Bangwei Zhou (bz2280)

Professor Paisley

ELEN 4720 HW1

16 February 2020

1. **Solution to Problem 1**

   (a) $p(x_1, ..., x_N \mid \lambda) = \prod_{i=1}^{n} p(x_i)$

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

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

   (b) $l(\lambda) = logp(x_1, ..., x_N \mid \lambda)$

   $= \sum_{i=1}^{n} x_i log\lambda - n\lambda + \sum_{i=1}^{n} log\frac{1}{x_i!}$

   $l\prime(\lambda) = 0 = \frac{\sum_{i=1}^{n} x_i}{\lambda} - n$

   $n\lambda = \sum_{i=1}^{n} x_i$

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

   $\hat{\lambda_{ML}} = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{X}$

   (c) $p(\lambda) = \frac{\beta\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$

   $\lambda_{MAP} = argmax_\lambda \ln p(x_1, ..., x_n \mid \lambda)$

   $= argmax_\lambda \ln \frac{p(\lambda|x)p(\lambda)}{p(x)}$

   $= argmax_\lambda$ log$p(\lambda \mid x)$ + log $p(\lambda)$ - log $p(x)$

   $= argmax_\lambda$ log $p(\lambda \mid x)$ + log $p(\lambda)$

   $= argmax_\lambda(\sum_{i=1}^{n} x_i log(\lambda) - n\lambda + (\alpha - 1)log\lambda - \beta\lambda)$

   $0 = \frac{\sum_{i=1}^{n} x_i}{\lambda} - n + \frac{\alpha-1}{\lambda} - \beta$

   $n + \beta = \frac{\sum_{i=1}^{n} x_i + (\alpha-1)}{\lambda}$

   $\hat{\lambda_{MAP}} = \frac{\sum_{i=1}^{n} x_i + (\alpha-1)}{n+\beta}$

   (d) $p(\lambda \mid x) \propto p(x \mid \lambda)p(\lambda)$

   $\propto \lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda} \prod_{i=1}^{n} \frac{1}{x_i!} \lambda^{\alpha-1} e^{-\beta\lambda}$

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

   $\propto \lambda^{\sum_{i=1}^{n} x_i + \alpha - 1} e^{-(n+\beta)\lambda}$

   Thus, $p(\lambda \mid x) \propto Gamma(\sum_{i=1}^{n} x_i + \alpha, n + \beta)$

(e) $E[\lambda \mid \alpha, \beta] = \frac{\sum_{i=1}^{n} x_i + \alpha}{n+\beta}$

$Var(\lambda \mid \alpha, \beta) = \frac{\sum_{i=1}^{n} x_i + \alpha}{(n+\beta)^2}$

Thus, when $\alpha = \beta = 0$, $\hat{\lambda_{ML}}$ is the mean of the posterior,

as $E[\lambda \mid \alpha, \beta] = \hat{\lambda_{ML}} = \frac{\sum_{i=1}^{n} x_i}{n}$.

Whereas, $\hat{\lambda_{MAP}}$ is the mode of the posterior, when $\alpha \geq 1$,

since the mode of gamma is $\frac{\alpha - 1}{\beta}$.

By the same logic, when $\alpha = \beta = 0$, $Var(\lambda \mid \alpha, \beta) = \frac{\sum_{i=1}^{n} x_i}{n^2} = \frac{1}{n} \frac{\sum_{i=1}^{n} x_i}{n} = \frac{1}{n} \hat{\lambda_{ML}}$.

