HW1

Yinke Lu
Y14372

problem 1.

(a) $PDF = f(x) = \dfrac{e^{-\lambda}\lambda^{x}}{x!}$ $x = 0, 1, \ldots$

Joint likelihood is $P(x_1 \ldots x_n | \lambda)$

$$= \prod_{i=1}^{n} f(x_i|\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}$$

$$= e^{-\lambda n} \cdot \lambda^{\sum_{i=1}^{n} x_i} \cdot \prod_{1}^{n} \frac{1}{x!}$$

(b) log-likelihood $= \ln\left(e^{-\lambda n} \cdot \lambda^{\sum_{i=1}^{n} x_i} \cdot \prod_{1}^{n} \frac{1}{x!}\right)$

$$= -n\lambda + \left(\sum_{i=1}^{n} x_i\right)\ln(\lambda) + \ln\left(\prod_{i=1}^{n} \frac{1}{x_i}\right)$$

take the derivitive: $\nabla_\lambda \ln \prod_{i=1}^{n} f(x_i|\lambda) = -n + \left(\frac{\sum_{i=1}^{n} x_i}{\lambda}\right) + 0 = 0.$

$$\lambda^{MLE} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$$

(c) $\lambda_{MAP} = \underset{\lambda}{\operatorname{argmax}} \ln P(\lambda | x_1 \ldots x_n)$

$P(\lambda)$ $= \underset{\lambda}{\operatorname{argmax}} \ln \dfrac{P(x_1 \ldots x_n|\lambda) P(\lambda)}{P(x_1 \ldots x_n)}$

$= gamma(a, b)$

$= \sim \ln P(x_1 \ldots x_n|\lambda) + \ln P(\lambda) = L$

$\dfrac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}$

take the deriwitave of logliklihood of L.

$$\nabla L_a = \frac{\partial}{\partial a}\left[\ln P(x_1 \dots x_n \mid \lambda) + \ln P(\lambda)\right]$$

$$= \frac{\partial}{\partial \lambda}\Big[\underbrace{\ln \lambda \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \ln x_i! - \lambda n}_{\text{possion}} + \underbrace{\ln \frac{b^a}{\Gamma(a)} + (a-1)\ln \lambda - b\lambda}_{\text{Gamma}}\Big]$$

$$= \frac{\sum_{i=1}^{n} x_i}{\lambda} + \frac{a-1}{\lambda} - b - n = 0.$$

$$\lambda_{MAP} = \frac{a + \sum_{i=1}^{n} x_i - 1}{n+b}.$$

(d) $$P(\lambda \mid x_1 \dots x_n) = \frac{P(x_1 \dots x_n \mid \lambda)\, P(\lambda)}{P(x_1 \dots x_n)}$$

gamma(a, b)

$$\propto P(x_1 \dots x_n \mid \lambda)\, P(\lambda)$$

$$\propto \lambda^{\sum_{i=1}^{n} x_i + a - 1} e^{-(b+n)\lambda} \Rightarrow \text{gamma}\Big(\sum_{i=1}^{n} x_i + a,\ b+n\Big)$$

(e) $$\hat{\lambda}_{posterior} = \frac{\sum_{i=1}^{n} x_i + a}{n+b}$$

$$Var(\lambda_{posterior}) = \frac{\sum_{i=1}^{n} x_i + a}{(n+b)^2}$$

$$\hat{\lambda}_{posterior} = \left(\frac{n}{n+b}\right)\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) + \frac{a}{n+b}$$

$$= \frac{n}{n+b}\lambda_{MLE} + \frac{a}{n+b}$$

$$\hat{\lambda}_{posterior} = \frac{a+\sum_{i=1}^{n} x_i - 1}{n+b} + \frac{1}{n+b} = \lambda_{MAP} + \frac{1}{n+b}$$

