Probabilidades Bayesianas – Aplicación en Regresores.

- Corroborar nuestras nociones de incertidumbre a partir de evidencia

Thomas Bayes

1701–1761

Thomas Bayes was born in Tunbridge Wells and was a clergyman as well as an amateur scientist and a mathematician. He studied logic and theology at Edinburgh University and was elected Fellow of the Royal Society in 1742. During the 18^{th} century, issues regarding probability arose in connection with gambling and with the new concept of insurance. One particularly important problem concerned so-called inverse probability. A solution was proposed by Thomas Bayes in his paper 'Essay towards solving a problem in the doctrine of chances', which was published in 1764, some three years after his death, in the *Philosophical Transactions of the Royal Society*. In fact, Bayes only formulated his theory for the case of a uniform prior, and it was Pierre-Simon Laplace who independently rediscovered the theory in general form and who demonstrated its broad applicability.

$y = f(x) + \eta.$ $\quad D = \{x \in \mathcal{X}, y \in \mathcal{Y}\}$ $\quad \eta$: ruido

VEROSIMILITUD

$$p(f|D) = \frac{p(D|f)\,p(f)}{p(D)}$$

PRIOR

POSTERIOR

EVIDENCIA

$$p(D) = \int_{f \in \mathcal{F}} p(D,f)\,\partial f = \int_{f \in \mathcal{F}} p(D|f)\,p(f)\,\partial f$$

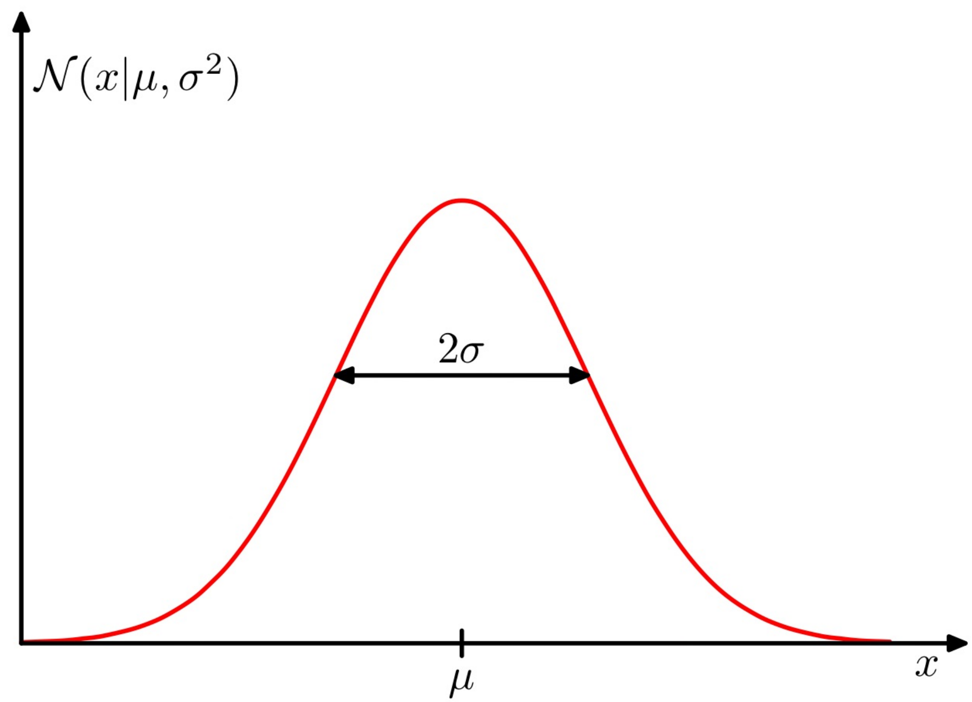
Para emplear apropiadamente Bayes, necesitamos estudiar algunas distribuciones, en especial la Gaussiana.

