Demostrar :

La expresión de la probabilidad condicional

$p(t_* \mid f(\mathbf{x}_*), f(\mathbf{X}))$.

Para una nueva muestra $\mathbf{x}_* \in \mathbb{R}^p$, tenemos que :

Datos observados

$$\begin{bmatrix} \mathbf{t} \\ t_* \end{bmatrix} \sim \mathcal{N}\left(\left[\begin{bmatrix} \mathbf{t} \\ t_* \end{bmatrix} \middle| \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right], \begin{bmatrix} \mathbf{K} + \sigma_\varepsilon \mathbf{I} & \mathbf{K}_* \\ \mathbf{K}_*^T & k(\mathbf{x}_*, \mathbf{x}_*) + \sigma_\varepsilon^2 \end{bmatrix} \right)$$

Datos predichos.

con $\mathbf{K}_* = [k(\mathbf{x}_*, \mathbf{X})]$

- La probabilidad condicional $p(t_* \mid f(\mathbf{x}_*), f(\mathbf{X}))$ se puede determinar como :

$$p(t_* \mid f(\mathbf{x}_*), f(\mathbf{X})) = \mathcal{N}(t_* \mid m(\mathbf{x}_*), \operatorname{cov}(f(\mathbf{x}_*), f(\mathbf{X})))$$

con :

$$m(\mathbf{x}_*) = \mathbf{K}_*^T (\mathbf{K} + \sigma_\varepsilon^2 \mathbf{I})^{-1} \mathbf{t}$$

$$\operatorname{cov}(f(\mathbf{x}_*), f(\mathbf{X})) = k(\mathbf{x}_*, \mathbf{x}_*) + \sigma_\varepsilon^2 - \mathbf{K}_*^T (\mathbf{K} + \sigma_\varepsilon^2 \mathbf{I})^{-1} \mathbf{K}_*$$

Para aclarar :

$\mathbf{K}$ es la matriz de covarianza entre los puntos de entrenamiento $\mathbf{X}$

K_* es el vector de covarianzas entre X_* y los puntos en X

$k(X_*, X_*)$ es la varianza del punto X_*

σ_ε^2 es la varianza del ruido gaussiano.

Entonces:

$$p(t_* | f(x_*), f(X)) = \mathcal{N}(t_* | m(x_*), \operatorname{cov}(f(x_*), f(X)))$$

Donde la media y la covarianza están dadas por:

- La media condicional:

$$m(X_*) = K_*^T (K + \sigma_\varepsilon^2 I)^{-1} t$$

- La covarianza condicional:

$$\operatorname{cov}(f(x_*), f(X)) = k(X_*, X_*) + \sigma_\varepsilon^2 - K_*^T (K + \sigma_\varepsilon^2 I)^{-1} K_*$$

Entonces :

Para encontrar la distribución condicional $p(t_* | t)$ aplicamos la fórmula de la distribución condicional de una distribución normal conjunta :

$$\begin{pmatrix} a \\ b \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix}, \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix} \right)$$

La distribución condicional de b dado a es:

$$b|a \sim \mathcal{N}(\mu_b + \Sigma_{ba}\Sigma_{aa}^{-1}(a - \mu_a), \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})$$

para el caso de t y t_*, obtenemos:

- La media condicional:

$$m(\mathbf{x}_*) = \mathbf{K}_*^T (K + \sigma_\varepsilon^2 I)^{-1} t$$

Esto se deriva directamente de la fórmula anterior donde $a = t$ y $b = t_*$, y la media de t_* depende de t.

- La covarianza condicional:

$$\operatorname{Cov}(f(x_*), f(\mathbf{X})) = K(\mathbf{x}_*, \mathbf{x}_*) + \sigma_\varepsilon^2 - K_*^T (K + \sigma_\varepsilon^2 I)^{-1} \mathbf{K}_*$$

Esta es la varianza de t_*, que se ajusta por la información contenida en t.