Parcial TAM

①

Modelo de regresión

$t_n = \phi(x_n)w^T + \eta_n$ con $\{t_n \in \mathbb{R}, x_n \in \mathbb{R}^P\}_{n=1}^N$ $w \in \mathbb{R}^Q$ $\phi\ \mathbb{R}^P \to \mathbb{R}^Q$ $Q \geq P$ $\eta_n \sim N(\eta_n | 0, \sigma_n^2)$

- Minimos cuadrados

Modelo vectorizado

$$t = \Phi w + \eta$$

F. objetivo

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

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

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

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

$$w^* = (\Phi^T \Phi)^{-1} \Phi^T t \quad \text{si } \Phi^T \Phi \text{ es invertible}$$

- Minimos cuadrados regularizados

Modelo vectorizado

$$t = \Phi w + \eta \quad n \sim N(0, \sigma_n^2 I)$$

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

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

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

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

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

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

- Maxima verosimilitud

$$t_n = \phi(x_n)^T w + \eta_n$$

$$p(t_n | x_n, w, \sigma^2) = N(t_n | \phi(x_n)^T w, \sigma^2)$$

Verosimilitud conjunta

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

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

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

$$W_{ML} = \underset{w}{\arg\min}\ \|t - \Phi w\|^2$$

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

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

$$\nabla_w L = -2\Phi^T t + 2\Phi^T \Phi w = 0$$

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

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

Estimación de σ^2 por ML

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

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

- Maximo aposteriori

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

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

Posterior W

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

Log posterior

$$\log p(w|t) = \log p(t|w) + \log p(w) + Cte$$

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

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

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

Estimador MAP

$$L_{MAP}(w) = \|t - \Phi w\|^2 + \lambda\|w\|^2 \qquad \lambda = \frac{\sigma^2}{\sigma_w^2}$$

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

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

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

$$w = (\Phi^T \Phi + \lambda I)^{-1} \Phi^T t \qquad \lambda = \frac{\sigma^2}{\sigma_w^2}$$

- Bayesiano con modelo lineal Gaussiano

Prior sobre los parametros

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

Verosimilitud

$$p(t \mid w) = N(t \mid \Phi w, \sigma^{2} I) \quad ó \quad N(t \mid \Phi w, \beta^{-1} I) \qquad \beta = \frac{1}{\sigma^2}$$

Posterior de w

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

Como ambos son gaussianas la posterior también

$$p(w \mid t) = N(w \mid m_N, S_N)$$

$S_N^{-1} = \alpha I + \beta \Phi^T \Phi$ precisión posterior

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

Predicción bayesiana

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

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

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

$\sigma_*^2 = \phi(x_*)^T S_N \phi(x_*) + \beta^{-1}$ Varianza predictiva

- Regresion rigida Kernel

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

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

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

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

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

$$L(a) = \|t - \Phi \Phi^T a\|^2 + \lambda \|\Phi a\|^2 = \|t - Ka\|^2 + \lambda a^T K a$$

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

$$\nabla_a L = -2K(t - Ka) + 2\lambda K a = -2Kt + 2K(K + \lambda I) a = 0$$

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

$$(K + \lambda I) a = t$$

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

- Procesos Gaussianos

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

$$E[f(x)] = 0 \qquad Var[f(x)] = \sigma_w^2 \|\phi(w)\|^2$$

para dos entrados x, x'

$$\mathrm{Cov}(f(x), f(x')) = \sigma_w^2 \phi(x)^T \phi(x') = k(x, x')$$

$$f(\cdot) = \mathcal{GP}(0, k(\cdot,\cdot))$$

Como

$$t_n = f(x) + \eta_n \qquad t = [t_1, \dots, t_N)$$

$$t \sim N(0, K + \sigma_n^2 I)$$

Predicción

$$p(t_* | x_*, X, t) = N(\mu_*, \sigma_*^2) \qquad k_* = [k(x_*, x_1) \dots, k(x_*, x_N)]^T$$

$K = N \times N$

$k_{**} = k(x_*, x_*)$

$\mu_* = k_*^T (K + \sigma_n^2 I)^{-1} t$ Media de la predicción

$\sigma_*^2 = k_{**} - k_*^T (K + \sigma_n^2 I)^{-1} k_*$ Varianza predictiva