$t_n = \phi(x_n) w + \eta_n$, datos i.i.d, $\Phi = \begin{bmatrix} \phi(x_1) \\ \phi(x_2) \\ \vdots \\ \phi(x_N) \end{bmatrix}$

Minimos Cuadrados

Problema

$$\arg\min_{w} \|t - \Phi w\|_2^2$$

Solucion

$$\hat{w} = (\Phi^T \Phi)^{-1} \Phi^T t$$

Discucion:

Modelo Frecuentista, no asume ningun prior ni tiene incertidumbre, unica solucion que mejor se ajuste a los datos

Minimos Cuadrados Regularizados:

Problema

$$\arg\min_{w} \|t - \Phi w\|_2^2 + \lambda \|w\|_2^2$$

Solucion

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

Discucion.

Igual que minimos cuadrados, se añade un termino de regularizacion que permite optimizar para el w mas simple!! . Tambien se encarga de que la matriz de valores singulares de la descomposicion SVD no diverga

Maxima Verosimilitud:

Problema

$$\arg\max_{w} \left[-\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma_n^2) - \frac{1}{2\sigma_n^2}\sum_{n=1}^{N} \|t_n - \phi(x_n) w\|_2^2 \right]$$

Solucion

$$\hat{w} = (\Phi^T \Phi)^{-1} \Phi^T t$$

$$\arg\min_{w} \|t - \Phi w^T\|_2^2$$

Discucion (ML)

Que la solucion sea igual a minimos cuadrados es un caso especial. Maxima verosimilitud busca una distribucion en el espacio $\mathbb{R}^Q$ de Φ que maximize la probabilidad de obtener los datos originales. Sigue siendo un modelo frecuentista

Maximo a-posteriori

Problema

$$\arg\min_{w} \left[\frac{1}{2\sigma^2} \sum_{n=1}^{N} (t_n - \phi(x_n) w)^2 + \frac{1}{2\sigma_w^2} \|w\|_2^2 \right]$$

Solucion

$$\hat{w} = (\Phi^T \Phi + \frac{\sigma_n^2}{\sigma_w^2} \mathbb{I})^{-1} \Phi^T t$$

Discucion

Maximo a-posteriori es un modelo bayesiano en este caso que la solucion sea igual a minimos cuadrados regularizados es un caso especial de asumir un prior gaussiano.

MAP no es tan vulnerable al sobreajuste como Max Verosimilitud dado el termino de regularizacion.

Lineal gaussiano

Problema

$$p(t^* | \mathbf{t}, \phi, w) = \int p(t^* | w)\, p(w | \mathbf{t})\, dw$$

Solucion

$$p(t^* | t, w) = \mathcal{N}(t_* | \phi^T(x_*) m_N, \beta^{-1} + \phi^T(x_*) S_N \phi(x_*))$$

Discucion

El modelo lineal gaussiano deja de asumir que el prior $p(w)$ viene de una unica distribucion y ~~se~~ asigna cada parametro como una variable aleatoria propia sobre todo el prior se aplica una gaussiana multivariada. El resultado de este modelo es una "region" de estimacion, en contraste a los valores puntuales de los anteriores modelos