Distribución Gaussiana Univariada

$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{|x-\mu|^2}{2\sigma^2}\right)$$

$$\mathcal{N}(x|\mu,\sigma^2) \geq 0$$

$$\int_{-\infty}^{\infty} \mathcal{N}(x|\mu,\sigma^2)\, \partial x = 1$$



Ejercicio: Demostrar

$$\mathbb{E}\{x\} = \int_{-\infty}^{\infty} \mathcal{N}(x|\mu,\sigma^2)\, x\, \partial x = \mu$$

$$\mathbb{E}\{x^2\} = \int_{-\infty}^{\infty} \mathcal{N}(x|\mu,\sigma^2)\, x^2 \partial x = \mu^2+\sigma^2$$

$$\mathrm{var}\{x\} = \sigma^2$$

NOTA: $\mathcal{E}\{x\} = \int_{-\infty}^{\infty} x\,p(x)\,\partial x = \mu$; $\quad \text{var}\{x\} = \mathcal{E}\{(x-\mu)^2\} = \int_{-\infty}^{\infty} p(x)(x-\mu)^2\,\partial x$

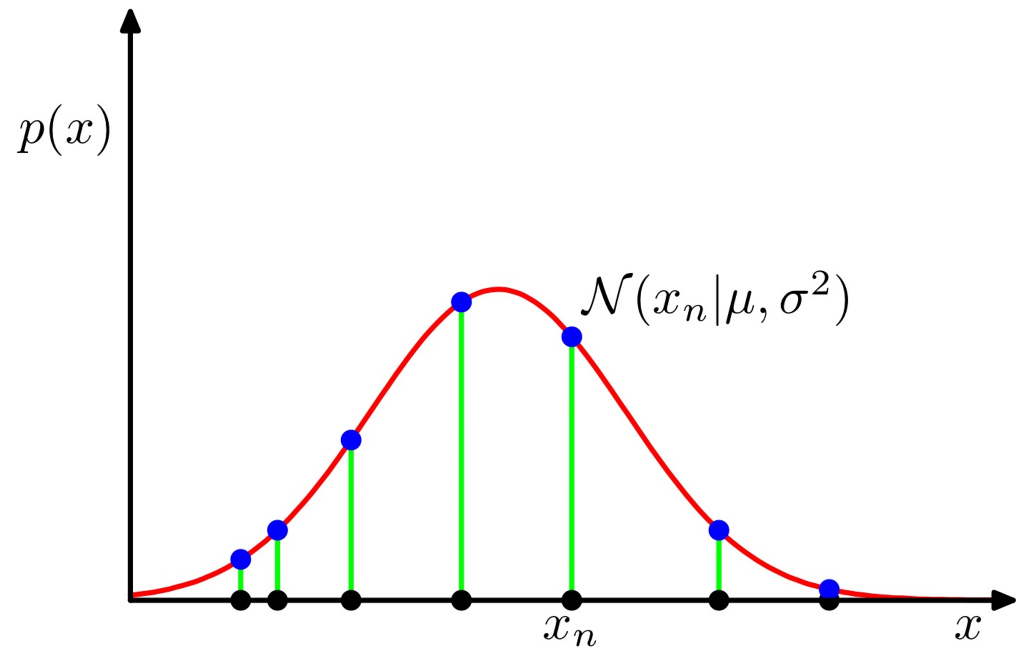
$$\begin{aligned}
\text{var}\{x\} &= \mathcal{E}\{(x-\mu)^2\} = \mathcal{E}\{x^2 - 2x\mu + \mu^2\} \\
&= \mathcal{E}\{x^2\} - 2\,\mathcal{E}\{x\mu\} + \mathcal{E}\{\mu^2\} \\
&= \mathcal{E}\{x^2\} - 2\,\mathcal{E}\{x\}\mu + \mu^2 \\
&= \mathcal{E}\{x^2\} - 2\mu^2 + \mu^2 \\
\text{var}\{x\} &= \mathcal{E}\{x^2\} - \mu^2
\end{aligned}$$

Para un vector de datos:

$\mathbf{x} = \{x_1, x_2, \ldots, x_N\}$

Asumiendo muestras independientes e identicamente distribuidas (i.i.d)

