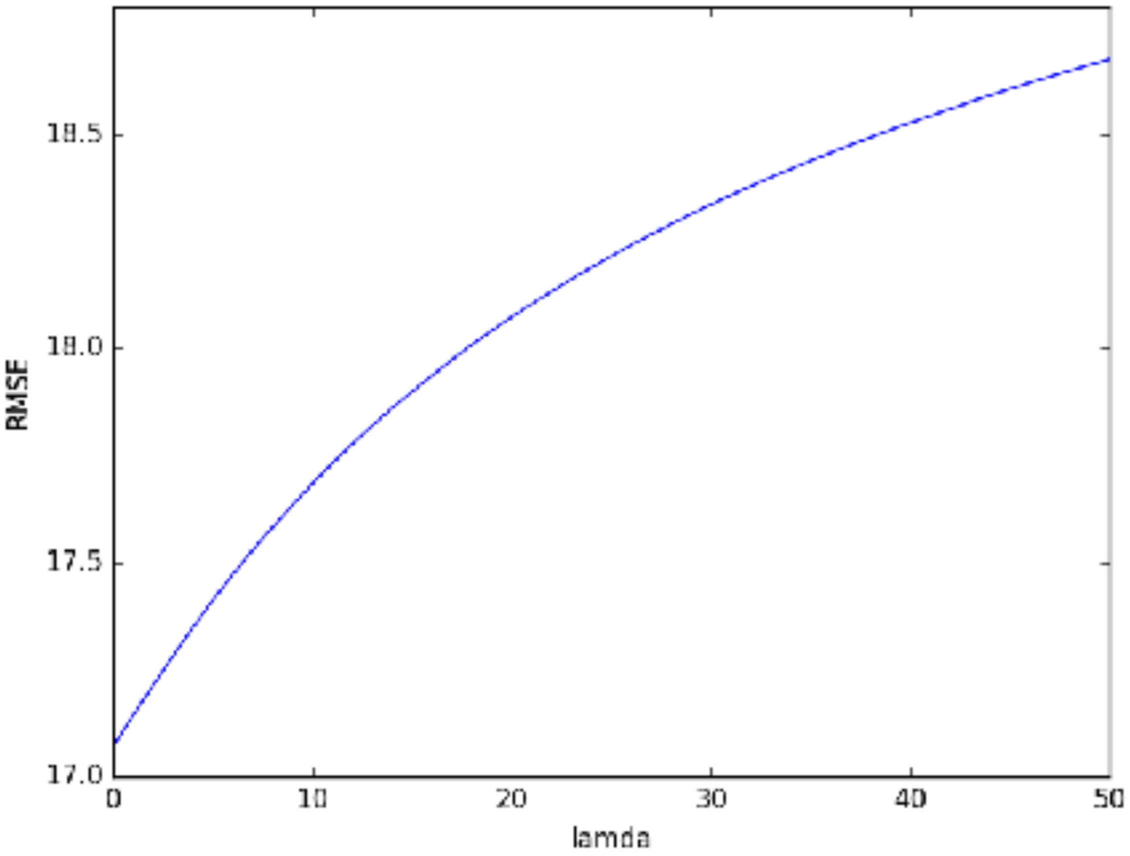
Problem 2:

(a) : The seven values in w as a function of df(lamda) is as follow:
(di denote dimension i)



