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

$$w \in \mathbb{R}^Q \qquad \phi : \mathbb{R}^P \longrightarrow \mathbb{R}^Q \qquad Q \geq P \qquad \text{y}$$

$$\eta_n \sim N(\eta_n \mid 0, \sigma_\eta^2)$$

Presente el problema de optimización (inferencia) para

- Mínimos Cuadrados
- Mínimos Cuadrados Regularizados
- Máxima verosimilitud
- Máximo a Posteriori
- Bayesiano

Mínimos Cuadrados

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

No prior
No Bayes

$$\| t - \phi(x) w \|_2^2 \quad \text{en nuestro caso} \quad J(w) = \| t - \phi w^T \|_2^2$$

$$\| t - \phi w^T \|_2^2 = \langle t_n - \phi w^T, t_n - \phi w^T \rangle$$

$$= (t_n - \phi w^T)^T \cdot (t_n - \phi w^T) = (t_n^T - (\phi w^T)^T) \cdot (t_n - \phi w^T)$$

$$= t_n^T t_n - \underbrace{t_n^T (\phi w^T) - (\phi w^T)^T t_n}_{-2 t_n (\phi w^T)^T} + (w^T \phi)^T (\phi w^T)$$

Minimizar la función de costo $\arg\min(J(\omega))$ Igualar a 0

$$\frac{\partial J(\omega)}{\partial \omega} = t_n{}^T t_n - 2t^T \phi \omega^T + (\omega^T \phi)^T (\phi \omega^T) = 0$$

$$\frac{\partial J(\omega)}{\partial \omega} = t_n{}^T t_n - 2t^T \phi \omega^T + \omega \phi^T \phi \omega^T$$

$$0 - 2t^T \phi + 2\omega \phi^T \phi = 0$$

$$2\omega \phi^T \phi = 2t^T \phi \qquad \boxed{\omega = t^T \phi (\phi^T \phi)^{-1}}$$

Mínimos Cuadrados Regularizados

Función de costo a utilizar $\| y - \phi \omega^T \|_2^2 + I \lambda \|\omega\|_2^2$

$y = t$ en nuestro caso

$\langle a, a \rangle = a^T a$

$$\langle (t - \phi \omega^T), (t - \phi \omega^T) \rangle = (t - \phi \omega^T)^T \cdot (t - \phi \omega^T)$$

$$I \lambda \langle \omega, \omega \rangle = I \lambda (\omega^T \cdot \omega)$$

$$(t - \phi \omega^T)^T \cdot (t - \phi \omega^T) + I \lambda (\omega^T \omega)$$

$$t^T t \underbrace{- t^T \phi \omega^T - t \phi^T \omega}_{-2t^T \phi \omega^T} + \phi^T \omega \phi \omega^T + I \lambda \omega^T + I \lambda \omega$$

$$J(\omega) = t^T t - 2t^T \phi \omega^T + \phi^T \omega \phi \omega^T + I \lambda \omega^T + I \lambda \omega$$

Minimizamos función de costo $J(\omega)$

$$\frac{\partial(J\omega)}{\partial\omega}=0 \Rightarrow 0-2t^T\phi+2\omega\phi^T\phi+2I\lambda\omega=0$$

$$2\omega\phi^T\phi+2I\lambda\omega=2t^T\phi$$

$$\omega\phi^T\phi+I\lambda\omega=t^T\phi$$

$$\omega(\phi^T\phi+I\lambda)=t^T\phi$$

$$\boxed{\omega=t^T\phi\left(\phi^T\phi+I\lambda\right)^{-1}}$$

Máxima Verosimilitud

Verosimilitud Gaussiana = $p(x)=\prod_{n=1}^{N}N(x_n|\mu,\sigma^2)$

Para este caso $p(t_n|\phi\omega^T,\sigma^2)\rightarrow N(t_n|\phi\omega^T,\sigma^2)$

$$\omega_{ML}=\underset{\omega}{\arg\max}\ \log\prod_{n=1}^{N}N(t_n|\phi\omega^T,\sigma^2)$$

Maximizar $\rightarrow$
los pesos

Sabemos lo siguiente $\Rightarrow$ $p(x|y)=\frac{1}{\sqrt{2\pi\sigma^2}}\cdot\exp\left(\frac{\|x-x_n\|_2^2}{2\sigma^2}\right)$

Se reemplaza $\Rightarrow$

$$\log\prod_{n=1}^{N}\frac{1}{\sqrt{2\pi\sigma^2}}\cdot\exp\left(\frac{\|t_n-\phi(x_n)\omega^T\|_2^2}{2\sigma^2}\right)$$

$$\log\left(\prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}}\right) + \cancel{\log}\left[\prod_{n=1}^{N} \cancel{\exp}\left(\frac{\|t_n - \phi(x_n) w^T\|_2^2}{2\sigma^2}\right)\right]$$

$$\log\left(\frac{1}{(2\pi\sigma^2)^{N/2}}\right) - \sum_{n=1}^{N} \frac{\|t_n - \phi(x_n) w^T\|_2^2}{2\sigma^2}$$

$$J(w) = -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\cdot \|t_n - \phi(x_n) w^T\|_2^2$$

$$\frac{\partial}{\partial w}\left(J(w)\right) = 0$$

$$\cancelto{\text{constante}}{-\frac{N}{2}\log(2\pi\sigma^2)} - \frac{1}{2\sigma^2}\cdot \|t_n - \phi(x_n) w^T\|_2^2 = 0$$

$$\frac{1}{2\sigma^2}\cdot\left\langle t_n - \phi(x_n) w^T, t_n - \phi(x_n) w^T\right\rangle = 0$$

$$\frac{1}{2\sigma^2}\cdot \left(t_n - \phi(x_n) w^T\right)^T\cdot\left(t_n - \phi(x_n) w^T\right)$$

$$= \frac{1}{2\sigma^2}\cdot\left(t_n^T - \phi(x_n)^T w\right)\left(t_n - \phi(x_n) w^T\right)$$

$$= \frac{1}{2\sigma^2}\cdot\left(t_n^T t_n - 2t_n^T \phi(x_n) w^T + w\,\phi(x_n)^T \phi(x_n) w\right)$$

$$\frac{1}{\cancel{2}\sigma^2}\cdot\left(-\cancel{2}t_n^T \phi(x_n) + \cancel{2}\phi(x_n)^T \phi(x_n) w\right)$$

$$\frac{1}{\sigma^2}\cdot \phi(x_n)^T \phi(x_n)\, w = t_n^T \phi(x_n)$$

$$\boxed{w = \left(\phi(x_n)^T \phi(x_n)\right)^{-1} t_n^T \phi(x_n)}$$