$x_n \sim \mathcal{N}(x_n|\mu, \sigma^2)$



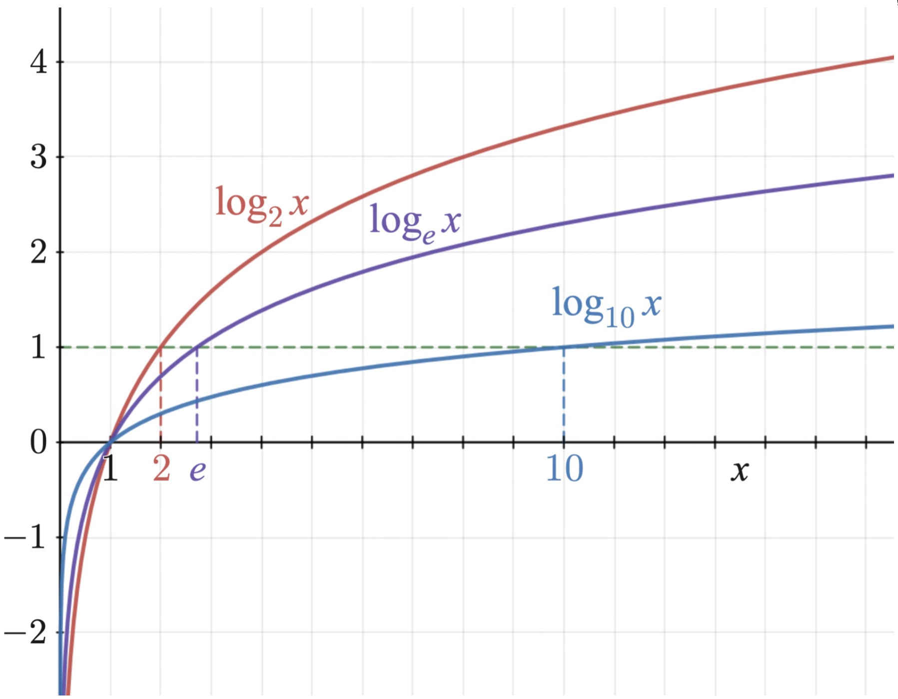
$$p(\mathbf{x}) = p(x_1, x_2, \ldots, x_N) = p(x_1)p(x_2)\cdots p(x_N)$$

$$= \mathcal{N}(x_1|\mu,\sigma^2)\,\mathcal{N}(x_2|\mu,\sigma^2)\cdots\mathcal{N}(x_N|\mu,\sigma^2)$$

$$\boxed{p(\mathbf{x}) = \prod_{n=1}^{N} \mathcal{N}(x_n|\mu,\sigma^2)}$$ Verosimilitud Gaussiana

¿Cómo estimar μ y σ^2?

NOTA: Se pueden emplear log probabilidades para facilitar los cálculos y la inferencia



Función monotónica

$$\log(p(x)) = \log\left(\prod_{n=1}^{N} \mathcal{N}(x_n \mid \mu, \sigma^2)\right)$$

$$= \log\left(\prod_n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{|x_n-\mu|^2}{2\sigma^2}\right)\right)$$

$$= \log\left(\prod_n \frac{1}{\sqrt{2\pi\sigma^2}}\right) + \log\left(\prod_n \exp\left(-\frac{|x_n-\mu|^2}{2\sigma^2}\right)\right)$$

$$= \log\left(\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\right) + \log\left(\exp\left(-\sum_n \frac{|x_n-\mu|^2}{2\sigma^2}\right)\right)$$

$$= -\frac{N}{2}\log(2\pi\sigma^2) - \sum_{n=1}^{N} \frac{|x_n-\mu|^2}{2\sigma^2}$$

$$\log(p(x)) = -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{n=1}^{N} |x_n-\mu|^2$$

Para encontrar μ, σ^2: Máxima log-verosimilitud

ML: Maximum likelihood

$$\mu_{ML} = \arg\max_{\mu} \log(p(\mathbf{x})) = -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_n (x_n - \mu)^2$$

$$\frac{\partial}{\partial \mu}\log(p(\mathbf{x})) = 0 \Rightarrow \frac{\partial}{\partial \mu}\left\{-\frac{1}{2\sigma^2}\sum_n (x_n-\mu)^2\right\}$$

$$\Rightarrow -\frac{1}{2\sigma^2}\sum_n \frac{\partial}{\partial \mu}\left\{(x_n-\mu)^2\right\} = -\frac{1}{2\sigma^2}\sum_n 2(x_n-\mu)(-1)$$

$$\frac{1}{\sigma^2}\sum_n (x_n - \mu) = 0$$

$$\sum_{n=1}^{N} x_n = \sum_{n=1}^{N} \mu = N\mu$$

$$\boxed{\mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} x_n}$$

NOTA: Estimador $\mu_{ML} \Rightarrow$ Estimador media muestral.