2. **Solution to Problem 2**

(a) $E[w_{RR}] = E[(\lambda I + X^T X)^{-1} X^T y]$

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

$= (\lambda I + X^T X)^- 1 X^T X_w$

(b) $W_{RR} = (\lambda I + X^T X)^{-1} X^T y$

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

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

$= (\lambda (X^T X)^{-1} + I)^{-1} (X^T X) w_{LS}$

$= (\lambda (X^T X)^{-1} + I)^{-1} w_{LS}$

Let $Z = (\lambda (X^T X)^{-1} + I)^{-1}$,

then $Var(w_{RR}) = Var(Z w_{LS})$

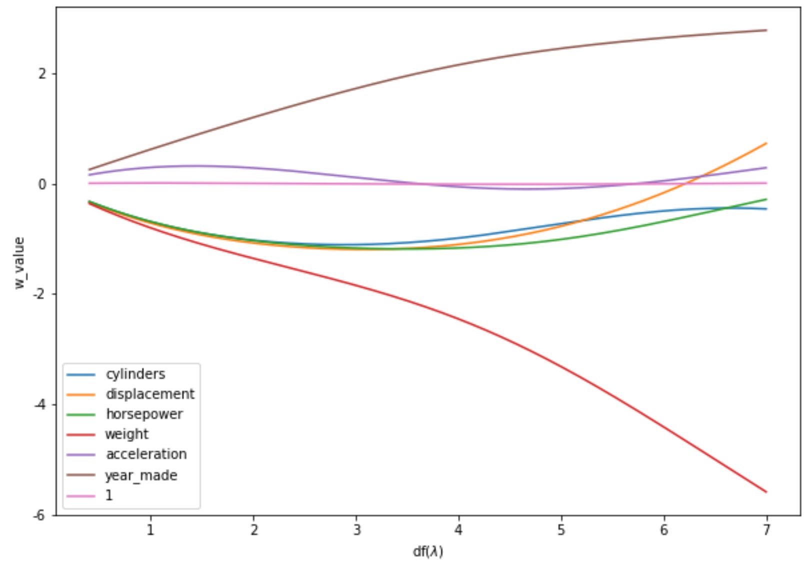
$= Z Var(w_{LS}) Z^T$

$= Z \sigma^2 (X^T X)^{-1} Z^T$

$= \sigma^2 Z (X^T X)^{-1} Z^T$

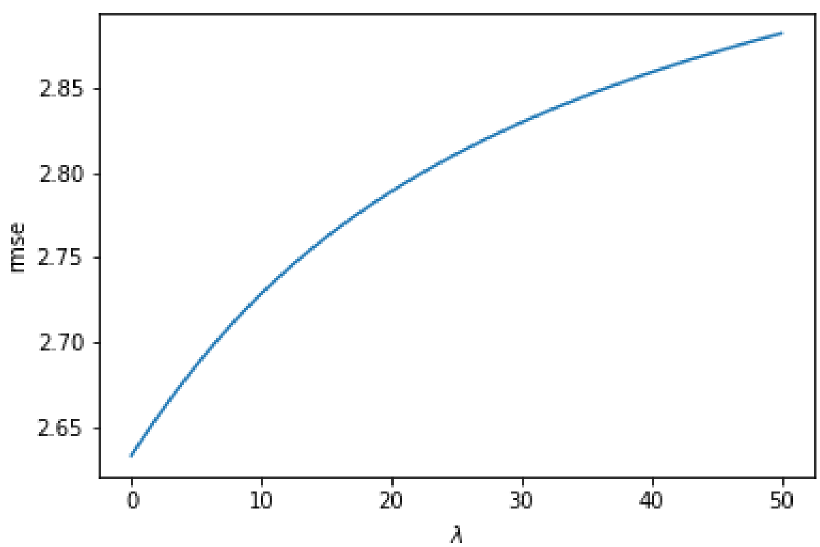
3. **Solution to Problem 3**

(a)

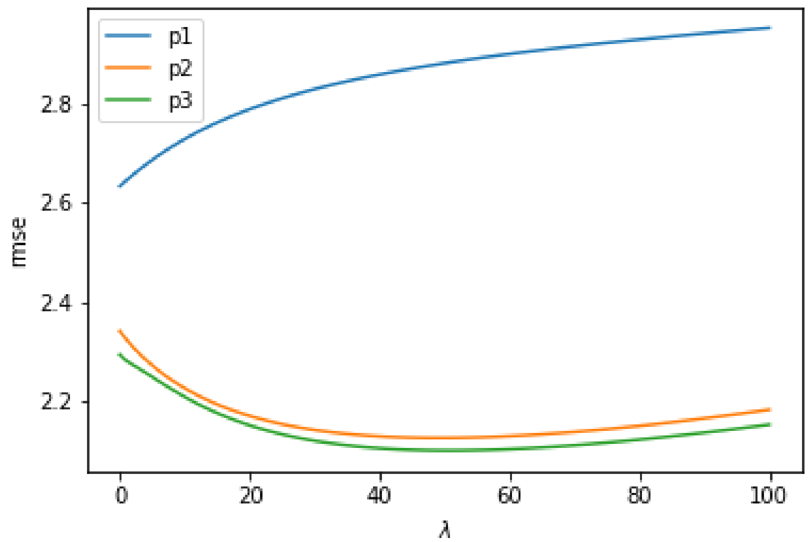


(b) Clearly, year_made and weight are the two dimensions that stood out over the others. They are the most important features in determining the miles per gallon a car will get. Since year_made is positive, it means that the miles per gallon will increase as car years increase; since weight is negative, the miles per gallon will decrease as car weight increases.

(c)



(d) RMSE increases as λ increases. The lowest RMSE is when λ equals zero. Also, as hinted in part (a), when $\lambda = 0$, $w_{RR} = w_{LS}$. Thus, $\lambda = 0$ also gives the least squares solution that generates the least RMSE. Thus, as a conclusion, we should use least squares instead of ridge regression.

(e) 

It seems that we should always pick p = 3, as it always gives the lowest RMSE; in addition, the idea value of λ is no longer 0. Instead, the ideal value is somewhere around 50, as the RMSE score is the lowest there.