HW 3

1 Ex.3.5

β^c与β之间的关系可以写成

$$\beta_0^c = \beta_0 + \sum_{j=1}^{p} \bar{x}_j \beta_j$$
$$\beta_j^c = \beta_j, j = 1, 2, \cdots, p$$

将 $\sum_{i=1}^{N}[y_i - \beta_0^c - \sum_{j=1}^{p}(x_{ij} - \bar{x}_j)\beta_j^c]^2 + \lambda \sum_{j=1}^{p}(\beta_j^c)^2$ 关于β_0^c求导，有

$$\beta_0^c - \frac{1}{N}\sum_{i=1}^{N} y_i = 0$$

于是$\beta_0^c = \bar{y}$。令

$$\tilde{y}_i = y_i - \beta_0^c,$$
$$\tilde{x}_{ij} = x_{ij} - \bar{x}_j,$$

$$\hat{\beta}^c = \arg\min_{\beta^c} \left\{ \sum_{i=1}^{N}[y_i - \beta_0^c - \sum_{j=1}^{p}(x_{ij} - \bar{x}_j)\beta_j^c]^2 + \lambda \sum_{j=1}^{p}(\beta_j^c)^2 \right\}$$

可以写为

$$\hat{\beta}^c = \arg\min_{\beta^c} \left\{ \sum_{i=1}^{N}[\tilde{y}_i - \sum_{j=1}^{p}(\tilde{x}_{ij})\beta_j^c]^2 + \lambda \sum_{j=1}^{p}(\beta_j^c)^2 \right\}$$

上式形式与原始ridge回归问题一致，于是该问题的解为

$$\hat{\beta}_c = (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}} + \lambda \mathbf{I})^{-1} \tilde{\mathbf{X}}^T \mathbf{y}$$

将上述对应关系代入lasso，

$$\hat{\beta}^{\text{lasso}} = \arg\min_{\beta^c} \left\{ \sum_{i=1}^{N}[y_i - \beta_0^c - \sum_{j=1}^{p}(x_{ij} - \bar{x}_j)\beta_j^c]^2 \right\}$$
$$s.t. \sum_{j=1}^{p} |\beta_j^c| \leq t$$

与原先lasso的形式完全一致。

2 Ex.3.6

根据高斯分布概率密度函数的定义，有

$$P(\mathbf{y} \mid \mathbf{X}, \beta) = \frac{1}{(2\pi)^{\frac{N}{2}} \sigma^N} \exp\left(-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) \right)$$
$$P(\beta) = \frac{1}{(2\pi)^{\frac{p}{2}} \tau^{\frac{p}{2}}} \exp\left(-\frac{1}{2\tau} \beta^T \beta \right)$$

由贝叶斯公式

$$P(\beta \mid \mathbf{y}, \mathbf{X}) = \frac{P(\mathbf{y} \mid \beta, \mathbf{X}) P(\beta)}{P(\mathbf{y})}$$
$$\log P(\beta \mid \mathbf{y}, \mathbf{X}) = -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) - \frac{1}{2\tau} \beta^T \beta + C$$

令$\lambda = \frac{\sigma^2}{\tau}$，$\beta$的均值与众数对应着ridge的解。即

$$\min_{\beta} (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) + \lambda \beta^T \beta,$$

3 Ex.3.7

根据高斯分布概率密度函数的定义，有

$$P(\mathbf{y} \mid \beta) = \frac{1}{(2\pi)^{\frac{N}{2}} \sigma^N} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_j x_{ij}\beta_j \right)^2 \right)$$

$$P(\beta) = \frac{1}{(2\pi)^{\frac{p}{2}} \sigma^p} \exp\left(-\frac{1}{2\tau^2} \sum_{j=1}^{p} \beta_j^2 \right)$$

由贝叶斯公式，

$$P(\beta \mid \mathbf{y}) = \frac{P(\mathbf{y} \mid \beta) P(\beta)}{P(\mathbf{y})}$$

代入概率密度，并取负对数

$$\begin{aligned} -\ln(P(\beta \mid \mathbf{y})) &= \frac{1}{2\sigma^2} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_j x_{ij}\beta_j)^2 + \frac{1}{2\tau^2} \sum_{j=1}^{p} \beta_j^2 + C \\ &= \frac{1}{2\sigma^2} \left(\sum_{i=1}^{N} (y_i - \beta_0 - \sum_j x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right) + C \end{aligned}$$

其中$\lambda = \frac{\sigma^2}{\tau^2}$，所以$P(\beta \mid \mathbf{y})$的负对数与$\sum_{i=1}^{N} (y_i - \beta_0 - \sum_j x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$呈正比例关系。