PARCIAL 1 TAM

1. Teniendo el modelo de regresión

$$t_n = \phi(x_n) w^T + \eta_n,$$

con $\{t_n \in \mathbb{R},\ x_n \in \mathbb{R}^P\}_{n=1}^{N}$

target — vector de entrada (características)

$w \in \mathbb{R}^Q \rightarrow$ parametros (vector de pesos)

$\phi : \mathbb{R}^P \rightarrow \mathbb{R}^Q, \quad Q \geq P$

$\eta_n \sim N(\eta_n | 0, \sigma_n^2) \Rightarrow$ ruido gaussiano

Así tenemos la representación matricial del modelo como:

$$\underbrace{\Phi = \begin{bmatrix} \phi(x_1)^T \\ \phi(x_2)^T \\ \vdots \\ \phi(x_n)^T \end{bmatrix}}_{\text{matriz de diseño } \in \mathbb{R}^{N \times Q}} \qquad \underbrace{t = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{bmatrix}}_{\text{Vector de Salida } \in \mathbb{R}^N} \qquad \underbrace{\eta = \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_N \end{bmatrix}}_{\text{ruido}} \qquad \underbrace{w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_Q \end{bmatrix}}_{\text{Vector de parametros } \in \mathbb{R}^Q}$$

$$t = \Phi w + \eta$$

Solución por mínimos cuadrados

$$J(w) = \sum_{n=1}^{N} (t_n - \phi(x_n)^T w)^2$$

$$= (t - \Phi w)^T (t - \Phi w)$$

$$= t^T t - t^T \Phi w - (\Phi w)^T t + (\Phi w)^T (\Phi w)$$

$$= t^T t - 2 t^T \Phi w + w^T \Phi^T \Phi w$$

derivamos

$$\frac{dJ(w)}{dw} = -2\Phi^T t + 2\Phi^T \Phi w = 0$$

$$\Phi^T \Phi w = \Phi^T t$$

asumiendo que $\Phi^T \Phi$ es invertible $\Rightarrow \boxed{(\Phi^T \Phi)^{-1} \Phi^T t}$

Minimos Cuadrados Regularizados

$$J(\omega) = \sum_{n=1}^{N} (t_n - \phi(x_n)^T \omega)^2 + \lambda R(\omega)$$

Con $\lambda \geq 0$ y $R(\omega) \to$ funcion de penalización:

hiperparametro controla el grado de regularizacion

- Ridge (ℓ_2): $R(\omega) = \|\omega\|_2^2 = \sum_{j=1}^{Q} \omega_j^2$
- Lasso (ℓ_1): $R(\omega) = \|\omega\|_1 = \sum_{j=1}^{Q} |\omega_j|$

Por Ridge

$$J(\omega) = (t - \Phi\omega)^T (t - \Phi\omega) + \lambda \omega\omega$$

$$J(\omega) = t^T t - 2t^T \Phi \omega + \omega^T \Phi^T \Phi \omega + \lambda \omega^T \omega$$

Con gradiente

$$\nabla_\omega J_{ridge} = -2\Phi^T t + 2\Phi^T \Phi \omega + 2\lambda \omega$$

Asi tenemos

$$-2\Phi^T t + 2\Phi^T \Phi \omega + 2\lambda \omega = 0$$

$$-\Phi^T t + \underbrace{(\Phi^T \Phi + \lambda I)\omega}_{\Phi^T t} = 0$$

I: QxQ Identidad

entonces:

$$\boxed{(\Phi^T \Phi + \lambda I)^{-1} \Phi^T t} \quad \text{si} \quad \lambda \geq 0$$

MAXIMA VEROSIMILITUD (MLE)

modelo base: $t_n = \Phi(x_n)^T \omega + \eta_n$; $\eta_n \sim N(0, \sigma_\eta^2)$

función de verosimilitud:

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

Aplicamos Log para simplicar el producto

$$\mathcal{L}_n(\omega) = \ln P(t_n|\omega, \sigma_\eta^2)$$

$$= \sum_{n=1}^{N} \ln\left[\frac{1}{\sqrt{2\pi\sigma_\eta^2}} \exp\left(-\frac{(t_n - \Phi(x_n)^T\omega)^2}{2\sigma_\eta^2}\right)\right]$$

$$= \sum_{n=1}^{N} \left(-\frac{1}{2}\ln(2\pi\sigma_\eta^2) - \frac{(t_n - \Phi(x_n)^T\omega)^2}{2\sigma_\eta^2}\right)$$

$$= -\frac{N}{2}\ln(2\pi\sigma_\eta^2) - \frac{1}{2\sigma_\eta^2}\sum_{n=1}^{N}(t_n - \Phi(x_n)^T\omega)^2$$

problema de optimización:

