Equivalence of ℓ_2-regularization and Gaussian priors

$Y_i \sim N(\beta, 1)$

$\beta \sim N(0, 1)$

$$P(\beta \mid y) = \frac{P(y \mid \beta)\, P(\beta)}{\int d\beta \text{ '' } \quad \text{''}} \propto_\beta P(y \mid \beta)\, P(\beta)$$

$$\hat{\beta}^{MAP} = \underset{\beta}{\operatorname{argmax}} \log P(\beta \mid y) = \underset{\beta}{\operatorname{argmin}} \left[-\log P(y \mid \beta)\, P(\beta) \right]$$

$$= \underset{\beta}{\operatorname{argmin}} - \log \left[\prod_i \exp\left(-\tfrac{1}{2}(y_i - \beta)^2\right) \right] \exp\left(-\tfrac{1}{2}\beta^2\right)$$

$$= \underset{\beta}{\operatorname{argmin}} \underbrace{\sum_i (y_i - \beta)^2}_{SSE} + \underbrace{\beta^2}_{\ell_2\text{-reg.}}$$

Normal-normal conjugacy

$$p(\mu \mid y) \propto_\mu \left[\prod_i \mathcal{N}(y_i ; \mu, \sigma^2)\right] \mathcal{N}(\mu ; \mu_0, \sigma_0^2)$$

$$\propto_\mu \left[\prod_i \exp\left(-\frac{1}{2\sigma^2}(y_i - \mu)^2\right)\right] \exp\left(-\frac{1}{2\sigma_0^2}(\mu_0 - \mu)^2\right)$$

$$\propto_\mu \exp\left(-\frac{1}{2}\left[\frac{n\mu^2}{\sigma^2} - \frac{2\mu \sum_i y_i}{\sigma^2}\right] + \left[\frac{\mu^2}{\sigma_0^2} - \frac{2\mu\mu_0}{\sigma_0^2}\right]\right)$$

$$\exp\left(-\frac{1}{2}\left[\mu^2 \underbrace{\left[\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right]}_{\triangleq J_n} - 2\mu \underbrace{\left[\frac{y_\cdot}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}\right]}_{\triangleq h_n}\right]\right)$$

$$\propto_\mu \exp\left(-\frac{1}{2}\left[\mu^2 J_n - 2\mu h_n\right]\right)$$

"complete the square"

$$\mu^2 J_n - 2\mu h_n = \left(\mu J_n^{1/2} - J_n^{-1/2} h_n\right)^2 + \text{cons}_\mu$$

$$= \left(\mu^2 J_n - 2\mu h_n + J_n h_n^2\right) - J_n h_n^2$$

$$\propto_\mu \exp\left(-\frac{1}{2}\left(\mu J_n^{1/2} - J_n^{-1/2} h_n\right)^2\right)$$

$$\propto_\mu \exp\left(-\frac{1}{2}\left(\frac{\mu - \overbrace{J_n^{-1} h_n}^{\triangleq \mu_n}}{\underbrace{J_n^{-1/2}}_{\triangleq \sigma_n}}\right)^2\right)$$

posterior params

$$\propto_\mu \mathcal{N}(\mu ; \mu_n, \sigma_n^2)$$

Definition of "kernel" (unnormalized density)

$$P(\mu \mid y) = \frac{f(\mu \mid y)}{\int f(\mu \mid y)\, d\mu}$$

$$P(\mu ; \mu_0, \sigma_0) = \frac{1}{\sqrt{2\pi}\,\sigma_0} \exp\left(-\frac{1}{2}\left(\frac{\mu - \mu_0}{\sigma_0}\right)^2\right)$$

$$= \int \ldots \, d\mu = 1$$

$$= \int \underbrace{\exp\left(-\frac{1}{2}\left(\frac{\mu-\mu_0}{\sigma_0}\right)^2\right)}_{\triangleq f(\mu)} d\mu = \underbrace{\sqrt{2\pi}\,\sigma_0}_{Z}$$

$$\mathcal{N}(\mu ; \mu_0, \sigma_0) = \frac{f(\mu)}{\int f(\mu)\, d\mu}$$

Posterior predictive in normal-normal model

$$p(y_{n+1} \mid y_{1:n}) = \int p(y_{n+1}, \mu \mid y_{1:n})\, d\mu$$

$$= \int \underbrace{p(y_{n+1} \mid \mu)}_{\text{likelihood}} \underbrace{p(\mu \mid y_{1:n})}_{\text{posterior}}\, d\mu$$

$$= \int \mathcal{N}(y_{n+1}; \mu, \sigma^2)\, \mathcal{N}(\mu; \mu_n, \sigma_n^2)\, d\mu$$

$$= \mathcal{N}(y_{n+1}; \mu_n, \sigma^2 + \sigma_n^2)$$

$$y \sim \mathcal{N}(\mu, \sigma) \quad || \rightarrow \quad \tilde{y} = y - \mu$$

$$\mu \sim \mathcal{N}(\mu_0, \sigma_0)$$

$$\tilde{y} \sim \mathcal{N}(0, \sigma)$$

$$\mu \sim \mathcal{N}(\mu_0, \sigma_0)$$

$$y = \tilde{y} + \mu$$

$$\mathbb{E}[\tilde{y} + \mu]$$

$$= \mathbb{E}[\tilde{y}] + \mathbb{E}[\mu]$$

Gamma-normal conjugacy

$$P(\tau \mid y, \mu, a, b) \propto_\tau \text{Gamma}(\tau; a, b) \prod_i \mathcal{N}(y_i; \mu, \tau^{-1})$$

$$\propto_\tau \underbrace{\tau^{a-1} \exp(-b\tau)}_{\text{kernel of gamma prior}} (\sqrt{\tau})^n \exp\left(-\frac{\tau}{2} \sum_i (y_i - \mu)^2\right)$$

$$\propto_\tau \underbrace{\tau^{a+\frac{n}{2}-1} \exp\left(-\tau\left(b + \frac{1}{2}\sum_i (y_i - \mu)^2\right)\right)}_{\text{kernel of a gamma}}$$

$$\propto_\tau \text{Gamma}\left(\tau; a + \frac{n}{2}, b + \frac{1}{2}\sum_i (y_i - \mu)^2\right)$$

Graphical model

$$\tau \sim \Gamma(a, b)$$

$$\mu \sim \mathcal{N}(\mu_0, (\kappa_0 \tau)^{-1})$$

$$Y_i \sim \mathcal{N}(\mu, \tau^{-1})$$