Ex. 3.6

$$\mathbb{P}(\beta \mid y) \propto \mathbb{P}(y \mid \beta)\,\mathbb{P}(\beta)$$

Where, $y \sim N(X\beta, \sigma^2 \mathbb{I})$

$\beta \sim N(0, \tau \mathbb{I})$

and since Normal is a conjugate prior for the Normal we know that $\mathbb{P}(\beta \mid y) \sim N(\mu, \Sigma)$

Now, $\mathbb{P}(\beta \mid y) \propto \exp(-\tfrac{1}{2}\sigma^{-2}(y - X\beta)^T(y - X\beta))\exp(-\tfrac{1}{2}\tau^{-1}\beta^T\beta)$

$$\propto \exp(-\tfrac{1}{2}\sigma^{-2}(y - X\beta)^T(y - X\beta) - \tfrac{1}{2}\tau^{-1}\beta^T\beta)$$

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

$$\propto \exp(-\tfrac{1}{2}\sigma^{-2}((X\beta)^TX\beta - (X\beta)^Ty - y^TX\beta) + \tau^{-1}\beta^T\beta) \quad (*)$$

and we know this is proportional to $N(\mu, \Sigma)$

e.g.

$$\exp(-\tfrac{1}{2}(\beta - \mu)^T\Sigma^{-1}(\beta - \mu))$$

$$= \exp(-\tfrac{1}{2}[\beta^T\Sigma^{-1}\beta - \mu^T\Sigma^{-1}\beta - \beta^T\Sigma^{-1}\mu + \mu^T\Sigma^{-1}\mu])$$

$$\propto \exp(-\tfrac{1}{2}[\beta^T\Sigma^{-1}\beta - \mu^T\Sigma^{-1}\beta - \beta^T\Sigma^{-1}\mu]) \quad (**)$$

Now match β terms to obtain μ and Σ - use $(*)$ and $(**)$

1. $\sigma^{-2}((X\beta)^TX\beta) + (\tau^{-1}\beta^T\beta) = \beta^T\Sigma^{-1}\beta$

$$\sigma^{-2}(X^TX) + (\tau^{-1}\mathbb{I}) = \Sigma^{-1}$$

$$\sigma^{-2}\left[X^TX + \frac{\sigma^2}{\tau}\mathbb{I}\right] = \Sigma^{-1}$$

$$\Rightarrow \Sigma = \left(\sigma^{-2}\left[X^T X + \frac{\sigma^2}{\tau}\mathbb{I}\right]\right)^{-1}$$

$$\Sigma = \sigma^2\left(X^T X + \frac{\sigma^2}{\tau}\mathbb{I}\right)^{-1}$$

② $\sigma^{-2}\left((X\beta)^T y + y^T(X\beta)\right) = \mu^T\Sigma^{-1}\beta + \beta^T\Sigma^{-1}\mu$

Now we will use the facts that:

(a) $(X\beta)^T y = y^T(X\beta)$

(b) Since $X^T X$ is symetric Σ^{-1} is also symetric and $\mu^T\Sigma^{-1}\beta = \beta^T\Sigma^{-1}\mu$

$$\Rightarrow 2\sigma^{-2}\left((X\beta)^T y\right) = 2\beta^T\Sigma^{-1}\mu$$

$$\sigma^{-2}\beta^T X^T y = \sigma^{-2}\beta^T\left(X^T X + \frac{\sigma^2}{\tau}\mathbb{I}\right)\mu$$

$$X^T y = \left(X^T X + \frac{\sigma^2}{\tau}\mathbb{I}\right)\mu$$

$$\Rightarrow \mu = \left(X^T X + \frac{\sigma^2}{\tau}\mathbb{I}\right)^{-1} X^T y$$

So if $\lambda = \frac{\sigma^2}{\tau}$ the ridge regression estimate is the mean of the Posterior distribution and, since it is Normal, the mode too