Modelo de regresión

Sea el modelo de regresión

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

$\phi : \mathbb{R}^P \to \mathbb{R}^Q$, $Q \geq P$ y $\eta_n \sim N(\eta_n | 0, \sigma_n^2)$.

Presente el problema de optimización (inferencia) y la solución del mismo para los modelos

$$\boxed{t_n = \phi(X_n) W^T + \eta_n} \qquad \eta_n \sim N(\eta_n | 0, \sigma^2)$$

Por hipotesis $E[\eta_n] = 0$. y $Var(\eta_n) = \sigma^2$

entonces

$$t_n \sim N((E[\eta_n] ; Var(\eta_n)).$$

$$
\begin{aligned}
E[t_n] &= E[\phi(x_n) w^T + \eta_n] \\
&= E[\phi(X_n) w^T] + \underbrace{E[\eta_n]}_{0} \longrightarrow \text{por hip} \\
&= \phi(X_n) w^T
\end{aligned}
$$

$$
\begin{aligned}
Var(t_n) &= Var(\phi(X_n) w^T + \eta_n) \\
&= Var(\cancel{\phi(x_n) w^T})^{\to 0} + Var(\eta_n) \\
&= \sigma^2 .
\end{aligned}
$$

luego se concluye que $t_n \sim N(\phi(X_n) w^T ; \sigma^2)$

$$f(t) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(t_n - u)^2}{2\sigma^2}}$$

$$f(t_n) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(t_n - \phi(X_n) w^T)^2}{2\sigma^2}}$$

1) $$L(\phi(X_n), w^T / \underset{\sim}{X}) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(t_n - \phi(X_n) w^T)^2}{2\sigma^2}} \cdots \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(t_n - \phi(X_n) w^T)^2}{2\sigma^2}}$$

$$L(\phi(X_n), w^T / \underset{\sim}{X}) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n e^{-\frac{\sum_{n=1}^{N} (t_n - \phi(X_n) w^T)^2}{2\sigma^2}}$$

$$\ell(\phi\cancel{(X_n), w^T / \underset{\sim}{X}}) = n \ln\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right) - \frac{\sum_{n=1}^{N} (t_n - \phi(X_n) w^T)^2}{2\sigma^2}$$

$$\frac{\partial \ell(\phi(X_n), w^T / \underset{\sim}{X})}{\partial w^T} = 0 - \frac{\sum_{n=1}^{N} \cancel{2}(t_n - \phi(X_n) w^T)(-\phi(X_n))}{\cancel{2}\sigma^2}$$

$$\Rightarrow \sum_{n=1}^{N} (t_n - \phi(X_n) w^T)(\phi(X_n)) = 0$$

$$\Rightarrow \sum_{n=1}^{N} (\phi(X_n)^T t_n - \hat{w} \cdot \phi(X_n) \phi(X_n)^T) = 0$$

$$\Rightarrow \sum_{n=1}^{N} \phi(X_n)^T t_n - \sum_{n=1}^{N} \phi(X_n) \phi(X_n)^T \cdot \hat{w} = 0$$

$$\hat{W}_{M.V.} = \frac{\sum_{n=1}^{N} \phi(X_n)^T t_n}{\sum_{n=1}^{N} \phi(X_n) \cdot \phi(X_n)^T}$$

② $$\frac{\partial P(\phi(X_n), W^T / X)}{\partial \phi(X_n)} = 0 - \frac{\sum_{n=1}^{N} \cancel{2}(t_n - \phi(X_n) W^T)(-W^T)}{\cancel{2}\sigma^2}$$

$$\Rightarrow \sum_{n=1}^{N} (t_n W^T - \phi(X_n) W^T \cdot W^T = 0$$

$$\Rightarrow \sum t_n^T \cdot W - \hat{\phi}(X_n) \sum_{n=1}^{N} W \cdot W = 0$$

$$\Rightarrow t_n^T \sum_{n=1}^{N} W = \hat{\phi}(X_n) \sum W^2 = 0$$

$$\hat{\phi}(X_n)_{m.v.} = \frac{\sum t_n^T \cdot W}{\sum W^2}$$

$$\hat{\phi}(X_n)_{m.v.} = \frac{\sum_{n=1}^{N} t_n^T \left(\frac{\sum \phi(X_n)^T t_n}{\sum \phi(X_n)\phi(X_n)^T} \right)}{\sum_{n=1}^{N} \left(\frac{\sum \phi(X_n)^T t_n}{\sum \phi(X_n)\phi(X_n)^T} \right)^2}$$

Por tanto, $\hat{w}$ y $\hat{\phi}(X_n)$ son los estimadores de máxima verosimilitud de W^T y $\phi(X_n)$.