EJERCICIO: Demostrar $\sigma^2_{ML} = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu_{ML})^2$

Algunos tips de estimadores:

Un parámetro se puede estimar de muchas formas.

¿Cuál es la mejor?

Sea $\theta \in \mathbb{R}$ el parámetro ideal $\Rightarrow$ CASI NUNCA CONOCIDO

Sea $\hat{\theta} \in \mathbb{R}$ el parámetro estimado

Error cuadrático medio:

$$\mathrm{ecm}(\theta,\hat{\theta}) = \mathcal{E}\{|\theta-\hat{\theta}|^2\} = \mathcal{E}\{|\hat{\theta}-\theta|^2\}$$

$$= \mathcal{E}\{(\hat{\theta}-\mathcal{E}\{\hat{\theta}\}+\mathcal{E}\{\hat{\theta}\}-\theta)^2\}$$

$$= \mathcal{E}\{(\hat{\theta}-\mathcal{E}\{\hat{\theta}\})^2+2(\hat{\theta}-\mathcal{E}\{\hat{\theta}\})(\mathcal{E}\{\hat{\theta}\}-\theta)+(\mathcal{E}\{\hat{\theta}\}-\theta)^2\}$$

$$\mathrm{ecm}(\theta,\hat{\theta}) = \mathcal{E}\{(\hat{\theta}-\mathcal{E}\{\hat{\theta}\})^2\}+2\,\mathcal{E}\{\hat{\theta}\mathcal{E}\{\hat{\theta}\}-\hat{\theta}\theta-\mathcal{E}\{\hat{\theta}\}\mathcal{E}\{\hat{\theta}\}+\mathcal{E}\{\hat{\theta}\}\theta\}$$

$$+\,\mathcal{E}\{(\mathcal{E}\{\hat{\theta}\}-\theta)^2\}$$

$\mathrm{Var}\{\hat{\theta}\} = \mathcal{E}\{(\hat{\theta}-\mathcal{E}\{\hat{\theta}\})^2\}$ $\Rightarrow$ VARIANZA DEL ESTIMADOR

$b(\hat{\theta}) = \mathcal{E}\{\hat{\theta}\}-\theta$ $\Rightarrow$ SESGO DEL ESTIMADOR (BIAS)

$$\text{ecm}(\theta,\hat{\theta}) = \text{var}\{\hat{\theta}\} + b^2(\hat{\theta}) + 2\left[\underline{\mathcal{E}\{\hat{\theta}\}\mathcal{E}\{\hat{\theta}\}} - \underline{\mathcal{E}\{\hat{\theta}\}\theta} - \underline{\mathcal{E}\{\hat{\theta}\}\mathcal{E}\{\hat{\theta}\}} + \underline{\mathcal{E}\{\hat{\theta}\}\theta}\right]$$

$$\boxed{\text{ecm}(\theta,\hat{\theta}) = \text{var}\{\hat{\theta}\} + b^2(\hat{\theta})}$$

EJERCICIO: Sea $x_n = A + w_n$, con $A \in \mathbb{R}$ y $w_n \sim \mathcal{N}(w_n | \mu, \sigma^2)$.

Sea $\hat{A} = \frac{1}{N}\sum_{n=1}^{N} x_n$ ESTIMADOR 1

$\tilde{A} = x_1$ ESTIMADOR 2

¿Cuál es mejor estimador desde $\{x_1, x_2, \ldots, x_N\}$?

DEMUESTRE MATEMÁTICAMENTE SU RESPUESTA

NOTA: $\mathcal{E}\{ax+b\} = a\,\mathcal{E}\{x\} + b \;;\; a,b \in \mathbb{R}$ CTES.

$$\begin{aligned}
\text{var}\{ax+b\} &= \mathcal{E}\{(ax+b-\mathcal{E}\{ax+b\})^2\}. \\
&= \mathcal{E}\{(ax+\underline{b}-a\,\mathcal{E}\{x\}-\underline{b})^2\} \\
&= \mathcal{E}\{(a(x-\mathcal{E}\{x\}))^2\} \\
&= \mathcal{E}\{(a^2(x-\mathcal{E}\{x\})^2\} \\
&= a^2\,\mathcal{E}\{(x-\mathcal{E}\{x\})^2\} \\
\text{var}\{ax+b\} &= a^2\,\text{var}\{x\}
\end{aligned}$$