Maximizamos $\mathcal{L}_n(\omega)$ con respecto a ω, $-\frac{N}{2}\ln(2\pi\sigma_\eta^2) \rightarrow$ contante

entonces es equivalente a

$$\min_{\omega} \mathcal{L}(\omega) = \min_{\omega} \sum_{n=1}^{N}(t_n - \Phi(x_n)^T\omega)^2$$

entonces podemos usar $J(\omega) = \|t - \Phi\omega\|_2^2$

$$=(t - \Phi\omega)^T(t - \Phi\omega) = -2\Phi^T t + 2\Phi^T\Phi\omega = 0$$

$$\boxed{= (\Phi^T\Phi)^{-1}\Phi^T t} \rightarrow \text{igual a minimos cuadrados}$$

MAXIMO A-POSTERIOR (MAP)

$$\hat{\omega}_{MAP} = \arg\max_{w} P(\omega|t)$$

Asumiendo independencia y distribucion gaussiana

$$P(t|\omega) \propto P(t_n | \phi(x_n)^T\omega, \sigma_n^2)$$

$$= (2\pi\sigma_n^2)^{-\frac{N}{2}} \exp\left(-\frac{1}{2\sigma_n^2} \|t - \phi\omega\|_2^2\right)$$

suponiendo prior gaussiano

$$P(\omega) = (2\pi\sigma_\omega^2)^{-\frac{Q}{2}} \exp\left(-\frac{1}{2\sigma_\omega^2} \omega^T\omega\right)$$

no normalizados

$$= \exp\left(-\frac{1}{2\sigma_n^2} \|t - \phi\omega\|_2^2 - \frac{1}{2\sigma_\omega^2} \omega^T\omega\right)$$

Problema de optimizacion

$$\hat{\omega}_{MAP} = \arg\min_{w} \left\{\frac{1}{\sigma_n^2} \|t - \phi\omega\|_2^2 + \frac{1}{\sigma_\omega^2} \omega^T\omega\right\}$$

Como $\lambda = \frac{\sigma_n^2}{\sigma_\omega^2} > 0$ entonces

$$\hat{\omega}_{MAP} = \arg\min_{w} \left\{\|t - \phi\omega\|_2^2 + \lambda \omega\omega\right\}$$

Asi:

$$J(\omega) = \|t - \phi\omega\|_2^2 + \lambda\omega^T\omega$$

$$J(\omega) = t^T t - 2t^T\phi\omega + \omega^T\phi^T\phi\omega + \lambda\omega^T\omega$$ → minimos cuadrados regularizados

entonces $\boxed{(\phi^T\phi + \lambda I)^{-1}\phi^T t}$

Powered by CamScanner

BAYESIANO PARA UN MODELO LINEAL GAUSSIANO

para parametro ω, con distribucion gausiana $p(\omega) = N(\omega | 0, \Sigma_\omega)$

matriz de covarianza

$\Sigma_\omega = \alpha^{-1} I$

Tenemos:

modelo $t_n = \phi(x_n)\omega^T + \eta_n$

$$p(t | X, \omega \sigma_n^2) = N(t | \phi\omega, \sigma_n^2 I)$$

- encontremos la distribucion posterior de ω,

$$p(\omega | t, X, \sigma_n^2) \propto p(t | X, \omega, \sigma_n^2) p(\omega)$$

asi que maximisamos el logaritmo de la probabilidad postario

$$\log p(\omega | t) = \log p(t | X, \omega, \sigma_n^2) + \log(p\omega) +$$

$$= -\frac{1}{2\sigma_n^2} \| t - \phi\omega \|^2 - \frac{1}{2}\omega^T \Sigma_\omega^{-1} \omega$$

asi minimizamos

$$J(\omega) = \frac{1}{2\sigma_n^2} \| t - \phi\omega \|^2 + \frac{1}{2}\omega^T \Sigma_\omega^{-1} \omega.$$

conuina, verosimilitud + la regularizacion con

$$S_N = (S_m^{-1} + \beta \phi^T \phi)^{-1}$$
$$M_N = S_N (S_m^{-1} \mu_m + \beta \phi^T t)$$

prior $\rightarrow$ $\mu_m = 0$ y $S_0 = \alpha^{-1} I$.

$$S_N = (\alpha I + \beta \Phi^T \Phi)^{-1}$$
$$M_N = \beta S_N \Phi^T t = (\alpha \beta^{-1} I + \Phi^T \Phi)^{-1} \Phi^T t$$

siendo $\alpha \beta^{-1} = \alpha \sigma_n^2$ se vuelve maximo a posteriori

REGRESION RIGIDA KERNEL

minimizamos el error cuadratico

$$J(w) = \sum_{n=1}^{N} (t_n - \phi(x_n) w^T)^2 + \lambda \|w\|^2$$

$$w = \sum_{n=1}^{N} \alpha_n \phi(x_n)$$

sustituimos para x

$$f(x) = \sum_{n=1}^{N} \alpha_n \underbrace{K(x_n, x)}_{\phi(x_n)^T \phi x} \rightarrow \text{Kernel}$$

asi problema de optimizacion

$$J(\alpha) = \|t - K\alpha\|^2 + \lambda \alpha^T K \alpha$$ matriz de kernel

minimizamos $J(\alpha)$

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

$$f(x_*) = k(x_*, x)(K + \lambda I)^{-1} t$$

$\sigma_n^2 \rightarrow$ no esta explicita en la prediccion