Regression: modeling conditional expectations

Ex: $\mathbb{E}\, Y|X = X\beta$ $\quad Y \in \mathbb{R}^n$, $X \in \mathbb{R}^{n\times p}$, $\beta \in \mathbb{R}^p$

Common model

$$Y \mid X, \beta, \sigma^2 \sim N(X\beta, \sigma^2 I)$$

β & σ^2 unknown.

We'll turn our attention to common priors for β.

Common prior setup: independent exp. power priors on β entries.

$$p(\beta) = \prod_{j=1}^{p} p(\beta_j)$$

$$p(\beta_j) \propto \exp\left\{ - \left| \frac{\beta_j}{\tau} \right|^{\alpha} \right\}$$

$\alpha = 2 \Rightarrow$ normal
$\alpha = 1 \Rightarrow$ Laplace
$\alpha \in (0,1) \Rightarrow$ "Bridge"

Following Ex 1: (considering σ^2 fixed, known right now)

$$p(\beta \mid Y, X) \propto p(Y \mid \beta, X)\, p(\beta)$$

posterior of β

$\hat{\beta}_{MAP} := \underset{\beta}{\operatorname{argmax}}\; p(\beta \mid Y, X)$ corresponds to many well known regularization procedures.

$$\hat{\beta}_{MAP} = \underset{\beta}{\operatorname{argmax}}\; \log p(\beta \mid Y, X)$$

$$= \underset{\beta}{\operatorname{argmax}}\; -\frac{1}{2}\|Y - X\beta\|_2^2 - \sum_j \left|\frac{\beta_j}{\tau}\right|^{\alpha}$$

$$= \underset{\beta}{\operatorname{argmin}}\; \frac{1}{2}\|Y - X\beta\|_2^2 + \lambda \sum |\beta_j|^{\alpha} \quad ; \quad \lambda = \left|\frac{1}{\tau}\right|^{\alpha}$$

$\alpha = 2$: ridge regression
$\alpha = 1$: lasso
$\alpha = 0$: best subset selection

well known regularization procedures

Mike West (1987): exp. power prior can be represented as scale mixture of normals.

$$p(\theta) \propto e^{-|\theta|^{\alpha}} = \int_0^{\infty} e^{-s\theta^2/2} \, g(s) \, ds$$

for some mixing distr. $g(s)$.

often we parameterize:

$$\beta_j \mid \lambda_j, \tau \sim N(0, \lambda_j^2 \tau^2)$$

λ_j: local shrinkage parameter

τ: global shrinkage parameter

$\lambda_j \sim g(\lambda)$

so $p(\beta_j \mid \tau) \propto \exp\left\{ - \left| \frac{\beta_j}{\tau} \right|^{\alpha} \right\} = \int \underbrace{p(\beta_j \mid \lambda_j, \tau)}_{\text{normal density}} g(\lambda) \, d\lambda$

when $\lambda_j \sim C^+(0,1)$ ← ½ Cauchy $\Rightarrow$ horseshoe prior

when $\lambda_j \sim \exp(2) \Rightarrow$ Laplace prior

$\lambda_j \sim$ inverse-Gamma $\Rightarrow$ student-t

See Carvalho, Polson & Scott (2009) "Handling sparsity via the Horseshoe" for further reading.