Parcial TAM

1. Sea el modelo de regresión $t_n = \phi(\mathrm{x}_n)\mathrm{w}^\top + \eta_n$, con $\{t_n \in \mathbb{R}, \mathrm{x}_n \in \mathbb{R}^P\}_{n=1}^N$, $\mathrm{w} \in \mathbb{R}^Q$, $\phi : \mathbb{R}^P \to \mathbb{R}^Q$, $Q \geq P$, y $\eta_n \sim \mathcal{N}(\eta_n|0, \sigma_\eta^2)$. Presente el problema de optimización y la solución del mismo para los modelos mínimos cuadrados, mínimos cuadrados regularizados, máxima verosilimitud, máximo a-posteriori, Bayesiano con modelo lineal Gaussiano, regresión rígida kernel y mediante procesos Gaussianos. Asuma datos i.i.d. Discuta las diferencias y similitudes entre los modelos estudiados.

$t_n = \phi(X_n) w^T + \eta_n$ con

$t_n \in \mathbb{R}$

$X_n \in \mathbb{R}^P$

$w \in \mathbb{R}^Q$

$\phi : \mathbb{R}^P \to \mathbb{R}^Q$

$Q > P$

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

Soluciones

- Mínimos Cuadrados

$$MSE = \| t_n - \phi(x_n) w^T \|_2^2$$

tomandolo en forma matricial:

$$MSE = \| t - \phi w^T \|^2 \quad \text{siendo} \begin{cases} t \in \mathbb{R}^{n \times 1} \\ \phi \in \mathbb{R}^{n \times Q} \\ w \in \mathbb{R}^{1 \times Q} \end{cases}$$

$$w^* = \underset{w}{\mathrm{argmin}} \| t - \phi w^T \|_2^2 = \frac{\partial}{\partial w} \| t - \phi w^T \|_2^2 = 0$$

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

$$\xrightarrow{0} \cancel{\frac{\partial}{\partial w} t^T t} - \frac{\partial}{\partial w} 2 t^T \phi w^T + \frac{\partial}{\partial w} w \phi^T w^T = 0$$

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

$$w\phi^T\phi = t^T\phi$$

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

- Mínimos Cuadrados Regularizados

$MSER = \|t - \phi w^T\|^2 + \lambda\|w\|^2$ siendo $\lambda \in \mathbb{R}$

$$w^* = \text{minarg}_w \; (t - \phi w^T)^2 + \lambda w^2$$

$$= \frac{\partial}{\partial w}\left((t - \phi w^T)^2 + \lambda w^2\right) = 0$$

$$\frac{\partial}{\partial w}\left(t^T t - 2t^T\phi w^T + w\phi^T\phi w^T + \lambda w w^T\right) = 0$$

$$-\cancel{2}t^T\phi + \cancel{2}w\phi^T\phi + \cancel{2}\lambda w = 0$$

$$-t^T\phi + w\left(\phi^T\phi + \lambda I\right) = 0$$

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

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

- Máxima Verosimilitud

$$t_n = \phi(x_n) w^T + \eta_n \quad \text{con} \quad \eta_n \sim N(\eta_n | 0, \sigma_n^2)$$

$$t_n \sim N(\phi(x_n) w^T, \sigma^2)$$

$$p(t_n | x_n, w, \sigma^2) = N(t_n | \phi(x_n) w^T, \sigma^2) \text{ Verosimilitud}$$

$$= \prod_{n=1}^{N} N(t_n | w^T \phi(x_n), \sigma^2)$$

$$\ln L(w, \sigma^2) = -\frac{N}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=1}^{N} (t_n - \phi(x_n) w^T)^2$$

$$\max_w \ln L(w, \sigma^2) \Leftrightarrow \min_w \sum_{n=1}^{N} [t_n - \phi(w_n) w^T]^2$$

Aplicando mínimos cuadrados

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

- Maximo a Posteriori

$$w = \arg\max_w P(w | t_n, D) \quad \text{siendo } D = \{(x_n, t_n)\}_{n=1}^{N} \text{ conjunto de datos}$$

$$P(w | t_n, D) = \frac{P(D | t_n, w) P(w)}{P(D | t_n)}$$

$$\ln(p(w | t_n, D)) = \ln(p(D | t_n, w)) + \ln p(w) + cte$$

$$= -\frac{1}{2\sigma^2} \|t - \phi w\|^2 - \frac{N}{2} \ln(2\pi\sigma^2) = \frac{-1}{2\sigma^2} \|t - \phi w\|^2 + cte$$

$$p(w) = N(w \mid 0, \alpha^{-1} I)$$

$$\ln(p(w)) = -\frac{\alpha}{2} w^T w + cte$$

$$\ln(p(w) \mid t_n, D) = -\frac{1}{2\sigma^2} \|t_n - \phi w\|^2 - \frac{\alpha}{2} w^T w + cte$$

