Assignment - 1

COMS 4721

- VIBHUTI MAHAJAN

vm2486.

Q1.

(a) joint likelihood function: (under iid assumption)

$$f_{x_1,x_2\ldots x_n}(x_1,x_2\ldots,x_n) = \prod_{i=1}^{n} p(x_i|\pi, r)$$

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

$$= \pi^{\sum_{i=1}^{n} x_i} (1-\pi)^{rn} \prod_{i=1}^{n} \binom{x_i - r - 1}{x_i}$$

$x_i \in \{0,1,\ldots\}$

(b) log-likelihood function:-

$$\log f(x) = \ell = \sum_{i=1}^{n} x_i \log \pi + rn \log(1-\pi) + \sum_{i=1}^{n} \log \binom{x_i - r - 1}{x_i}$$

$$\frac{\partial \ell}{\partial \pi} = \frac{\sum_{i=1}^{n} x_i}{\pi} - \frac{rn}{1-\pi} = 0 \longrightarrow \hat{\pi} = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i + rn}$$

$$\left.\frac{\partial^2 \ell}{\partial \pi^2}\right|_{\hat{\pi}} = -\frac{\sum x_i}{\pi^2} - \frac{rn}{(1-\pi)^2} < 0 \quad \forall \pi$$; hence $\hat{\pi}$ is the maxima.

$$\Rightarrow \hat{\pi}_{ML} = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i + rn} \; ; \; x_i \in \{0,1,2,\ldots\}$$

(c) $P(\pi) = \text{Beta}(a,b)$

$$\pi_{MAP} = \underset{\pi}{\text{argmax}} \log p(\pi | X, r)$$

$$= \underset{\pi}{\text{argmax}} \log \frac{P(X|\pi, r)\, P(\pi)}{P(X)} \; ; \; [X = (x_1, x_2, \ldots x_n)]$$

$$= \underset{\pi}{\text{argmax}} \ln p(X|\pi, r) + P(\pi) - P(X)$$

which is equivalent to: $\arg\max_{\pi} \ln p(x|\pi, r) + \ln p(\pi) \longrightarrow ⓐ$

$$ⓐ = \ln p(x|\pi, r) + \ln p(\pi) = \ln\left[\prod_{i=1}^{n} \binom{x_i + r - 1}{x_i} \pi^{x_i} (1-\pi)^{r}\right]$$

$$+ \ln\left(\frac{\pi^{a-1}(1-\pi)^{b-1}}{\beta(a,b)}\right); \pi \in [0,1]$$

$$= \sum_{i=1}^{n} \ln\binom{x_i + r - 1}{x_i} + (\ln \pi).\left(\sum_{i=1}^{n} x_i + a - 1\right)$$

$$+ (rn + b - 1).\ \ln(1-\pi) - \ln \beta(a,b)$$

Maximising this expression w.r.t. π:

$$\frac{dⓐ}{d\pi} = \left(\sum_{i=1}^{n} x_i + a - 1\right).\frac{1}{\pi} - (rn + b - 1)\left(\frac{1}{1-\pi}\right) = 0$$

$$\Rightarrow \hat{\pi}_{MAP} = \frac{\sum_{i=1}^{n} x_i + a - 1}{\sum_{i=1}^{n} x_i + a + rn + b - 2}$$

[Note that $\frac{dⓐ^2}{d\pi^2} < 0$ for $\pi = \hat{\pi}_{MAP}$ by previous parts analogy]

Hence $\hat{\pi}_{MAP}$ is point of maxima

(d) $$p(\pi|x, r) = \frac{p(x|\pi, r).\ p(\pi)}{p(x)}$$

$$p(x|\pi, r).\ p(\pi) = \frac{\pi^{\sum_{i=1}^{n} x_i + a - 1} (1-\pi)^{rn + b - 1} \prod_{j=1}^{n} \binom{x_i + r - 1}{x_i}}{\beta(a,b)}$$

$$p(x) = \int_{\pi} p(x|\pi, r).p(\pi) d\pi = \int_{0}^{1} p(x|\pi, r).\ p(\pi).\ d\pi$$