EJERCICIO: Demuestre que:

$$\mathbb{E}\{\mu_{ML}\} = \mu$$

$$\mathbb{E}\{\sigma^2_{ML}\} = \left(\frac{N-1}{N}\right)\sigma^2$$

con $x_n \sim \mathcal{N}(x_n|\mu,\sigma^2)$

¿Cuál debería ser la corrección sobre σ^2_{ML} para evitar el sesgo del estimador?

$$\mathbb{E}\left\{\frac{1}{N}\sum_n (x_n - \mu_{ML})^2\right\} = \mathbb{E}\left\{\frac{1}{N}\sum_n (x_n^2 - 2x_n\mu_{ML} + \mu_{ML}^2)\right\} \; ; \qquad \mu_{ML} = \frac{1}{N}\sum_n x_n$$

$$= \mathbb{E}\left\{\frac{1}{N}\sum_n x_n^2 - 2\mu_{ML}\frac{1}{N}\sum_n x_n + \frac{1}{N}\sum_n \mu_{ML}^2\right\}$$

$$= \mathbb{E}\left\{\frac{1}{N}\sum_n x_n^2\right\} - 2\,\mathbb{E}\{\mu_{ML}\mu_{ML}\} + \mathbb{E}\left\{\frac{1}{N} N \mu_{ML}^2\right\}$$

$$= \frac{1}{N}\sum_n \mathbb{E}\{x_n^2\} - 2\,\mathbb{E}\{\mu_{ML}^2\} + \mathbb{E}\{\mu_{ML}^2\}$$

$$\mathbb{E}\left\{\frac{1}{N}\sum_n (x_n - \mu)^2\right\} = \frac{1}{N} N\, \mathbb{E}\{x_n^2\} - \mathbb{E}\{\mu_{ML}^2\} = \mathbb{E}\{x_n^2\} - \mathbb{E}\{\mu_{ML}^2\}$$

$$\mathrm{var}\{x\} = \mathbb{E}\{x^2\} - \mathbb{E}^2\{x\} \Longrightarrow \mathbb{E}\{x_n^2\} = \mathrm{var}\{x_n\} + \mathbb{E}^2\{x_n\}.$$

$$x_n \sim \mathcal{N}(x_n \mid \mu, \sigma^2) \qquad \mathbb{E}\{x_n^2\} = \sigma^2 + \mu^2$$

$$\mathbb{E}\{\mu_{ML}^2\} = \mathrm{var}\{\mu_{ML}\} + \mathbb{E}^2\{\mu_{ML}\}$$

$$\mathbb{E}\{\mu_{ML}\} = \mathbb{E}\left\{\frac{1}{N}\sum_n x_n\right\} \qquad \mathbb{E}\{\mu_{ML}^2\} = \frac{\sigma^2}{N} + \mu^2$$

$$= \frac{1}{N}\sum_n \mathbb{E}\{x_n\} = \frac{1}{N}\sum_n \mu = \frac{1}{N} N\mu$$

$$\mathbb{E}\{\mu_{ML}\} = \mu$$

$$\mathrm{var}\{\mu_{ML}\} = \mathrm{var}\left\{\frac{1}{N}\sum_n x_n\right\} = \frac{1}{N^2}\sum_n \mathrm{var}\{x_n\}$$

$$= \frac{1}{N^2} N \sigma^2$$

$$\mathrm{var}\{\mu_{ML}\} = \frac{\sigma^2}{N}$$

$$\mathcal{E}\left\{\frac{1}{N}\sum_n (x_n-\mu_{ML})^2\right\} = \mathcal{E}\{x_n^2\} - \mathcal{E}\{\mu_{ML}^2\} = \sigma^2+\mu^2-\left(\frac{\sigma^2}{N}+\mu^2\right)$$

$$= \sigma^2-\frac{\sigma^2}{N}+\mu^2-\mu^2$$

$$= \sigma^2\left(1-\frac{1}{N}\right)$$

$$\mathcal{E}\left\{\frac{1}{N}\sum_n (x_n-\mu_{ML})^2\right\} = \sigma^2\,\frac{(N-1)}{N}$$

