Parcial 1 - Cristian Camilo Osorio Mesa

Mínimos cuadrados (OLS)

$$t_n = \phi(x_n) w^T + \eta_n$$
$$\eta_n \sim \mathcal{N}(0, \sigma_\eta^2)$$

$$\mathcal{L}(w) = \sum_{n=1}^{N} (t_n - w^T \phi(x_n))^2 = \|t - \Phi w\|^2$$

$$w_{OLS} = \min_w \mathcal{L}(w) \Rightarrow \nabla_w \mathcal{L} = 0$$

$$\nabla_w \mathcal{L} = -2\Phi^T (t - \Phi w) = 0$$
$$2\Phi^T w \Phi = 2\Phi^T t$$

$$w_{OLS} = (\Phi^T \Phi)^{-1} \Phi^T t$$

Mínimos cuadrados regularizados (Ridge)

$$t_n = \phi_n(x_n) w^T$$

$$\mathcal{L}(w) = \sum_{n=1}^{N} (t_n - w^T \phi(x_n))^2 + \lambda \|w\|^2$$

Penalización L_2

$$\mathcal{L}(w) = \|t - \Phi w\|^2 + \lambda \|w\|^2$$

$$w_{Ridge} = \min_w \mathcal{L}(w) \Rightarrow \nabla_w \mathcal{L}(w)$$

$$\nabla_w \mathcal{L} = -2\Phi^T (t - \Phi w) + 2\lambda w = 0$$
$$2\Phi^T t = 2\Phi^T w \Phi + 2\lambda w$$

$$w_{Ridge} = (\Phi^T \Phi + \lambda I) \Phi^T t$$

Máxima verosimilitud (MLE)

$$p(t|w) = \prod_{n=1}^{N} \mathcal{N}(t_n | w^T \phi(x_n), \sigma_\eta^2)$$

$$\mathcal{L}(w) = -\frac{1}{2\sigma_\eta^2} \|t - \Phi w\|^2$$

$$w_{MLE} = \max_w \mathcal{L}(w)$$

$$\log p(t|w) = -\frac{1}{2\sigma_\eta^2} \sum_{n=1}^{N} (t_n - w^T \phi(x_n))^2$$

$$w_{MLE} = w_{OLS} = (\Phi^T \Phi)^{-1} \Phi^T t$$

Máximo a posterior (MAP)

$$t_n = w^T \phi_n(x_n) + \eta_n$$
$$w \sim \mathcal{N}(0, \tau^2 I)$$

$$p(w|t) \propto p(t|w) p(w)$$

$$\log p(w|t) = -\frac{1}{2\sigma_\eta^2} \|t - \Phi w\|^2 - \frac{1}{2\tau^2} \|w\|^2 - C$$

$$w_{MAP} = \max_w \log p(w|t)$$

$$w_{MAP} = \left(\Phi^T \Phi + \frac{\sigma_\eta^2}{\tau^2} I\right) \Phi^T t$$

$$w_{MAP} = w_{Ridge} \; ; \; \sigma_\eta = \tau$$

Modelo Bayesiano lineal gausiano

$$t_n = w^T \phi(x_n) + \eta_n$$
$$\eta_n \sim \mathcal{N}(0, \sigma_\eta^2)$$

$$w \sim \mathcal{N}(0, \tau^2 I)$$

$$p(w|t) = \mathcal{N}(w | \mu_w, \Sigma_w)$$

$$\mu_w = \frac{1}{\sigma_\eta^2} \Sigma_w \Phi^T t \qquad \Sigma_w^{-1} = \frac{1}{\sigma_\eta^2} \Phi^T \Phi + \frac{1}{\tau^2} I$$

Regresión Rígida kernel (Kernel Ridge)

$$k(x, x') = \phi(x)^T \phi(x')$$
$$w = \Phi^T a$$

$$f(x) = \sum_{i=1}^{N} a_i k(x, x_i)$$

$$\min_a \|t - Ka\|^2 + \lambda a^T K a$$

$$a = (K + \lambda I)^{-1} t$$

Procesos Gaussianos (GP)

$$f(x) \sim GP(0, k(x, x'))$$

$$X = [x_1, \dots, x_N]$$
$$f = [f(x_1), \dots, f(x_N)]^T$$
$$f \sim \mathcal{N}(0, K)$$
$$t|f \sim \mathcal{N}(f, \sigma_\eta I)$$

$$p(t_* | t, x_*) = \mathcal{N}(t_* | \mu_*, \sigma_*^2)$$
$$\mu_* = k_*^T (K + \sigma_\eta^2 I)^{-1} t$$
$$\sigma_*^2 = k(x, x') - k_*^T (K + \sigma_\eta^2 I)^{-1} k_*$$