After some algebraic manipulation: $= \int_{0}^{1}$

$$\cancel{= \pi^{\sum x_i}}$$

$$= \prod_{i=1}^{n} \binom{x_i + r - 1}{x_i} \frac{\beta\left(\sum_{i=1}^{n} x_i + a, rn + b\right)}{\beta(a,b)} \int_0^1 \frac{\pi^{\sum_{i=1}^{n} x_i + a - 1} (1-\pi)^{rn+b-1}}{\beta\left(\sum_{i=1}^{n} x_i + a, rn+b\right)} d\pi$$

$$= \prod_{i=1}^{n} \binom{x_i + r - 1}{x_i} \frac{\beta\left(\sum_{i=1}^{n} x_i + a, rn + b\right)}{\beta(a,b)}$$

$$\Rightarrow P(\pi | x, r) = \frac{\pi^{\sum_{i=1}^{n} x_i + a - 1} (1-\pi)^{rn+b-1}}{\beta\left(\sum_{i=1}^{n} x_i + a, rn+b\right)} \cdot \frac{\cancel{\prod_{i=1}^{n} \binom{x_i + r - 1}{x_i} \frac{\beta\left(\sum_{i=1}^{n} x_i + a, rn + b\right)}{\beta(a,b)}}}{\cancel{\prod_{i=1}^{n} \binom{x_i + r - 1}{x_i} \frac{\beta\left(\sum_{i=1}^{n} x_i + a, rn + b\right)}{\beta(a,b)}}}$$

for $\pi \in [0,1]$

hence $P(\pi | x, r) \sim \text{Beta}\left(\sum_{i=1}^{n} x_i + a, rn + b\right)$; $x_i \in \{0, 1, 2, \dots\}$

(d) $$E[\pi | x, r] = \frac{\sum_{i=1}^{n} x_i + a}{\sum_{i=1}^{n} x_i + a + rn + b}$$

$$Var[\pi | x, r] = \frac{\left(\sum_{i=1}^{n} x_i + a\right)(rn+b)}{\left(\sum_{i=1}^{n} x_i + a + rn + b\right)^2 \left(\sum_{i=1}^{n} x_i + a + rn + b + 1\right)}$$

□ Note $\hat{\pi}_{ML} = \hat{\pi}_{MAP} = \dfrac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i + rn}$ if $a = 1$ & $b = 1$

i.e if π has a diffused prior [uniform (0,1)] then $\hat{\pi}_{ML} = \hat{\pi}_{MAP}$

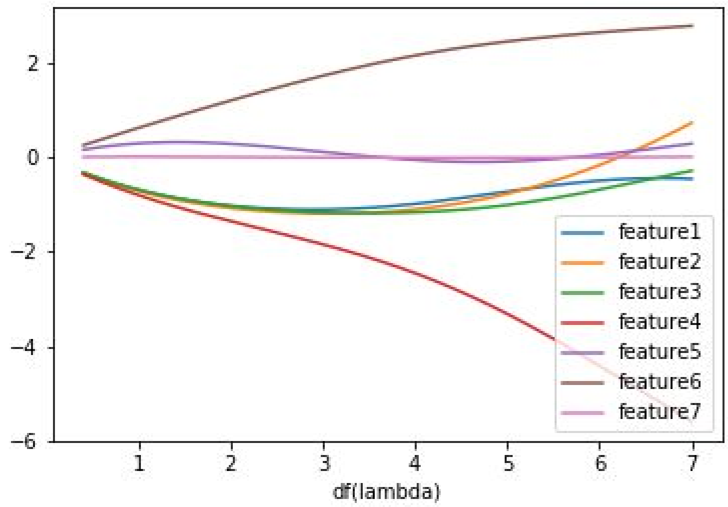
□ $\hat{\pi}_{ML} = E[\pi | x, r] = \dfrac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i + rn}$ if $a = 0$ & $b = 0$

i.e $P(\pi) \sim \text{Beta}(0,0) \Rightarrow \pi$ has a non-informative prior.