$$\sigma^2_{ML} = \sigma^2\,\frac{(N-1)}{N}$$

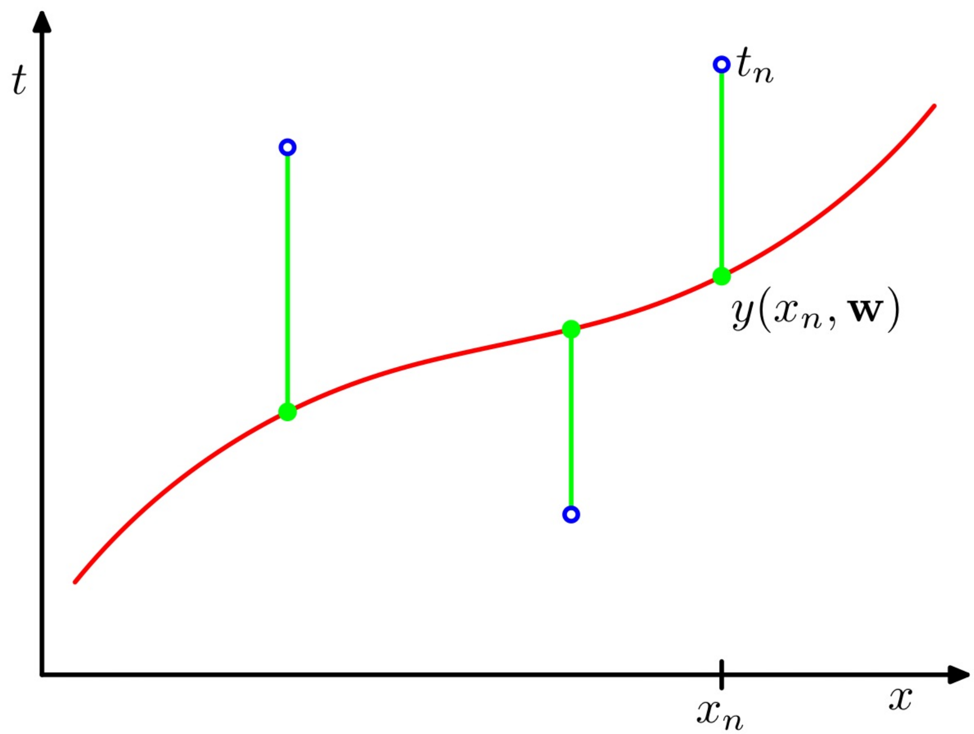
FACTOR DE CORRECCIÓN:

$$\frac{N}{N-1}\sigma^2_{ML} = \sigma^2$$

$$\frac{N}{N-1}\,\frac{1}{N}\sum_n (x_n-\mu_{ML})^2 = \sigma^2$$

$$\frac{1}{N-1}\sum_n (x_n-\mu_{ML})^2 = \sigma^2$$

Ajuste probabilístico Bayesiano → REGRESIÓN



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

REPASO POR MÍN. CUADRADOS.

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

$$\phi(x_n) \in \mathbb{R}^Q$$

$$y_n = \phi(x_n) w \qquad w \in \mathbb{R}^Q$$

$$w^* = \underset{w}{\arg\min}\ \mathcal{E}\{ \| t - \Phi w \|_2^2 \}$$

$$t \in \mathbb{R}^N, \ \Phi \in \mathbb{R}^{N \times Q}.$$

Asumiendo i.i.d.: $\mathcal{E}\{x_n\} \simeq \frac{1}{N} \sum_{n=1}^{N} x_n$ ⇒ MEDIA MUESTRAL

$$\mathcal{E}\{\|t-\phi w\|_2^2\} = \frac{1}{N}\|t-\phi w\|_2^2 = \frac{1}{N}\langle t-\phi w, t-\phi w\rangle$$

$$= \frac{1}{N}\left[t^T t - t^T\phi w - (\phi w)^T t + (\phi w)^T\phi w\right]$$

$$= \frac{1}{N}\left[t^T t - 2t^T\phi w + w^T\phi^T\phi w\right]$$

$$\frac{\partial}{\partial w}\mathcal{E}\{\|t-\phi w\|_2^2\} = 0 \Rightarrow \frac{1}{N}\left[-2(t^T\phi)^T + 2\phi^T\phi w\right] = 0$$

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

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

NOTA: Descomposición espectral.

