

1. Assume $y \sim N(X\beta, s^2\mathbf{I})$ where X is $n \times p$.

(a) The prior distribution whose MAP corresponds to the ridge estimator is

$$\beta \sim N(0, \lambda^{-1}\mathbf{I}_p)$$

In this prior, λ controls the amount of shrinkage. The variance of the prior is solely dependant on λ so as it increase, the variance decreases and becomes tighter around the mean(in this case, 0).

(b) The prior distributions whose MAP corresponds to the LASSO estimator are

$$p(\beta|\sigma^2) \propto \prod_{j=1}^p \frac{\lambda}{2\sqrt{\sigma^2}} e^{-\lambda|\beta_j|/\sqrt{\sigma^2}}$$

$$p(\sigma^2) \propto \frac{1}{\sigma^2}$$

In this prior, λ also controls the amount of shrinkage, but it is not as straight forward as the cases before. Here we have a Laplace prior with variance $2\frac{\sigma^2}{\lambda^2}$. So as λ increases the variance will go to 0. So λ controls again the shrinkage, but in this case σ^2 will also affect the how much λ influences shrinkage.

(c) The elastic net estimator is

$$p(\beta|\sigma^2) \propto \exp\left\{-\frac{1}{2\sigma^2}(\lambda_1\|\beta\|_1 + \lambda_2\|\beta\|_2^2)\right\}$$

$$p(\sigma^2) \propto \frac{1}{\sigma^2}$$

The λ parameters control the blend between a Normal and Laplace distribution. When $\lambda_2 > \lambda_1$, the prior takes more of a form of a normal distribution which corresponds to the prior for the ridge estimator and when $\lambda_1 > \lambda_2$, the prior takes more of a form of a Laplace distribution which corresponds to the prior for the LASSO estimator.