Problem 1

a) joint likelihood:

$$p(x_1 x_2 \ldots x_N \mid \pi, r) = \prod_{i=1}^{N}\left(\binom{x_i + r - 1}{x_i}\pi^{x_i}(1-\pi)^r\right)$$

b) Derive the maximum likelihood estimate for π :

$$\log\, p(x_1 x_2 \ldots x_N \mid \pi, r) = \sum_{i=1}^{N}\log\left(\binom{x_i + r - 1}{x_i}\pi^{x_i}(1-\pi)^r\right)$$

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

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

$$\frac{\partial \log p(x_1 x_2 \ldots x_N \mid \pi, r)}{\partial \pi} = \sum_{i=1}^{N} x_i \bullet \frac{1}{\pi} - \frac{Nr}{1-\pi}$$

Set $\frac{\partial \log p(x_1 x_2 \ldots x_N \mid \pi, r)}{\partial \pi} = 0 \;=>\; \sum_{i=1}^{N} x_i \bullet \frac{1}{\pi} - \frac{Nr}{1-\pi} = 0 \;=> \pi = \frac{\sum_{i=1}^{N} x_i}{\sum_{i=1}^{N} x_i + Nr}$

$$\therefore \hat{\pi}_{ML} = \frac{\sum_{i=1}^{N} x_i}{\sum_{i=1}^{N} x_i + Nr}$$

c) Derive MAP for π :

$$p(\pi \mid x_1 x_2 \ldots x_N) = \frac{p(x_1 x_2 \ldots x_N \mid \pi) \bullet p(\pi)}{p(x_1 x_2 \ldots x_N)} = \frac{\prod_{i=1}^{N}\left(\binom{x_i + r - 1}{x_i}\pi^{x_i}(1-\pi)^r\right) \bullet \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\pi^{a-1}(1-\pi)^{b-1}}{\int_0^1 p(x_1 x_2 \ldots x_N \mid \pi) \bullet p(\pi) d\pi}$$

The denominator is irrelevant of π , so

$$p(\pi \mid x_1 x_2 \ldots x_N) \propto \prod_{i=1}^{N}\left(\binom{x_i + r - 1}{x_i}\pi^{x_i}(1-\pi)^r\right) \bullet \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\pi^{a-1}(1-\pi)^{b-1}$$

Let

$$l = \log\left[\prod_{i=1}^{N}\left(\binom{x_i + r - 1}{x_i}\pi^{x_i}(1-\pi)^r\right) \bullet \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\pi^{a-1}(1-\pi)^{b-1}\right]$$

$$= \sum_{i=1}^{N}\log\binom{x_i + r - 1}{x_i} + \log\pi\sum_{i=1}^{N} x_i + Nr\log(1-\pi) + \log\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} + (a-1)\log\pi + (b-1)\log(1-\pi)$$

Set $\frac{\partial l}{\partial \pi} = 0$ => $\sum_{i=1}^{N} x_i \bullet \frac{1}{\pi} - \frac{Nr}{1-\pi} + \frac{a-1}{\pi} - \frac{b-1}{1-\pi} = 0$ => $\pi = \frac{\sum_{i=1}^{N} x_i + a - 1}{\sum_{i=1}^{N} x_i + Nr + a + b - 2}$

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

d) derive the posterior distribution of π :

$$p(\pi \mid x_1 x_2 \dots x_N) = \frac{p(x_1 x_2 \dots x_N \mid \pi) \bullet p(\pi)}{\int_0^1 p(x_1 x_2 \dots x_N \mid \pi) \bullet p(\pi) d\pi} = \frac{\prod_{i=1}^{N} \left(\begin{pmatrix} x_i + r - 1 \\ x_i \end{pmatrix} \pi^{x_i} (1-\pi)^r \right) \bullet \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \pi^{a-1} (1-\pi)^{b-1}}{p(x_1 x_2 \dots x_N)}$$

$$\propto \prod_{i=1}^{N} \left(\begin{pmatrix} x_i + r - 1 \\ x_i \end{pmatrix} \pi^{x_i} (1-\pi)^r \right) \bullet \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \pi^{a-1} (1-\pi)^{b-1}$$

$$\propto \prod_{i=1}^{N} \left(\pi^{x_i} (1-\pi)^r \right) \bullet \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \pi^{a-1} (1-\pi)^{b-1}$$

$$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \pi^{\sum_1^N x_i + a - 1} (1-\pi)^{Nr+b-1} \sim \mathrm{Beta}(\sum_{i=1}^{N} x_i + a, Nr + b)$$

e) the mean and variance of π_{MAP} :