$$\arg\min_w -\ln(p(w) \mid t_n, D) = \frac{1}{2\sigma^2} \|t_n - \phi w\|^2 - \frac{\alpha}{2} w^T w + cte$$

$$= \frac{\partial}{\partial w}\left(-\ln(p(w) \mid t_n, D)\right) = 0$$

$$= \frac{1}{\sigma^2} \phi^T (\phi w - D) + \alpha w = 0$$

$$\boxed{w = (\phi^T \phi + \lambda I)^{-1} \phi^T D}$$

- Bayesiano con Modelo Lineal Gaussiano

consideraciones:

w: se considera aleatorio

t_n: se modela como funcion lineal de la entrada X_n mas ruido gaussiano

La Verosimilitud en forma matricial: $p(t \mid w) = N(t \mid \phi w, \sigma^2 I)$

Apriori: $p(w) = N(w \mid 0, \alpha^{-1} I)$

aplicando teorema de Bayes

$$p(t \mid w) \sim \exp\left(\frac{-1}{2\sigma^2} \|t - \phi w\|^2\right)$$

$$P(w) \sim \exp\left(-\frac{\alpha}{2}\|w\|^2\right)$$

$$P(w|t) \propto \exp\left(-\frac{1}{2\sigma^2}\|t-\phi w\|^2 - \frac{\alpha}{2}\|w\|^2\right)$$

$P(w|t) = N(w | m_N, S_N)$ $\therefore$ $W_N \rightarrow$ media posterior

$S_N \rightarrow$ Varianza posterior

$$S_N^{-1} = \alpha I + \frac{1}{\sigma^2}\phi^T\phi \; ; \; m_n = \frac{1}{\sigma^2} S_N \phi^T t$$

$$\boxed{P(t_* | x_*, t) = N(t_* | \mu_*, \sigma_*^2) \rightarrow \mu_* = \phi(x_*)^T m_n \quad \sigma_*^2 = \phi(x_*)^T S_N \phi(x_*) + \sigma^2}$$

- Regresión Rigida con Kernel

el Kernel se define como $K(x, x') = \phi(x)^T \phi(x')$

se asume que w esta en el subespacio generado: $w = \phi^T a$

y ya retomando la forma de RMSE

con $a \in \mathbb{R}^N$ vector de coeficientes a determinar

$$\|t - \phi w\|^2 + \lambda \|w\|^2 = \|t - \phi\phi^T a\|^2 + \lambda \|\phi^T a\|^2$$

$K = \phi\phi^T \in \mathbb{R}^{N \times N}$ cada $K_{ij} = K(x_i, x_j)$

$$\|t - Ka\|^2 + \lambda a^T K a$$

$$\operatorname{argmin}_a \left(\|t - Ka\|^2 + \lambda a^T K a \right) = 0$$

$$K(Ka + \lambda a - t) = 0$$

$$a = (K + \lambda I)^{-1} t$$

La predicción: es $\boxed{Y_* = K_*^T a}$

- Regresion mediante proceso Gausianos

No se busca un vector, se modela toda la funcion $f(x)$ como variable aleatoria, si cualquier conjunto finito de puntos $\{x_1, x_2, \ldots, x_n\}$ genera un vector aleatorio

$f = [f(x_1), f(x_2), \ldots, f(x_n)]^T$ t.q. $f \sim N(\mu, K)$ siendo $\mu \to E[f(x)]$

$K_{ij} = K(x_i, x_j)$

Matriz de covarianza

$D = \{(x_1, t_1), \ldots, (x_n, t_n)\}$ $t_n = f(x_n) + \eta_n$

$$\eta_n \sim N(0, \sigma_n^2) \rightarrow \text{ruido}$$

queremos predecir el valor de $f(x_*)$ para una entrada x_* sabiendo que:

$$\begin{bmatrix} f \\ f_* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(x,x) & K_* \\ K_*^T & K(x_*, x_*) \end{bmatrix}\right)$$

siendo: $f = [f(x_1), \ldots, f(x_n)]$

$$K_* = [K(x_1, x_*), \ldots, K(x_N, x_*)]^T$$

$$K \in \mathbb{R}^{N \times N}$$

$$f_* = f(X_*)$$

$$t \sim N(0, K + \sigma_n^2 I)$$

Se quiere calcular la distribucion posterior de f_* dado t

$$P(f_* \mid t, X, X_*) \sim N(\mu_*, \sigma_*^2) \qquad \mu_* = K_*^T (K + \sigma_n^2 I)^{-1} t$$

$$\sigma_*^2 = K(X_*, X_*) - K_*^T (K + \sigma_n^2 I)^{-1} K_*$$

El proceso Gaussiano elige funciones plausibles dadas las observaciones.

$$f(X_*) \sim N\Big(\underbrace{K_*^T (K + \sigma_*^2 I)^{-1} t}_{\text{media}},\ \underbrace{K(X_*, X_*) - K_*^T (K + \sigma_n^2 I)^{-1} K_*}_{\text{Varianza}}\Big)$$