$$\underset{p\times p}{\phi^T\phi} = V\Delta V^T \rightarrow \text{eig} \qquad V^T V = I$$

$$\underset{N\times p}{\phi} = \underset{N\times N}{U}\,\underset{N\times p}{S}\,\underset{p\times p}{V^T} \rightarrow \text{svd} \qquad U^T U = I$$

$$\Delta = \begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p\end{bmatrix}$$

$$w = (V\Delta V^T)^{-1}(USV^T)^T t = V\Delta^{-1}\underbrace{V^T V}_{I}(US)^T t$$

$$w = V\Delta^{-1}S^T U^T t$$

$$\boxed{S^2 = \Delta}$$

EJERCICIO DEMOSTRAR.

$$S = \left[\begin{array}{cccc|ccc} \sigma_1 & & & & \sigma_{P+1} & & \\ & \sigma_2 & & & & \ddots & \\ & & \ddots & \sigma_P & & & \sigma_N \end{array}\right]$$

$N > P \qquad \mathrm{rank}(\Phi) = \min(N, P)$

$\tilde{U} \in \mathbb{R}^{N \times P}$

$S^2 = \Delta \quad ; \quad S = \Delta^{1/2} \quad ; \quad \Delta^{-1} = S^{-2}$

$$\boxed{w = V S^{-2} S \tilde{U}^T t = V S^{-1} \tilde{U}^T t}$$

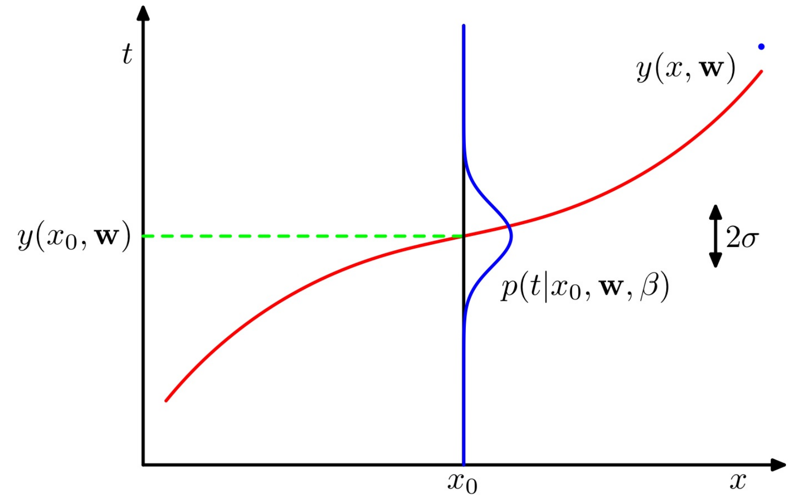
$$w = V \begin{bmatrix} \sigma_1^{-1} & & 0 \\ & \sigma_2^{-1} & \\ 0 & \ddots & \sigma_P^{-1} \end{bmatrix} \tilde{U}^T t$$

EJERCICIO: Repetir para MÍN CUADRADOS REGULARIZADOS.

$$ecmr(t, \Phi w) = \mathcal{E}\{\|t - \Phi w\|_2^2\} + \lambda \|w\|_2^2$$

REGRESIÓN POR MÁX VEROSIMILITUD.

$t_n = f(x_n) + \eta$; $x_n \in \mathbb{R}^P$; $t_n, \eta \in \mathbb{R}$



Asumiendo ruido blanco Gaussiano

$\eta \sim \mathcal{N}(\eta | \mu_n, \sigma_\eta^2)$; con $\mu_n = 0$.

$\eta = t_n - f(x_n)$

donde $f(x_n) = \phi_n w$; $\phi_n \in \mathbb{R}^Q$ $w \in \mathbb{R}^Q$; $\phi_n = \phi(x_n)$; $\phi: \mathbb{R}^P \to \mathbb{R}^Q$.

$$\eta = t_n - \phi_n w \sim \mathcal{N}(\eta | 0, \sigma_\eta^2) = \mathcal{N}(t_n | \phi_n w, \sigma_\eta^2)$$

VEROSIMILITUD $\longleftarrow$ $\mathcal{N}(t_n | \phi_n w, \sigma_\eta^2) = \frac{1}{\sqrt{2\pi\sigma_\eta^2}} \exp\left(-\frac{|t_n - \phi_n w|^2}{2\sigma_\eta^2}\right)$

EJERCICIO: - Asumiendo conjunto de datos i.i.d.
$D = \{x_n \in \mathbb{R}^P, t_n \in \mathbb{R}\}_{n=1}^N$; encuentre w y σ_η^2 que maximizan la log verosimilitud.

