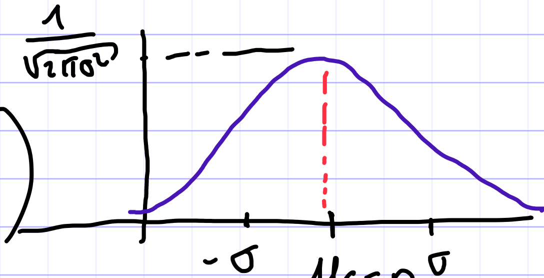
Modelo Bayesiano

$$y = f(x,w) + \epsilon \;; \qquad \epsilon \sim \mathcal{N}(\epsilon \mid 0, \sigma^2)$$

$$\mathcal{N}(\epsilon \mid 0, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\|\epsilon - \mu_\epsilon\|^2}{2\sigma^2}\right)$$

Varianza



$\epsilon = \mu_\epsilon = 0$

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

precisión

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\beta \|\epsilon - \mu_\epsilon\|^2\right)$$

$$\int_{-\infty}^{\infty} p(\epsilon)\,\partial\epsilon = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\|\epsilon - \mu_\epsilon\|^2}{2\sigma^2}\right) \partial\epsilon = 1 .$$

$$p(\epsilon) > 0$$

MODELO DE APRENDIZAJE

$$\epsilon = y - f(x,w) \sim \mathcal{N}(\epsilon \mid 0, \sigma^2) = \mathcal{N}(y - f(x,w) \mid \underset{\mu_\epsilon}{0}, \sigma^2)$$

ALEATORIO

DATOS (ALEATORIOS)

DETERMINÍSTICO

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\|y - f(x,w)\|^2}{2\sigma^2}\right)$$

?

$$p(y \mid f(x,w), \sigma^2) = \mathcal{N}(y \mid f(x,w), \sigma^2)$$

verosimilitud

$\{x_i, y_i\}_{i=1}^{N}$. independientes e idénticamente distribuidos $\rightarrow$ i.i.d.

DATASET

$x_i \perp x_{i-1} \perp x_{i-2} \qquad y_i \sim p(y_i) = p(y_{i+1}) =$

Ej: $\hat{y} = f(x, w) = xw$
$= \phi(x) w$

$$p(y \mid f(x,w), \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(- \frac{\| y - \phi(x) w \|_2^2}{2\sigma^2} \right)$$

$\hat{y}$

Estimación frecuentista - puntual

Estimación Bayesiana

$\mathbb{y} = [y_1 \; y_2 \cdots y_n]^T \in \mathbb{R}^N$; $\mathbb{X} = [x_1 \; x_2 \cdots x_N]^T \in \mathbb{R}^{N \times P}$

i.i.d $\quad p(\mathbb{y} \mid \mathbb{X}, \mathbb{w}, \sigma^2) = p(y_1, y_2, \cdots, y_N \mid \mathbb{X}, \mathbb{w}, \sigma^2)$

$p(a, b) = p(a \mid b) p(b) = p(a) p(b)$

a ind $b \quad p(a \mid b) = p(a)$

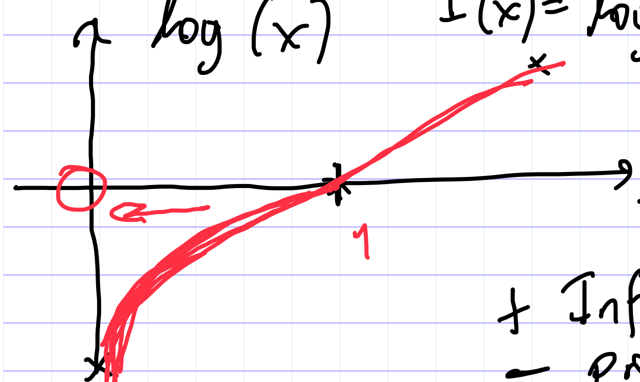
$p(y_1 \mid \mathbb{X}, \mathbb{w}, \sigma^2)\, p(y_2 \mid \mathbb{X}, \mathbb{w}, \sigma^2) \cdots p(y_N \mid \mathbb{X}, \mathbb{w}, \sigma^2) =$

$$\prod_{i=1}^{N} p(y_i \mid \phi(x_i) w, \sigma^2) = \prod_{i=1}^{N} \mathcal{N}(y_i \mid \phi(x_i) w, \sigma^2)$$

$\hat{y}_i$.

$\log(x)$ $\quad I(x) = \log_{2} \left(\frac{1}{p(x)} \right)$

10
n



Hartley → Medida de

Información.

bit.