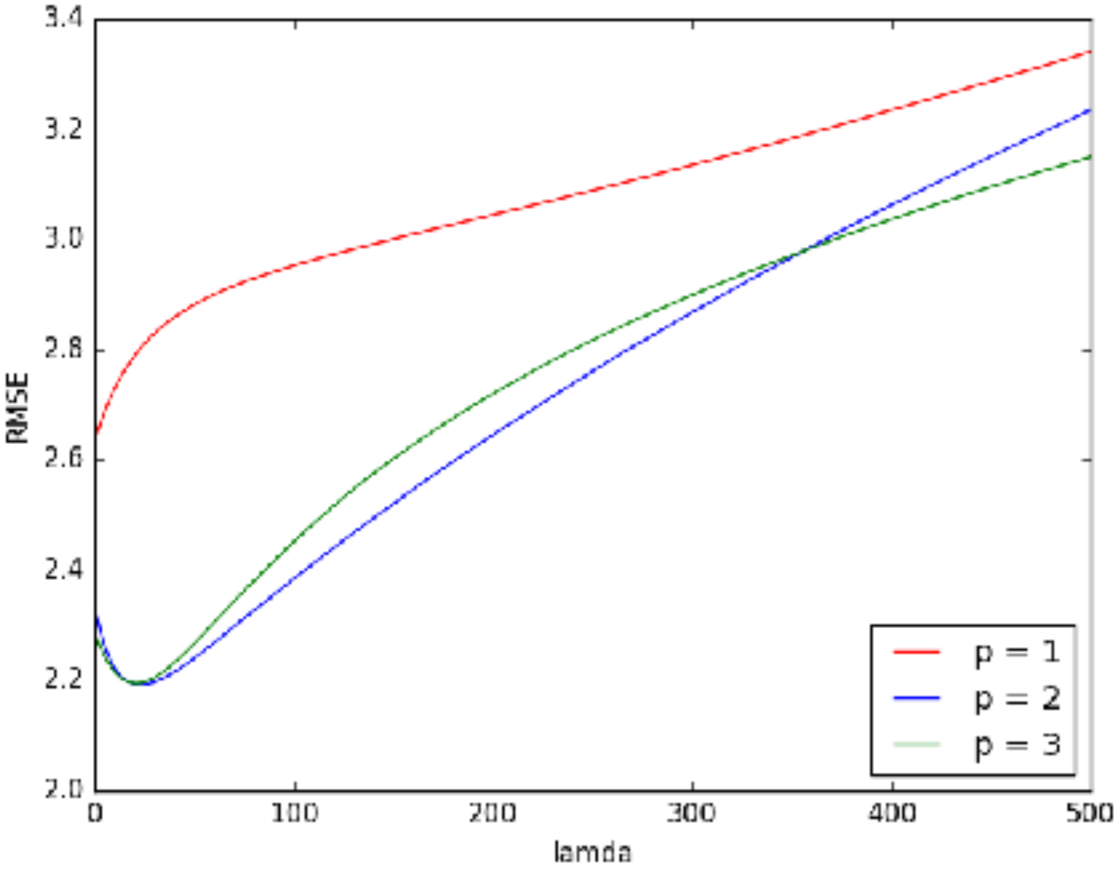
(b) Because these two variables can explain more variation in response than other variables when constraining the l2 norm of beta. On the other hand, these two coefficients is larger than others when lambda equal to zero. That means the changing of these two variables have larger influence on the value of y when we use linear regression to fit the model.

(c): The figure tell us that the RMSE is a monotonic increasing sequence as a function of lambda. The value that the bias will increase is larger than the value that the variance will decrease. So under this circumstance, we should use least square method instead of the ridge regression, considering the trade-of between bias-variance. We should choose lambda = 0 in this problem.



(d) We should choose the lambda and p which minimize the RMSE. We could find the location the the minimum in the figure is (23, 2.192574120864215) with p = 2. So we should choose lambda = 23 and p = 2 in this problem. My assessment of the ideal value of lambda change because the model could not be regression under p = 1 (like what we do in question (c)). The bias is too high when we choose p = 1. We get closer to the true model when we set p = 2. And we can tell from the figure that lambda = 23 is the best choice. So that is why we choose p = 2 and lambda = 23.

(If we could using a float number as lambda, we could use the numerical method to calculate the point where it minimum of the function. It should be near the integer 23.)


RMSE
lamda
p = 1
p = 2
p = 3