2. $$E[w_{rr}] = E\left[(X^T X + \lambda I)^{-1} X^T y\right]$$

$$= (\lambda I + X^T X)^{-1} X^T E[W]$$

$$= (X^T X + \lambda I)^{-1} X^T X w.$$

We know that $y = X\beta + \varepsilon$ where $\varepsilon \sim N(0, \sigma^2 I)$

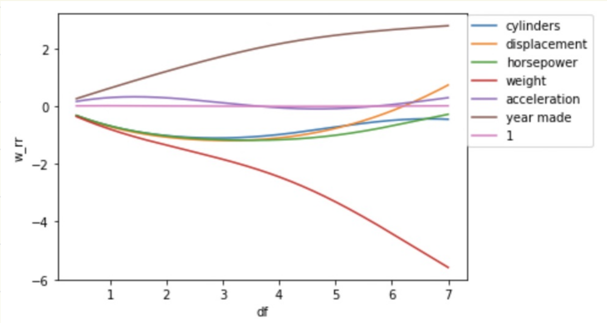
$$Var[y] = var[\varepsilon] = \sigma^2 I$$

$$Var[w_{rr}] = Var\left[(X^T X + \lambda I)^{-1} X^T y\right]$$

$$= \underbrace{(X^T X + \lambda I)^{-1} X^T}_{A} \underbrace{\sigma^2 I}_{var(y)} \underbrace{X (X^T X + \lambda I)^{-1}}_{A^T}$$

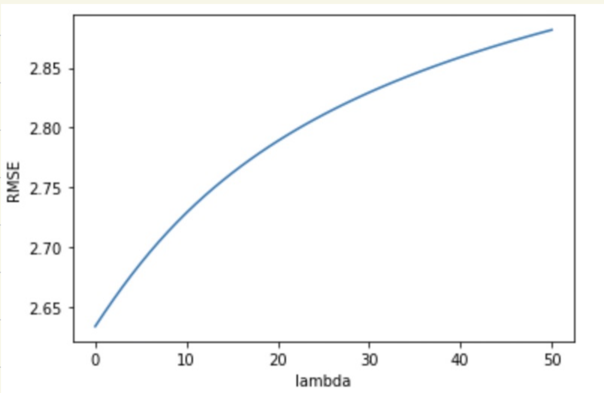
$$= \sigma^2 (X^T X + \lambda I)^{-1} X^T X (X^T X + \lambda I)^{-1}$$

3. (a)



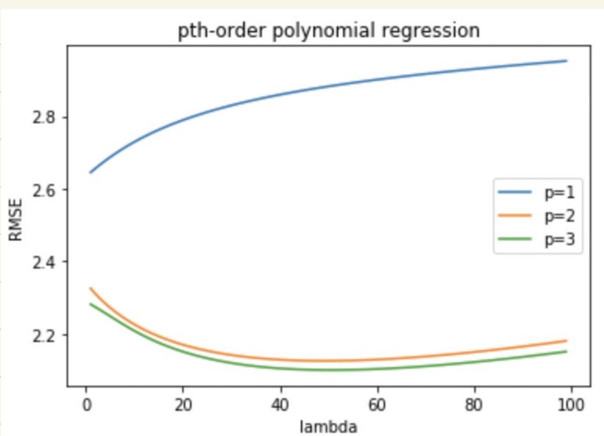
(b) weight is negatively correlated with MPG and year of made is positely correlaeed to MPG and they affect MPG level the most. From the plot, weight and year made dimension have high $|w_{ls}|$ value because they affect w_{ls} the most and thus they stands out.

(c).



when $\lambda \uparrow$ the RMSE also increase. So we want smaller λ value and will choose least square over ridge regression.

(d)



I would choose $p=3$ since it has the lowest RMSE value. But when λ value get larger, w.r.r value will approach 0. RMSE value is at lowest when $\lambda \approx 51$ and it can be the ideal value. after $\lambda > 51$, the RMSE is slowly raising.