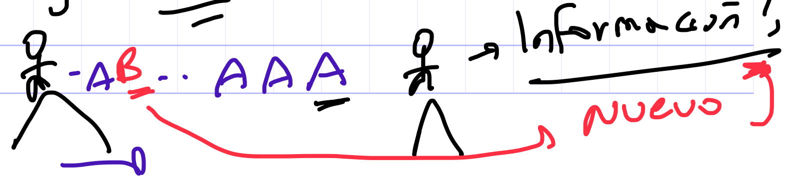
Incertidumbre.

+ Información
- probabilidad

$p(x = A) = 0.8$

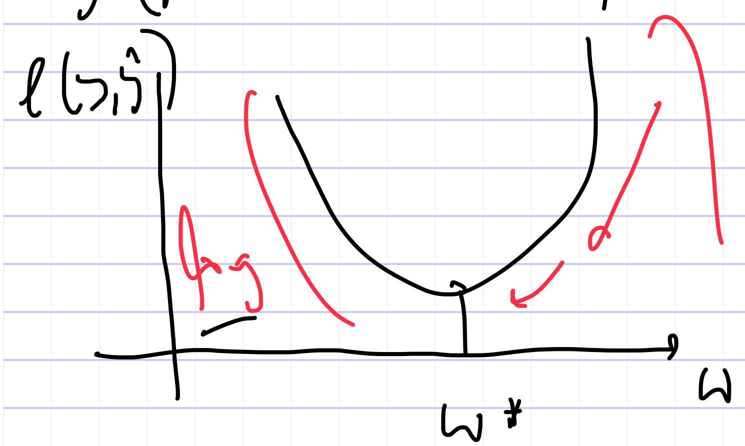
$p(x = B) = 0.1$

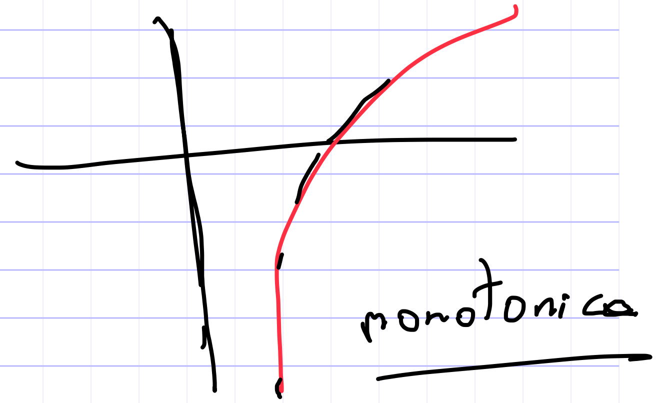
Llego $\underline{A}$



$$\log\left(\frac{1}{p(x)}\right) = -\log(p(x)) = I(x)$$

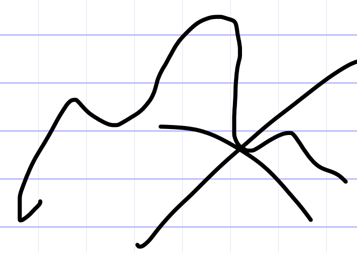
$\log(p(y \mid \phi w, \sigma^2))$ → Información
Aplico no linealidad.





monotonica

$\cancel{\log(\exp(}$



log-likelihood
log-verosimilitud.

$$\log(p(y \mid \phi(x) w, \sigma^2)) = \log\left(\prod_{i=1}^{N} p(y_i \mid \phi(x_i) w, \sigma^2)\right)$$

$$\log(a \cdot b) = \log(a) + \log(b)$$

$$= \sum_{i=1}^{N} \log(p(y_i \mid \phi(x_i) w, \sigma^2)) = \sum_{i=1}^{N} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-|y_i - \phi(x_i) w|^2}{2\sigma^2}\right)\right)$$

$$= \underbrace{\sum_{i=1}^{N} \log\left(\frac{1}{(2\pi\sigma^2)^{1/2}}\right)} + \underbrace{\sum_{i=1}^{N} \cancel{\log}\left(\cancel{\exp}\left(\frac{-|y_i - \phi(x_i) w|^2}{2\sigma^2}\right)\right)}$$

$$\sum_{i=1}^{N} \log\left(\frac{1}{(2\pi\sigma^2)^{1/2}}\right) = \log\left[\prod_{i=1}^{N} \frac{1}{(2\pi\sigma^2)^{1/2}}\right]$$

$$\prod_{i=1}^{N} a = a^N \longrightarrow \log\left(\left(\frac{1}{(2\pi\sigma^2)^{1/2}}\right)^N\right)$$

hiperparametro

$$\log\left(\frac{1}{(2\pi\sigma^2)^{N/2}}\right) = \cancel{\log(1)}^{0} - \log\left((2\pi\sigma^2)^{N/2}\right)$$

