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

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

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

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 $\arg\min(J(w))$ Igualar a 0 $\rightarrow$

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

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

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

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

Mínimos Cuadrados Regularizados

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

$y = t$ en nuestro caso

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

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

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

$$\underbrace{(t - \phi w^T)^T \cdot (t - \phi w^T)} + I\lambda(w^T w)$$

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

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

Created with Notewise

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

Objetivo: Maximizar la función de verosimilitud (L)

$$\hat{\theta} = \arg\max_{\theta} L(\theta; x)$$

datos

parámetro

$$t_n = \phi(x_n)\omega^T + \eta \qquad \eta = t_n - \phi(x_n)\omega^T$$

$$\eta \sim N(0, \sigma_\eta^2) \rightarrow N(\eta \mid \mu_\eta, \sigma_\eta^2) \quad \mu_\eta = 0$$

Función de verosimilitud para este caso $\Rightarrow$

$$P(t_n \mid \phi(x_n)\omega^T, \sigma_\eta^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(t_n - \phi(x_n)\omega^T)^2}{2\sigma^2}\right)$$

Dado probabilidad conjunta y utilizando Log-verosimilitud ⇒

$$L(\omega) = \prod_{n=1}^{N} p(t_n \mid \phi(x_n)\omega^T, \sigma^2)$$

$$\operatorname{Log} L(\omega) = \sum_{n=1}^{N} \operatorname{Log} p(t_n \mid \phi(x_n)\omega^T, \sigma^2)$$

$$\operatorname{Log} L(\omega) = -\frac{N}{2}\operatorname{Log}(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(t_n - \phi(x_n)\omega^T)^2$$

constante

$$\cancel{-\frac{N}{2}\operatorname{Log}(2\pi\sigma^2)} - \frac{1}{2\sigma^2}\sum_{n=1}^{N} \|t - \phi\omega\|_2^2 \rightarrow -2t^T\phi\omega^T + \omega\phi^T\phi\omega^T$$

Se deriva respecto a ω $\frac{\partial(\omega)}{\partial\omega} = 0$

$$\frac{-1}{2\sigma^2}\sum_{n=1}^{N}(t_n - \phi(x_n)\omega^T)^2$$

$$\frac{-1}{\cancel{2}\sigma^2}\sum_{n=1}^{N}\cancel{2}(t_n - \phi(x_n)\omega^T)(-\phi(x_n))$$

igualar a cero

$$\frac{1}{\sigma^2}\sum_{n=1}^{N}(t_n - \phi(x_n)\omega^T)(\phi(x_n)) = 0$$

$$\sum_{n=1}^{N} t_n\phi(x_n) - \sum_{n=1}^{N}\phi(x_n)\omega^T\phi(x_n) = 0$$

$$\phi(x_n)^2 = \phi(x_n)^T\phi(x_n)$$

$$\sum_{n=1}^{N} t_n\phi(x_n) = \sum_{n=1}^{N}\phi(x_n)\omega^T\phi(x_n)$$

$$t^T\phi = \omega\phi^T\phi$$

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

Máximo a Posteriori (MAP)

$$w_{map} = \arg\max_{w} \; P(w|t) \rightarrow P(t|w)\,P(w) = P(w|t)\,p(t)$$

El posterior
$$p(w|t) = \frac{\underbrace{P(t|w)}_{\text{Likelihood}}\;\underbrace{p(w)}_{\text{Prior}}}{p(t)}$$

Sabemos que

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

$$w_{map} = \arg\max_{w} \; \log\left(\prod_{n=1}^{N} N(t_n \,|\, \phi(x_n)\,w^T, \sigma_\eta^2) \prod_{q=1}^{Q} N(w_q \,|\, 0, \sigma_w^2) \right)$$

$$w_{MAP} = \arg\max_{w} \; -\frac{1}{2\sigma_\eta^2}\,\| t - \phi w^T \|_2^2 - \frac{1}{2\sigma_w^2}\,\| w \|_2^2$$

$$\lambda = \frac{\sigma_\eta^2}{\sigma_w^2} \qquad \boxed{w = t^T \phi\,(\phi^T\phi + I\lambda)^{-1}}$$

Diferencias

Mínimos Cuadrados	Minimiza la suma de los cuadrados de los errores. Sensible al sobreajuste
Mínimos Cuadrados Regularizados	Mínimos Cuadrados Regularizados, penaliza complejidad del modelo
Máxima Verosimilitud	Maximiza función de verosimilitud de los datos dados parámetros, propenso a sobreajuste
MAP	Maximiza probabilidad posterior, considerando prior y likelihood