BY S. VISHAL
CH18B020

CH5115 ASSIGNMENT-4

(5) a) Ridge regression:

$$\hat{\theta}_{RR} = \arg\min_{\theta} \|\underline{y} - \Phi\underline{\theta}\|_2^2 + \lambda \|\underline{\theta}\|_2^2 \quad \text{—— (1)}$$

Bayesian estimation:

i) Gaussian prior: $$f(\underline{\theta}) = \frac{1}{\left(\sqrt{2\pi\sigma_\beta^2}\right)^p} \exp\left(-\frac{1}{2} \frac{\sum_{i=1}^{p} \theta_i^2}{\sigma_\beta^2}\right)$$

where p is the length of $\underline{\theta}$ vector. —— (2)

ii) $$f(\underline{y}_N | \underline{\theta}) = \prod_{k=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_e} \exp\left(-\frac{1}{2} \frac{(y[k] - \varphi^T[k]\underline{\theta})^2}{\sigma_e^2}\right)$$

$(\because e(k) = y(k) - \varphi^T(k)\underline{\theta}$ & $e(k) \sim N(0, \sigma_e^2))$

$$\Rightarrow f(\underline{y}_N | \underline{\theta}) = a \exp\left(\frac{-1}{2\sigma_e^2} \sum_{k=1}^{N} (y[k] - \varphi^T(k)\theta)^2\right) \quad \text{—— (3)}$$

where 'a' is a constant

iii) $$f(\theta | \underline{y}_N) = \frac{f(\theta) f(y_N|\theta)}{f(y_N)}$$

Substituting from eqns (2) & (3) and collecting the constant terms in one place.

$$f(\theta | y_N) = C \exp\left(-\frac{1}{2}\left[\frac{\sum_{i=1}^{p} \theta_i^2}{\sigma_\beta^2} + \frac{\sum_{k=1}^{N} (y(k) - \varphi^T(k)\theta)^2}{\sigma_e^2}\right]\right)$$

$$= C \exp\left(-\frac{1}{2}\left(\frac{\|\underline{\theta}\|_2^2}{\sigma_\beta^2} + \frac{\|\underline{y}_N - \Phi\underline{\theta}\|_2^2}{\sigma_e^2}\right)\right)$$

where C is just a constant.

$$\Rightarrow f(\underline{\theta} \mid \underline{y_N}) = C \exp\left(-\frac{\sigma_e^2}{2}\left\{\left(\frac{\sigma_e^2}{\sigma_\beta^2}\right)\|\underline{\theta}\|_2^2 + \|\underline{y}_N - \Phi\theta\|_2^2\right\}\right) \quad (4)$$

Consider the MAP estimate :

$$\hat{\theta}_{BE} = \arg\max_{\theta} f(\underline{\theta} \mid \underline{y_N})$$

From (4),

$$\hat{\theta}_{BE} = \arg\max_{\theta} C \exp\left(-\frac{\sigma_e^2}{2}\left\{\left(\frac{\sigma_e^2}{\sigma_\beta^2}\right)\|\theta\|_2^2 + \|y - \Phi\theta\|_2^2\right\}\right)$$

$$= \arg\max_{\theta}\left(-\left\{\frac{\sigma_e^2}{\sigma_\beta^2}\|\theta\|_2^2 + \|y - \Phi\,\theta\|_2^2\right\}\right)$$

$$\Rightarrow \underline{\hat{\theta}_{BE}} = \arg\min_{\theta} \|y - \Phi\theta\|_2^2 + \frac{\sigma_e^2}{\sigma_\beta^2}\|\theta\|_2^2 \quad (5)$$

We find that the Bayesian estimate (eqn (5)) is identical to the Ridge Regression estimate (eqn (1))

ie $\boxed{\lambda = \frac{\sigma_e^2}{\sigma_\beta^2}}$

(5) b) $\min_{\theta} \|y - \Phi\theta\|_2^2 + \lambda\left[\alpha\|\theta\|_2^2 + (1-\alpha)\|\theta\|_1\right]$

(Elastic net)

$= \min_{\theta} \|y - \Phi\theta\|_2^2 + \lambda\left[\alpha\|\theta\|_2^2 + (1-\alpha)\|\theta\|_1\right]$

$= \min_{\theta} \|y - \Phi\theta\|_2^2 + (\lambda\alpha)\|\theta\|_2^2 + \lambda(1-\alpha)\|\theta\|_1$

$= \min_{\theta} \sum_{k=1}^{N} (y[k] - \underline{\varphi}^T(k)\underline{\theta})^2 + (\lambda\alpha.) \sum_{i=1}^{p} \theta_i^2$

$+ \lambda(1-\alpha)\|\theta\|_1$

$= \min_{\theta} \sum_{k=1}^{N} (y[k] - \varphi^T(k)\underline{\theta})^2 + \sum_{i=1}^{p} (\lambda\alpha\theta_i^2) + \lambda(1-\alpha)\|\theta\|_1$

$= \min_{\theta} \sum_{k=1}^{N} (y[k] - \varphi^T(k)\underline{\theta})^2 + \sum_{i=1}^{p} (0 - \lambda\alpha\theta_i)^2$

$+ \lambda(1-\alpha)\|\theta\|_1$

$= \min_{\theta} \sum_{k=1}^{N} (y[k] - \varphi^T(k)\underline{\theta})^2 + \sum_{i=1}^{p} (0 - \lambda\alpha\beta^T(i)\underline{\theta})^2$

$+ \lambda(1-\alpha)\|\theta\|_1$

where $\beta_j(i) = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{otherwise} \end{cases}$

(ith element of $\beta(i)$ in 1, rest of the elements are zeroes. $\beta(i) \in \mathbb{R}^{p\times 1}$)

$= \min_{\theta} \sum_{k=1}^{N} (y[k] - \varphi^T(k)\theta)^2 + \sum_{i=1}^{p} (0 - \lambda\alpha\beta^T(i)\theta)$

$+ \lambda(1-\alpha)\|\theta\|_1$

The second summation is of the form,

$$\sum \left(y[k] - \varphi^T(k)\, \underline{\theta} \right)^2 = 0$$

where $y[k] = 0$ & $\varphi(k) = \lambda\alpha\,\beta(k)$

So accordingly we augment the matrix $\underline{y}$ & Φ such that they become $(N+P) \times 1$ and $(N+P) \times P$ respectively

$$\underline{y}^* := \begin{bmatrix} \underline{y} \\ 0_{p\times 1} \end{bmatrix}$$

where $0_{p\times 1}$ is a column vector of zeroes of size $p \times 1$

$$\Phi^* = \begin{bmatrix} \Phi \\ \lambda\alpha I_{p\times p} \end{bmatrix}$$

where $I_{p\times p}$ is the identity matrix with dimensions $p\times p$.

$(= [\beta(1) \dots \beta(P)])$

Now we can write the elastic net optimisation problem as

$$\min_{\theta} \sum_{k=1}^{N+P} \left(y^*(k) - \varphi^{*T}(k)\, \underline{\theta} \right)^2 + \lambda(1-\alpha) \|\theta\|_1$$

$$= \min_{\theta} \| \underline{y}^* - \Phi^* \underline{\theta} \|_2^2 + \lambda(1-\alpha)\|\theta\|_1$$

which is a LASSO problem.