- Analice la solución desde descomposición svd.

Regresión por Máximo Aposteriori (MAP)

$t_n = f(x_n) + \eta$. Asumiendo $f(x_n) = \boldsymbol{\phi}_n \mathbf{w}$; $\phi(x_n) = \boldsymbol{\phi}_n$

$x_n \in \mathbb{R}^P$; $\boldsymbol{\phi}_n, \mathbf{w} \in \mathbb{R}^Q$; $t_n \in \mathbb{R}$; $\eta \sim N(\eta | 0, \sigma_\eta^2)$

$w_q \in \mathbf{w}$; $q \in \{1, 2, \ldots, Q\}$; $w_q \sim N(w_q | 0, \sigma_w^2)$

$$p(\mathbf{w} | \mathbf{t}, \boldsymbol{\phi}, \sigma_w^2, \sigma_\eta^2) \propto p(\mathbf{t} | \boldsymbol{\phi}, \mathbf{w}, \sigma_\eta^2)\, p(\mathbf{w} | \sigma_w^2)$$

NOTA: MAP es proporcional al posterior, sin incluir la evidencia.

EJERCICIO: Encuentre los $\mathbf{w}$ que maximizan el log-MAP.

¿Cuál es la relación entre MAP y MÍN CUADRADOS?

¿Qué pasa si se cambia el PRIOR en MAP?

Exprese la solución de MAP mediante descomposición espectral.

Ejercicio: - Consulte qué es la SIGNAL TO NOISE RATIO - SNR y cómo estimar la varianza de ruido blanco Gaussiano desde SNR.

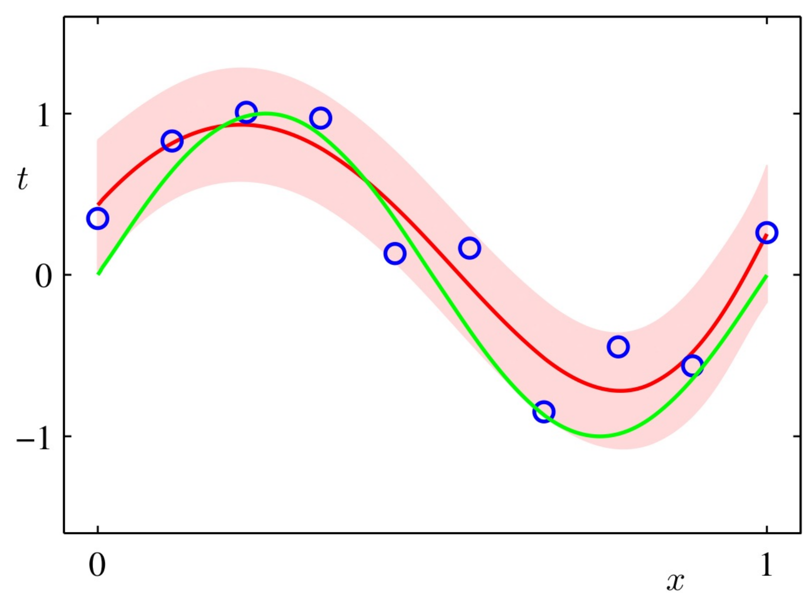
- Genere datos sintéticos a partir de un tono (cos/sin) contaminado con ruido blanco Gaussiano para $SNR \in \{1, 2, 5, 10\}$ dB.
- Implemente una solución en Python que construya regresores por mín cuadrados, log verosimilitud y log MAP sobre los datos simulados; utilizando 80% de los datos para ajustar regresores y 20% para evaluar salida estimada con las incertidumbres respectivas según el caso (ver fig.).

— SOLUCIÓN REGRESOR

o Datos simulados

— TONO SIN RUIDO

LA SOMBRA REPRESENTA SOLUCIÓN ± 1.96 std.



NOTA: Utilizar ϕ polinomial:

$$\phi(x) = [1, x, x^2, \ldots, x^{Q-1}] \in \mathbb{R}^Q$$