Minimos Cuadrados

$$t_n = \phi(X_n) W^T + \eta_n \qquad \eta_n \sim N(\eta_n | 0, \sigma^2)$$

$$E[\eta_n] = 0$$

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

$$\frac{\partial E}{\partial W^T} = \sum_{n=1}^{N} 2 (t_n - \phi(X_n) W^T)(-\phi(X_n))$$

$$\Rightarrow -2 \sum_{n=1}^{N} (t_n \phi(X_n) - W^T \phi(X_n)^2) = 0$$

$$\Rightarrow \sum_{n=1}^{N} (t_n \phi(X_n) - \hat{W}^T \phi(X_n)^2) = 0$$

$$\Rightarrow \sum_{n=1}^{N} t_n \phi(X_n) - \hat{W}^T \sum_{n=1}^{N} \phi(X_n)^2 = 0$$

$$\hat{W} = \frac{\sum t_n^T \phi(X_n)^T}{\sum_{n=1}^{N} \phi(X_n) \phi(X_n)^T}$$

$$\hat{\phi}(X_n) = \frac{t_n^T \sum_{n=1}^{N} \left(\frac{\sum t_n \phi(X_n)^T}{\sum \phi(X_n) \phi(X_n)^T} \right)}{\sum_{n=1}^{N} \left(\frac{\sum t_n \phi(X_n)^T}{\sum \phi(X_n) \phi(X_n)^T} \right)^2} \checkmark$$

Con la segunda derivada $\hat{W}$ y $\hat{\phi}(X_n)$ son los estimadores de un minimo.

Minimos Cuadrados Regularizados

$$t_n = \phi(X_n) w^T + \eta_n \qquad \eta_n \sim N(\eta_n | 0, \sigma^2)$$

$$l(\phi(x_n), w^T / \underset{\sim}{X}) = -\frac{\sum_{n=1}^{N} (t_n - \phi(x_n) w^T)^2 + \lambda |w|^2}{2\sigma^2}$$

$$= -\frac{\sum 2 (t_n - \phi(x_n) w^T)(-\phi(x_n)) + 2\lambda w}{2\sigma^2}$$

$$\Rightarrow \sum_{n=1}^{N} (t_n - \phi(x_n) w^T) \phi(x_n) + \lambda w = 0$$

$$\Rightarrow \sum_{n=1}^{N} (t_n \phi(x_n)^T - w \phi(x_n) \phi(x_n)^T) + \lambda w = 0$$

$$\Rightarrow \sum_{n=1}^{N} t_n \phi(x_n)^T - w \sum_{n=1}^{N} \phi(x_n) \phi(x_n)^T + \lambda w = 0$$

$$\Rightarrow \sum_{n=1}^{N} t_n \phi(x_n)^T = w \left(\sum_{n=1}^{N} \phi(x_n) \phi(x_n)^T + \lambda I \right)$$

$$\hat{W}_{M.V.} = \sum_{n=1}^{N} t_n \phi(x_n)^T \left(\sum_{n=1}^{N} \phi(x_n) \phi(x_n)^T + \lambda I \right)^{-1}$$

Modelo máximo a-posteriori

$$P(\phi(X_n), W^T, \underset{\sim}{X}) \propto P(t_n / \underset{\sim}{X}, W^T)\ P(W^T)$$

$$\log P(W^T / t_n, \phi(X_n) / \underset{\sim}{X}) \approx l\left(\prod_{n=1}^{N} N(t_n / \phi(X_n) W^T, \underset{\sim}{X}) \prod_{n=1}^{N} N(\eta_n / 0, \sigma_n^2)\right)$$

$$= \underbrace{l\left[\prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{\left(-\frac{(t_n - \phi(X_n W^T))^2}{2\sigma_n^2}\right)}\right] + l\left[\prod_{n=1}^{N} N(\eta_n / 0, \sigma_n^2)\right]}_{*}$$

$$+ l\left(\prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-\|\eta_n - 0\|^2}\right)$$

$$= * + l\left(\prod_{n=1}^{N} \frac{1}{(\sqrt{2\pi\sigma_n^2})^{1/2}}\right) + l\left(e^{-\frac{\Sigma|\eta_n|^2}{2\sigma_n^2}}\right)$$

$$= * + l\left(\frac{1}{(\sqrt{2\pi\sigma_n^2})^{N/2}}\right) + l\left(e^{\frac{\Sigma|\eta_n|^2}{2\sigma_n^2}}\right)$$

$$= * - \frac{N}{2}\left[\log(2\pi) + \log(\sigma_n^2)\right] - \frac{\sum_{n=1}^{N}|\eta_n|^2}{2\sigma_n^2}$$

$$= -\frac{N}{2}\left[\log(2\pi) + \log(\sigma_n^2)\right] - \frac{\|t_n - \phi(X_n) W^T\|^2}{2\sigma\eta_n}$$

$$-\frac{N}{2}\left[\log(2\pi + \log(\sigma_n^2)\right) - \frac{\|\eta_n\|^2}{2\sigma^2}$$

$$= MA - \left[\frac{2\sigma_n^2}{2\sigma_n^2}\|t_n - \phi(X_n) W^T\|^2 + \frac{2\sigma_n^2}{2\sigma_n^2}\|\eta\|^2\right] + cte$$

$$= Min\ \|t_n - \phi(X_n) W^T\|^2 \frac{2\sigma_n^2}{2\sigma_n^2}\|\eta\|^2\Big] + Cte$$

$$\lambda = \frac{\sigma_n^2}{\sigma_n^2} \quad \hat{W}_{MAP} = \left(\phi(X_n)^T \phi(X_n) + \frac{\sigma\lambda I}{N}\right)^{-1} \phi(X_n)^T t_n \rightarrow W_{M.C.R}$$