Linear Regressim

$$\mathbb{P}(Y|\mu,\Sigma) \propto |\Sigma|^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} (y_i-\mu)^T\Sigma^{-1}(y_i-\mu)\right\}$$

Trace Trick $$tr\left\{\sum_i (y_i-\mu)^T\Sigma^{-1}(y_i-\mu)\right\} = tr\left\{\Sigma^{-1}\underbrace{\sum_i (y_i-\mu_i)(y_i-\mu_i)^T}_{S}\right\}$$

Centering Trick $$\sum_i (y_i-\mu)^T\Sigma_i (y_i-\mu) = \sum_i (y_i-y^*+y^*-\mu)^T\Sigma_i^{-1}(y_i-y^*+y^*-\mu)$$

$$+\sum_i (y^*-\mu)^T\Sigma_i^{-1}(y^*-\mu)$$

where y^* chosen s.t. $\sum(y_i-y_i^*)^T\Sigma_i^{-1}(y^*-\mu) = \sum_i (y^*-\mu)^T\Sigma_i^{-1}(y_i-y^*)=0$

$$\Rightarrow \sum_i \Sigma_i^{-1}(y_i-y^*)=0 \Rightarrow \boxed{y^*=\left(\sum_i\Sigma_i^{-1}\right)^{-1}\left(\sum_i\Sigma_i^{-1}y_i\right)}$$

→ Weighted average of precisions

Therefore combing these tricks

$$\sum_i (y_i-\mu)^T\Sigma^{-1}(y_i-\mu) = \sum_i (y_i-y^*)^T\Sigma^{-1}(y_i-y^*) + n(\mu-y^*)^T\Sigma^{-1}(\mu-y^*)$$

In this setting $y^* = (n\Sigma)^{-1}\sum\Sigma_i y_i = \bar{y}$

$$\mathbb{P}(Y|\mu,\Sigma)\propto|\Sigma|^{-n/2}\exp\left\{-\frac{1}{2}\, tr\left[\Sigma^{-1}\underbrace{\sum_i(y_i-\bar y)(y_i-\bar y)^T}_{S} + (\mu-\bar y)^T\left(\frac{\Sigma^{-1}}{n}\right)(\mu-\bar y)\right]\right\}$$

This inspires

$$\mathbb{P}(\mu|\Sigma)\propto\exp\left\{(\mu-\mu_0)^T\left(\frac{\Sigma}{k_0}\right)^{-1}(\mu-\mu_0)\right\}\sim N(\mu_0,\Sigma/k_0)$$

$$\mathbb{P}(\mu|\Sigma,y) \propto \exp\left\{-\frac{1}{2}\left[(\mu-\bar{y})^T\left(\frac{\Sigma}{n}\right)^{-1}(\mu-\bar{y}) + (\mu-\mu_0)^T\left(\frac{\Sigma}{k_0}\right)^{-1}(\mu-\mu_0)\right]\right\}$$

Centering trick

$$\propto \exp\left\{-\frac{1}{2}\left[(\bar{y}-\hat{\mu})^T(\Sigma/k_0)^{-1}(\bar{y}-\hat{\mu}) + (\mu_0-\hat{\mu})^T\left(\frac{\Sigma}{n}\right)^{-1}(\mu_0-\hat{\mu})\right.\right.$$

$$\left.\left. + (\mu-\hat{\mu})^T\left[\left(\frac{\Sigma}{n}\right)^{-1} + \left(\frac{\Sigma}{k_0}\right)^{-1}\right](\mu-\hat{\mu})\right\}$$

$$\propto \exp\left\{\frac{-1}{2}(\mu-\hat{\mu})^T\left[\left(\frac{\Sigma}{n}\right)^{-1} + \left(\frac{\Sigma}{k_0}\right)^{-1}\right](\mu-\hat{\mu})\right\}$$

$$\sim N\left(\hat{\mu}, \frac{1}{n+k_0}\Sigma\right)$$

$$\hat{\mu} = (n\Sigma^{-1} + k_0\Sigma^{-1})(n\Sigma^{-1}\bar{y} + k_0\Sigma^{-1}\mu_0) = \frac{n}{n+k_0}\bar{y} + \frac{k_0}{n+k_0}\mu_0$$