▫ $$\text{Mode}\left[\pi \mid x, r\right] = \hat{\pi}_{MAP} = \frac{\sum_{i=1}^{n} x_i + a - 1}{\sum_{i=1}^{n} x_i + a + rn + b - 2}$$

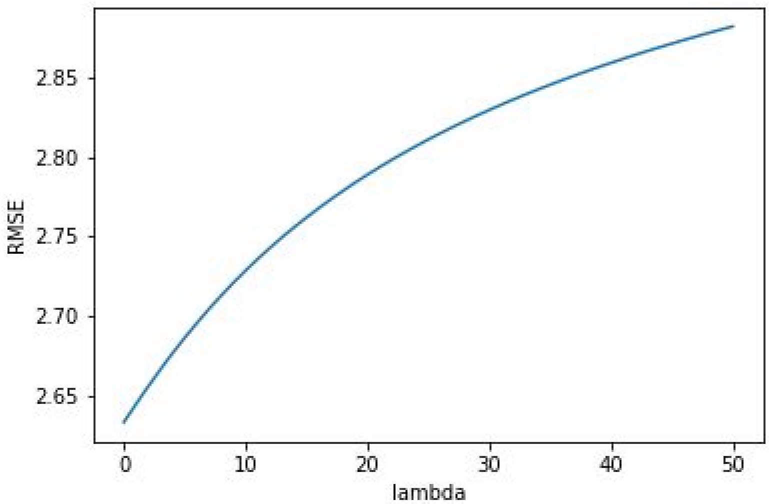
Problem 2:

A. ω_r v/s $df(\lambda)$



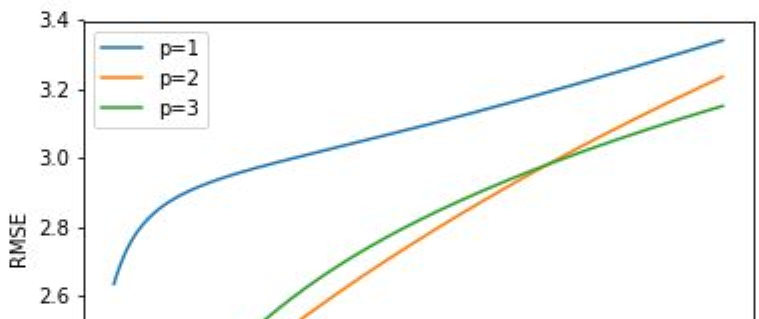
B. We are modelling the miles per gallon a car will get using six quantities (features) about that car.
The figure shows that the magnitude for feature 4(car weight) and feature 6(car year) are greater as compared to other features implying the higher correlation between miles per gallon used and these 2 quantities.
Feature 4 (car weight) has large negative value. This certifies the fact, which we already know to be true, that a heavier car would require more gallons to travel an equivalent amount of miles as a lighter car.
Feature 6 (car year) has large positive value. As the technology progresses over the years, cars become more efficient in terms of fuel consumption and mileage. Hence this plot certifies that fact.

C.



As we increase the weight of the bias (λ) , the value of the root mean squared error is rising. Hence for this particular problem of linear regression of polynomial degree 1, I would choose least squares solution, that is $\lambda = 0$.

D.



For higher degree polynomial models for linear regression, the behaviour of RMSE with lambda is definitely different than from p=1.

The results obtained are as follows:

for p=1, lambda = 0 gives the minimum RMSE of 2.63364357797
for p=2, lambda = 23 gives the minimum RMSE of 2.19257412086
for p=1, lambda = 21 gives the minimum RMSE of 2.19511512727

Hence, I would choose a polynomial of degree 2 linear regression model in this particular case since that gives the least value of the RMSE. Also, the ridge regression parameter λ that attains the minima is 23.
Hence, I would use ridge regression with parameter lambda = 2 and order of the polynomial = 2.