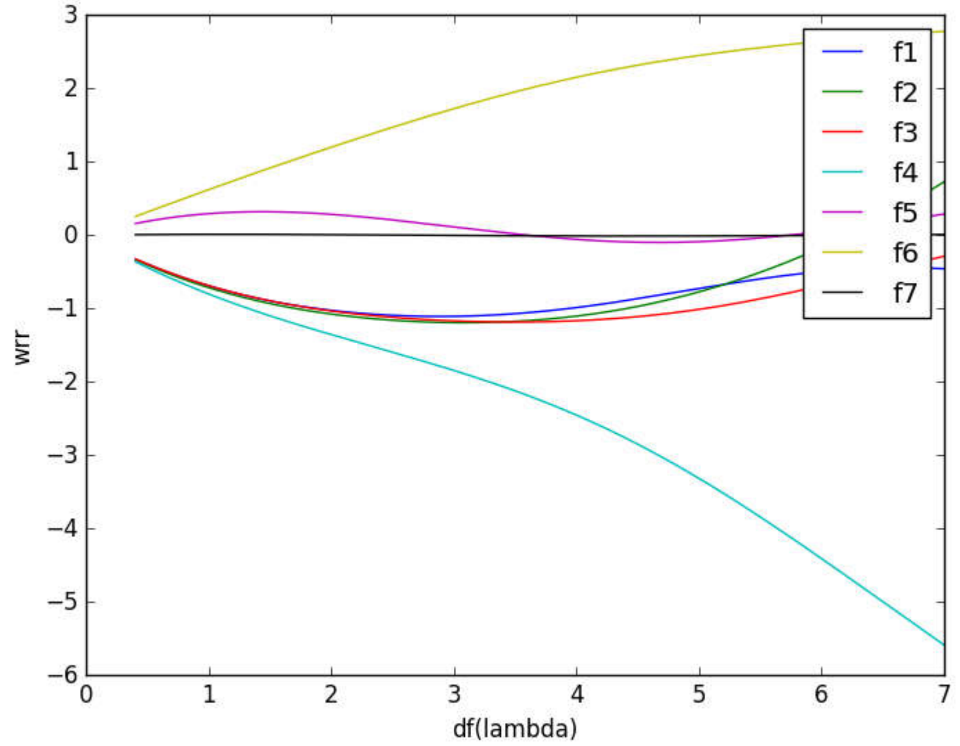
$$\because p(\pi \mid x_1 x_2 \dots x_N) \sim \mathrm{Beta}(\sum_{i=1}^{N} x_i + a, Nr + b)$$

$$\therefore E[\pi_{MAP}] = \frac{\sum_{i=1}^{N} x_i + a}{\sum_{i=1}^{N} x_i + a + Nr + b}, Var[\pi_{MAP}] = \frac{\left(\sum_{i=1}^{N} x_i + a \right) (Nr + b)}{\left(\sum_{i=1}^{N} x_i + a + Nr + b \right)^2 \left(\sum_{i=1}^{N} x_i + a + Nr + b + 1 \right)}$$

The mean of π_{MAP} is very similar to $\hat{\pi}_{MAP}$. $\hat{\pi}_{MAP}$ depends not only on given data, but also on the prior distribution of π. While as for $\hat{\pi}_{ML}$, it only depends on the given data.

Problem 2

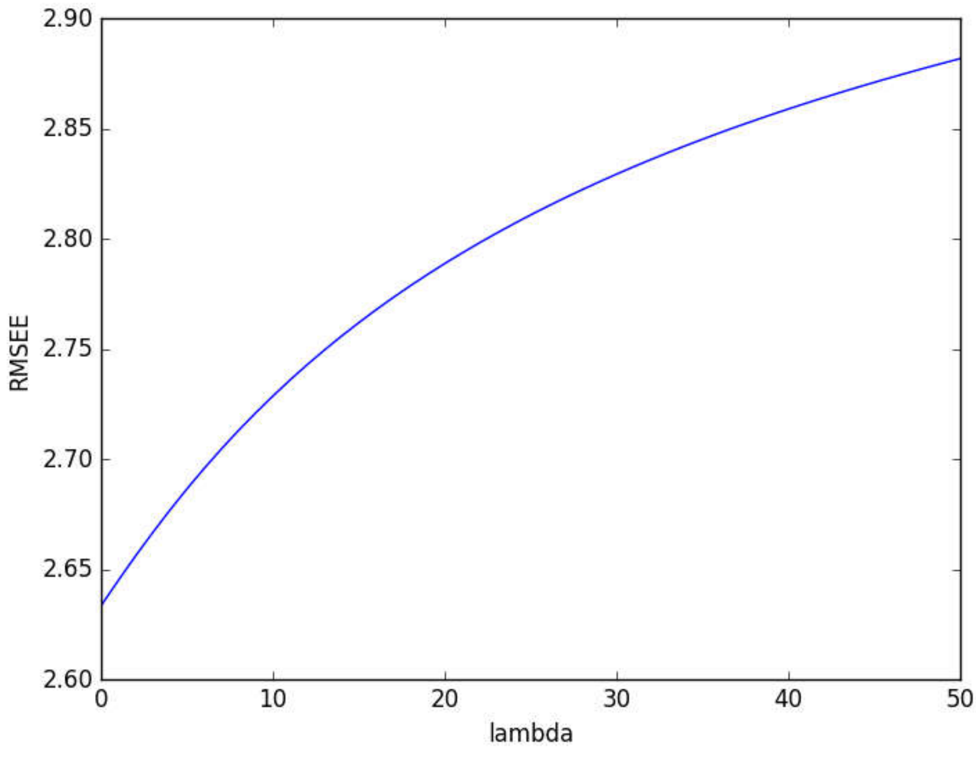
a)



b)

Feature 4 and Feature 6 have a greater impact on the predictive value y, for they have larger absolute values.

c)



RMSE increases when λ become larger, so we should choose $\lambda = 0$. That's to calculate w_{LS}