One can show

$$\mathbb{P}(y|\Sigma) \propto |\Sigma|^{-n/2}\exp\left(\frac{-1}{2}\operatorname{tr}(\Sigma^{-1}\tilde{S})\right) \quad \tilde{S} = S + \frac{nk_0}{n+k_0}(\bar{y}-\mu_0)(\bar{y}-\mu_0)^T$$

$$\mathbb{P}(\Sigma) \propto |\Sigma|^{-\frac{\nu_0+p+1}{2}}\exp\left(-\frac{1}{2}\operatorname{tr}(\Sigma^{-1}\Lambda_0)\right)$$

$$\sim \text{Inv-Wishart}_p(r_0, \Lambda_0^{-1})$$

$$\Sigma|y \sim \text{Inv-Wish}(n+r_0, (\Lambda_0+\tilde{S})^{-1})$$

Linear Regression

Always conditional on X. $e_i | \sigma^2 \overset{iid}{\sim} N(0, \sigma^2)$

$$Y | \beta, \sigma^2 \sim N(X\beta, \sigma^2 I_n)$$

$$\mathbb{P}(Y|\beta, \sigma^2) \propto |\sigma^2 I_n|^{-1/2} \exp\left\{-\frac{1}{2}(Y-X\beta)^T(\sigma^2 I_n)^{-1}(Y-X\beta)\right\}$$

$$\propto (\sigma^2)^{-N/2} \exp\left(-\frac{1}{2\sigma^2}(Y-X\beta)^T(Y-X\beta)\right)$$

By a similar trick:

$$(Y-X\beta)^T(Y-X\beta) = (Y-X\beta^* + X\beta^* - X\beta)^T(Y-X\beta^* + X\beta^* - X\beta)$$

$$= (Y-X\beta^*)^T(Y-X\beta^*) + (X\beta^* - X\beta)^T(X\beta^* - X\beta)$$

s.t. $(Y-X\beta^*)^T(X\beta^* - X\beta) = 0 \implies X^T(y - X\beta^*) = 0$

$= (\beta^* - \beta^T)X^T(Y-X\beta^*) = 0 \implies \boxed{X^T X\beta^* = X^T y}$

Thus, $\beta^* = \beta_{MLE}$ and

$$\mathbb{P}(Y|\beta, \sigma^2) \propto (\sigma^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2}\underbrace{(y-X\beta^*)^T(y-X\beta^*)}_{RSS} - \frac{1}{2\sigma^2}(\beta-\beta^*)^T X^T X(\beta-\beta^*)\right\}$$

$$\mathbb{P}(\beta | \sigma^2) \propto \exp\left\{-\frac{1}{2\sigma^2}(\beta-\beta_0)^T X^T X(\beta-\beta_0)\right\}$$

$$\beta | \sigma^2 \sim N(\beta_0, \underset{\text{scaling}}{g}\sigma^2 (X^T X)^{-1})$$ "Zeller's g prior"

In general $\beta \sim N(\beta_0, \Sigma_0)$

The joint post. is then

$$\mathbb{P}(\beta, \sigma^2 \mid y) \propto \exp\left(-\frac{1}{2\sigma^2}(y - X\beta)^T(y - X\beta)\right)\exp\left(-\frac{1}{2}(\beta - \beta_0)^T \Sigma_0^{-1}(\beta - \beta_0)\right\}$$

$$\times (\sigma^2)^{-\left(\frac{\nu_0}{2}+1\right)} \exp\left(\frac{-\nu\tau^2}{2\sigma^2}\right)$$

But

$$\frac{1}{\sigma^2}(y - X\beta)^T(y - X\beta) + (\beta - \beta_0)^T \Sigma_0^{-1}(\beta - \beta_0)$$

$$= \frac{1}{2\sigma^2}(y - X\beta^*)^T(y - X\beta^*) + \frac{1}{\sigma^2}(\beta - \beta^*)X^TX(\beta - \beta^*) + (\beta - \beta_0)^T \Sigma_0^{-1}(\beta - \beta_0)$$

$$= \text{---}\,''\text{---} + \frac{1}{\sigma^2}(\hat\beta - \beta^*)^T X^T X(\hat\beta - \beta^*) + (\hat\beta - \beta_0)^T \Sigma_0^{-1}(\hat\beta - \beta_0) + (\beta - \hat\beta)^T\left(\frac{1}{\sigma^2}X^TX + \Sigma_0^{-1}\right)(\beta - \hat\beta)$$

$$\boxed{\begin{aligned}\hat\beta &= \left(\frac{1}{\sigma^2}X^TX + \Sigma_0^{-1}\right)^{-1}\left(\frac{1}{\sigma^2}X^TX\beta^* + \Sigma_0^{-1}\beta_0\right)\\ &= \left(\frac{1}{\sigma^2}X^TX + \Sigma_0^{-1}\right)^{-1}\left(\frac{1}{\sigma^2}X^Ty + \Sigma_0^{-1}\beta_0\right)\end{aligned}}$$

$$\boxed{\beta \mid y, \sigma^2 \sim N\left(\hat\beta, \left(\frac{1}{\sigma^2}X^TX + \Sigma_0^{-1}\right)^{-1}\right)}$$