Solucion punto 1

Modelo de regresion:

$$t_n = \phi(X_n) W^T + n_n \quad \text{con} \quad \{t_n \in \mathbb{R}, X_n \in \mathbb{R}^P\}_{n=1}^{N}$$

$$W \in \mathbb{R}^Q ; \ \phi \mathbb{R}^P \to \mathbb{R}^Q , \ Q \geq P ; \ n_n \sim N(n_n | 0, \sigma_n^2)$$

- Minimos cuadrados

Modelo vectorizado

$$t = \phi w + n$$

F. Objetivo

$$L(W) = \sum_{n=1}^{N} (t_n - \phi(X_n)^T W)^2 - \| t - \phi W \|^2$$

$$L(W) = (t - \phi W)^T (t - \phi W) = t^T t - 2t^T \phi W + W^T \phi^T \phi W$$

$$\nabla_W L(W) = -2\phi^T t + 2\phi^T \phi W$$

$$\phi^T \phi W = \phi^T t$$

$$W^* = (\phi^T \phi)^{-1} \phi^T t \qquad \text{Si } \phi^{-T}\phi \text{ es invertible}$$

- Minimos regularizados

Modelo vectorizado

$$t = \phi w + n \qquad n \sim N(0, \sigma_n^2 I)$$

Primavera

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

$$L(w) = (t - \Phi w)^T (t - \Phi w) + \lambda w^T w$$

$$\nabla_w = -2\Phi^T t + 2\Phi^T \Phi w + 2\lambda w$$

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

$$(\Phi^T \Phi + \lambda I) w = \Phi^T t$$

$$w^* = (\Phi^T \Phi + \lambda I)^{-1} \Phi^T t$$

- Maxima verosimilitud

$$t_n = \Phi(x_n)^T w + n_n$$

$$p(t_n | x_n, w, \sigma^2) = \mathcal{N}(t_n | \Phi(x_n)^T w, \sigma^2)$$

Verosimilitud conjunta

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

$$p(t | w, \sigma^2) = \mathcal{N}(t | \Phi w, \sigma^2 I)$$

Log - Verosimilitud

$$\log p(t | w, \sigma^2) = -\frac{N}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \|t - \Phi w\|^2$$

Estimador de maxima verosimilitud para w

$$w_{ml} \rightarrow \arg\min_w \|t - \Phi w\|^2$$

$$l(w) = \|t - \Phi w\|^2 = (t - \Phi w)^T (t - \Phi w)$$

$$l(w) = T^T t - 2t^T \Phi w + w^T \Phi^T \Phi w$$

$$\nabla_{wL} = -2\Phi^T t + 2\Phi^T \Phi w = 0$$

$$\Phi^T \Phi w = \Phi^T t$$

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

Estmacon de σ^2 por ML

$$\frac{\partial}{\partial \sigma^2} \log p(t \mid w, \sigma^2) = \frac{-N}{-2\sigma^2} + \frac{1}{2\sigma^4} \|t - \Phi w\|^2 = 0$$

$$\sigma^2_{ML} = \frac{1}{N} \|t - \Phi w_{ML}\|^2$$

Maximo aposterior

$$p(t \mid w) = N(t + \Phi w, \sigma^2 I)$$

$$p(w) = N(w \mid 0, \sigma_w^2 I)$$

posterior w

$$p(w \mid t) \propto p(t \mid w)\, p(w)$$

log posterior

$$\log p(w \mid t) = \log_p (t \mid w) + \log p(w) + cte$$

$$\log p(t \mid w) = \frac{1}{2\sigma_w^2} \|t - \Phi w\|^2 + cte$$

$$\log p(w) = \frac{-1}{2\sigma^2_w} \|w\|^2 + cte$$

$$\log p(w|t) = -\frac{1}{2\sigma^2} \|t - \phi w\|^2 - \frac{1}{2\sigma^2_w} \|w\|^2 + cte$$

Estimador MAP

$$L_{MAP}(w) = \|t - \phi w\|^2 + l\|w\|^2 \; ; \; l = \frac{\sigma^2}{\sigma^2_w}$$

$$L(w) = t^T t - 2t^T \phi w + w^T \phi^T \phi w + l w^T w$$

$$\nabla L = -2\phi^T t + 2(\phi^T \phi + lI) w = 0$$

$$(\phi^T \phi + lI) w = \phi^T t$$

$$w = (\phi^T \phi + lI)^{-1} \phi^T t \qquad l = \frac{\sigma^2}{\sigma^2_w}$$

Bayesiano con modelo lineal Gaussiano

Prior sobre los parametros

$$p(w) = N(w | 0, \alpha^{-1} I)$$

Verosimilitud

$$p(t|w) = N(t | \phi w, \sigma^2 I) \quad ó \quad N(t | \phi w, \beta^{-1} I)$$

$$\beta = \frac{1}{\sigma^2}$$

Posterior de w

$$p(w|t) = \frac{p(t|w)\,p(w)}{p(t)}$$

Como ambos son Gaussianos, su posterior también.

$$p(w|t) = N(w|m_N, S_N)$$

$S_N^{-1} = \alpha I + \beta \Phi^T \Phi \longrightarrow$ Presición Posterior

$m_N = \beta S_N \Phi^T t \longrightarrow$ Media posterior.

Predicción Bayesiana.

$$p(t_* | x_*, t) = \int p(t_* | x_*, w)\, p(w|t)\, dw$$

$$p(t_* | x_*, t) = N(t_* | \mu_*, \sigma_*^2)$$

$\mu_* = \Phi(x_*)^T m_N \longrightarrow$ Media de predicción

$\sigma_*^2 = \Phi(x_*)^T S_m \Phi(x_*) + \beta^{-1} \longrightarrow$ Varianza predictiva

Regresión rígida Kernel

$$f(x) = \Phi(x)^T w$$

$$\min_w \|t - \Phi w\|^2 + l\|w\|^2$$

$$w = (\Phi^T \Phi + lI)^{-1} \Phi t$$

$$w = \sum_{n=1}^{N} a_n \Phi(x_n) = \Phi^T a \qquad a \in R^N$$

$$f(x) = \phi(x)^T \phi^T a = \sum_{n=1}^{N} a_n \underbrace{\phi(x)^T \phi(x_n)}_{K(x, x_n)} =$$

$$= \sum_{n=1}^{N} a_n K(x, x_n)$$

$$l(a) = \|t - \phi\phi^T a\|^2 + l\|\phi a\|^2 = \|t - Ka\|^2 + l a^T K a$$

$$\nabla_m l = -2K(t - Ka) + 2lKa = -2Kt + 2K(K + lI)a = 0$$

$$(K + lI)a = t$$

$$a = (K + lI)^{-1} t$$

$$K = \phi\phi^T \in \mathbb{R}^{N \times N}$$

$$K_{ij} = K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$$

Procesos Gaussianos

$$f(x) = \phi(x)^T w$$

$$E[f(x)] = 0$$

$$Var[f(x)] = \sigma_w^2 \|\phi(w)\|^2$$