ESL 3.6

$\beta \sim N(0, \tau I)$, $y|\beta \sim N(X\beta, \sigma^2 I)$

$\therefore$ $\therefore f(\beta|y) = \dfrac{f(y|\beta)\,\pi(\beta)}{f(y)} \propto f(y|\beta)\cdot\pi(\beta)$

即 $f(\beta|y) \propto \exp\{-\frac{1}{2\sigma^2}(y-X\beta)^T(y-X\beta) + \frac{1}{2\tau}\beta^T\beta\}$

其中 $\exp\{-\frac{1}{2\sigma^2}(y-X\beta)^T(y-X\beta) + \frac{1}{2\tau}\beta^T\beta$

$$= \exp\{-\frac{1}{2\sigma^2}[(y-X\beta)^T(y-X\beta) + \frac{\sigma^2}{\tau}\beta^T\beta]\}$$

$$= \exp\{-\frac{1}{2\sigma^2}(\beta^T(X^TX + \frac{\sigma^2}{\tau}I)\beta - 2y^TX\beta + y^Ty)\}$$

$$= \exp\{-\frac{1}{2\sigma^2}(\beta - X^Ty(X^TX + \frac{\sigma^2}{\tau}I)^{-1})^T(X^TX + \frac{\sigma^2}{\tau}I)(\beta - (X^TX + \frac{\sigma^2}{\tau}I)^{-1}X^Ty) + C\}$$

为 $N((X^TX + \frac{\sigma^2}{\tau}I)^{-1}X^Ty,\ C_2)$ 的核函数

故 $\beta|y \sim N((X^TX + \frac{\sigma^2}{\tau}I)^{-1}X^Ty,\ C_2)$

$E(\beta|y) = (X^TX + \frac{\sigma^2}{\tau}I)^{-1}X^Ty$

下面给出: Ridge Regression:

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

求导为零得: $-2X^T(y-X\beta) + 2\lambda\beta = 0$

即 $X^Ty = (X^TX + \lambda I)\beta$

故 $\beta = (X^TX + \lambda I)^{-1}X^Ty$

又 $\because \max f(\beta|y) \Leftrightarrow \max -\frac{1}{2\sigma^2}[(y-X\beta)^T(y-X\beta) + \frac{\sigma^2}{\tau}\beta^T\beta]$

$\Leftrightarrow \min\ y-X\beta)^T(y-X\beta) + \frac{\sigma^2}{\tau}\beta^T\beta$

令 $\lambda = \frac{\sigma^2}{\tau}$ 我们会现此时贝叶斯方法即为 $\lambda = \frac{\sigma^2}{\tau}$ 的岭回归方法.

综上, $\lambda = \frac{\sigma^2}{\tau}$ 时, 岭回归与贝叶斯求出的 $\hat{\beta}$ 均为 $(X^TX + \lambda I)^{-1}X^Ty$, 即 $E(\beta|y)$.

此时二者等价.

3.7.

$\because$ $y_i \sim N(\beta_0 + x_i^T\beta, \sigma^2)$，给定$\beta$时，

$\therefore$ $f(y_i) = \dfrac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\dfrac{(y_i - (\beta_0 + x_i^T\beta))^2}{2\sigma^2}\right\}$.

$y = (y_1, y_2, \cdots, y_N)$

$\therefore$ $f(y|\beta) = \prod f(y_i) = \dfrac{1}{(\sqrt{2\pi})^N \sigma^N} \exp\left\{-\dfrac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - (\beta_0 + x_i^T\beta))^2\right\}$

又$\because$ $\beta_j \overset{iid}{\sim} N(0, \tau^2)$

$\therefore$ $\pi(\beta) = \prod_{j=1}^{p} \pi(\beta_j) = \dfrac{1}{(\sqrt{2\pi}\tau)^p} \exp\left\{-\dfrac{1}{2\tau^2}\sum_{j=1}^{p}\beta_j^2\right\}$

由贝叶斯可得:

$$P(\beta|y) \propto P(y|\beta)\cdot\pi(\beta) \propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - \beta_0 - x_i^T\beta)^2 - \frac{1}{2\tau^2}\sum_{j=1}^{p}\beta_j^2\right\}.$$

故 $\ln p(\beta|y) \propto -\dfrac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - \beta_0 - x_i^T\beta)^2 - \dfrac{1}{2\tau^2}\sum_{j=1}^{p}\beta_j^2$

故 $\ln P(\beta|y) \propto -\dfrac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p}x_{ij}\beta_j)^2 - \dfrac{1}{2\tau^2}\sum_{j=1}^{p}\beta_j^2$

令$\lambda = \dfrac{\sigma^2}{\tau^2}$，则有

$$\ln P(\beta|y) \propto \left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p}x_{ij}\beta_j)^2 + \lambda\sum_{j=1}^{p}\beta_j^2\right]\right\}.$$

综上 $-\ln P(\beta|y) \propto \sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p}x_{ij}\beta_j)^2 + \lambda\sum_{j=1}^{p}\beta_j^2$，其中$\lambda = \dfrac{\sigma^2}{\tau^2}$

3.30.

令 $\tilde{y}=\begin{bmatrix} y \\ 0_{p\times 1} \end{bmatrix}$，$\tilde{X}=\begin{bmatrix} X \\ \delta I_{p\times p} \end{bmatrix}$

则有 $\tilde{y}-\tilde{X}\beta=\begin{bmatrix} y-X\beta \\ \delta\beta \end{bmatrix}$

$\therefore \|\tilde{y}-\tilde{X}\beta\|_2^2=\|y-X\beta\|_2^2+\delta^2\|\beta\|_2^2$

原目标函数可化为 $\|\tilde{y}-\tilde{X}\beta\|_2^2+\gamma\|\beta\|_1$

$\because \|\tilde{y}-\tilde{X}\beta\|_2^2+\gamma\|\beta\|_1=\|y-X\beta\|_2^2+\delta^2\|\beta\|_2^2+\gamma\|\beta\|_1$

令 $\delta^2=\lambda\alpha$，$\gamma=\lambda(1-\alpha)$，则与原目标函数一致

综上，原问题可化为LASSO问题：

$$\min_{\beta}\|\tilde{y}-\tilde{X}\beta\|_2^2+\gamma\|\beta\|_1,\text{ 其中 }\delta^2=\lambda\alpha,\ \gamma=\lambda(1-\alpha)$$