1) $t_n = \phi(\mathbb{X}_n)\mathbb{W}^T + \eta_n$

con $\{t_n \in \mathbb{R}, \mathbb{X}_n \in \mathbb{R}^P\}_{n=1}^N$, $\mathbb{W} \in \mathbb{R}^Q$

$\phi : \mathbb{R}^P \to \mathbb{R}^Q$

$Q \geq P$

$\eta_n \sim N(\eta_n | 0, \sigma_n^2)$

$t_n = \Phi W + \eta_n \quad \therefore \Phi \in \mathbb{R}^{N \times Q}$, $\eta \sim N(0, \sigma_n^2 I)$

Mínimos Cuadrados

$$\min_W J(w) = \frac{1}{2}\|t - \Phi w\|_2^2$$

Solución mínimos cuadrados

$$J(w) = \frac{1}{2}(t - \Phi w)^T (t - \Phi w)$$

$$J(w) = \frac{1}{2}(t^T t - 2w^T \Phi^T t + w^T \Phi^T \Phi w)$$

$$\frac{\partial J(w)}{\partial w} = \frac{1}{2}(-\cancel{2}\Phi^T t + \cancel{2}\Phi^T \Phi w) = 0$$

$$= \Phi^T t + \Phi^T \Phi W$$

$$\Phi^T \Phi W = \Phi^T t$$

$$W^* = (\underbrace{\Phi^T \Phi}_{})^{-1} \Phi^T t$$

Invertible

Mínimos cuadrados regularizados

$$\min_{w} J_\lambda(w) = \frac{1}{2}\|t - \Phi w\|_2^2 + \frac{1}{2}\lambda\|w\|_2^2$$

Solución mínimos cuadrados regularizados

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

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

$$\frac{\partial J_\lambda(w)}{\partial w} = \cancel{\frac{1}{2}}\left[-\cancel{2}\Phi^T t + \cancel{2}\Phi^T \Phi w + \cancel{2}\lambda w\right] = 0$$

$$= -\Phi^T t + \Phi^T \Phi w + \lambda w = 0$$

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

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

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

Máxima verosimilitud

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

$$P(t_n | x_n, w, \sigma_n^2) = \mathcal{N}(t_n | \phi(x_n) w^T, \sigma_n^2)$$

Para los datos $\{(x_n, t_n)\}_{n=1}^{N}$ ∴ i.i.d

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

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

Solución máxima verosimilitud

$$\max_{w} \log p(t \mid \Phi, w, \sigma_n^2) \Rightarrow \min_{w} \|t - \Phi w\|^2$$

Máximo a posteriori

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

$$p(t) \mid \Phi, w) = \frac{1}{(2\pi\sigma_n^2)^{N/2}} \exp\left(-\frac{1}{2\sigma_n^2} \|t - \Phi w\|^2\right)$$

Prior gaussiano

$$p(w) = \mathcal{N}(w \mid 0, \alpha^{-1} I) = \frac{\alpha^{Q/2}}{(2\pi)^{Q/2}} \exp\left(-\frac{\alpha}{2} w^T w\right)$$

Bayes

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

$$p(w \mid t, \Phi) = \exp\left(-\frac{1}{2\sigma_n^2} \|t - \Phi w\|^2\right) \exp\left(-\frac{\alpha}{2} w^T w\right)$$

$$\log p(w \mid t, \Phi) = -\frac{1}{2\sigma_n^2}\|t - \Phi w\|^2 - \frac{\alpha}{2} w^T w + C$$

Agrupa las constantes que no dependen de w

$$J(w) = \frac{1}{2\sigma_n^2}\|t - \Phi w\|^2 + \frac{\alpha}{2} w^T w$$

multiplico por $2\sigma_n^2$ para simplificar

$$\hat{J}(w) = \|t - \Phi w\|^2 + \lambda w^T w \quad \therefore \quad \lambda = \alpha \sigma_n^2$$

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

$\Downarrow$

Lleva a la solución de mínimos cuadrados legularizados

Regresión Rígida Kernel

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

función kernel

$$K(x_n, x_m) = \langle \phi(x_n), \phi(x_m) \rangle$$

$$t = K w + \eta$$

$$J(w) = \frac{1}{2}\left(\|t - K w\|^2 + \lambda w^T K w\right)$$

$$J(w) = t^T t - 2w^T K t + w^T K^T K w + \lambda w^T K w$$

$$\frac{\partial J(w)}{\partial w} = -2Kt + 2K^T K w + 2\lambda K w = 0$$

$$= -2Kt + 2(K^T K + \lambda K) w = 0$$

$$= (K^T K + \lambda K) w = Kt$$

Solución regresión rígida Kernel

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

Regresión con procesos gaussianos

$$t_n = f(x_n) + \eta_n$$

$\downarrow$ modela mis datos

$$f(x) \sim GP(m(x), K(x, x'))$$

$m(x) \rightarrow$ media del proceso en x $(m(x) = 0)$
$K(x, x') \rightarrow$ función de covarianza

$$K(x_n, x_m) = \mathrm{cov}(f(x_n), f(x_m))$$

Distribución conjunta de las observaciones

$$\begin{pmatrix} f \\ t \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} K(X,X) + \sigma_n^2 I & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) + \sigma_n^2 I \end{pmatrix}\right)$$

$X \rightarrow$ Matriz entrenamiento

X_* → conjunto para predicción

$K(\alpha, \beta)$ → Matriz de covarianzas

Condicional de la distribución del posterior

$$p(f_* \mid X, t, X_*) = \mathcal{N}(\mu_*, \Sigma_*)$$

Solución regresión mediante procesos gaussianos

$$\mu_* = K(X_*, X)(K(X, X) + \sigma_n^2 I)^{-1} t$$

$$\Sigma_* = K(X_*, X_*) - K(X_*, X)(K(X, X) + \sigma_n^2 I) K(X, X_*)$$