β^{-1}

$$-\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2)$$

$\sigma^2 = \beta^{-1}$

$$\sum_{i=1}^{N}\left(-\frac{|y_i - \phi(x_i)w|^2}{2\sigma^2}\right) = -\frac{1}{2\sigma^2}\sum_{i=1}^{N}|y_i - \phi(x_i)w|^2$$

y $\qquad \hat{y} = \phi w$

$$\|\underline{\underline{y}}\|_2^2 = y_1^2 + y_2^2 + \cdots + y_N$$

$$= -\frac{1}{2\sigma^2}\|y - \phi(x)w\|_2^2$$

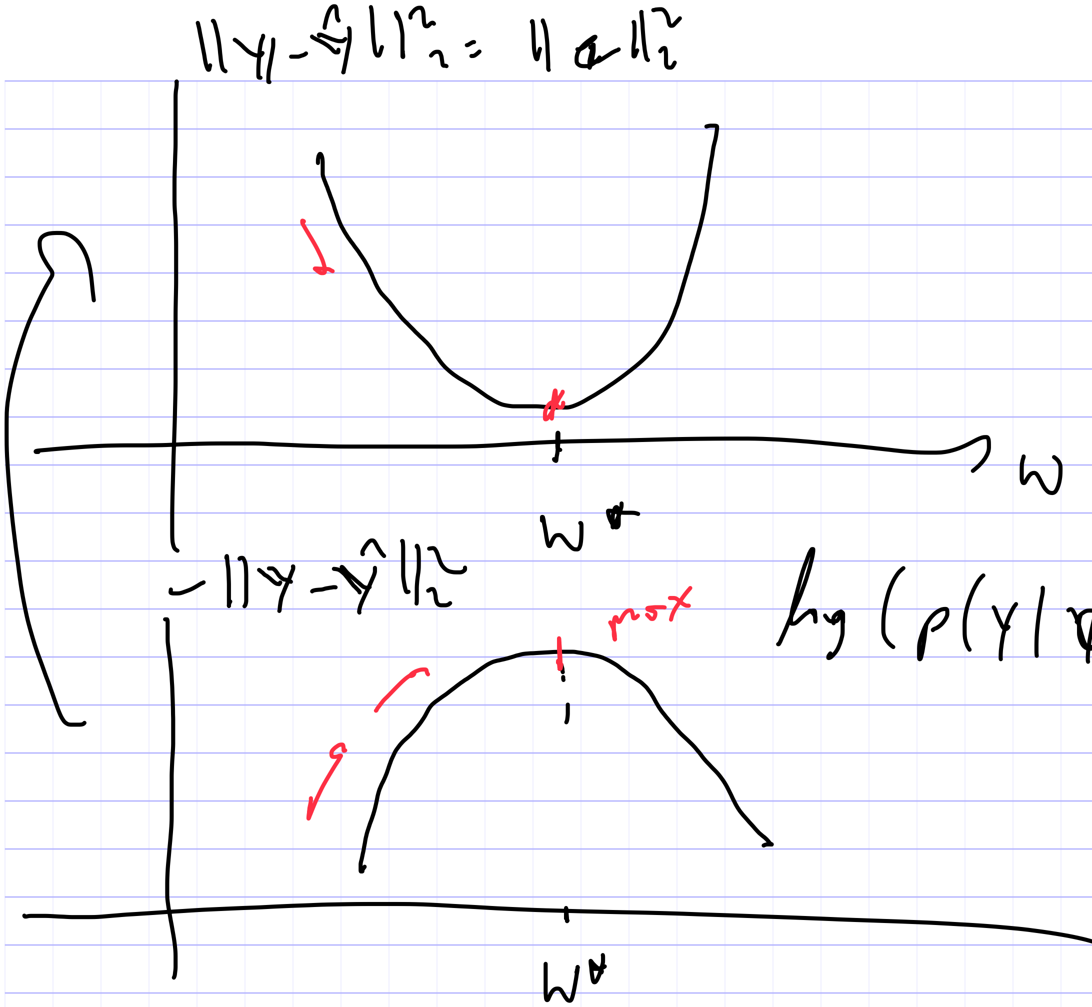
$\rightarrow$ cte $\rightarrow$ cte

$$\log(p(y \mid \phi(x)w, \sigma^2)) = -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2)$$

$$-\frac{1}{2\sigma^2}\|y - \phi(x)w\|_2^2$$

Fijar $\rightarrow$ estimar ?

$$w^* = \arg\max_{w}$$



$$\log\left(p(y \mid \phi(x)w, \sigma)\right)$$

$$\downarrow$$

$$\epsilon$$

$$y = f(x,w) + \epsilon \rightarrow \epsilon = y - f(x,w) \sim N(\epsilon \mid 0, \sigma^2)$$

$$p(\epsilon)$$

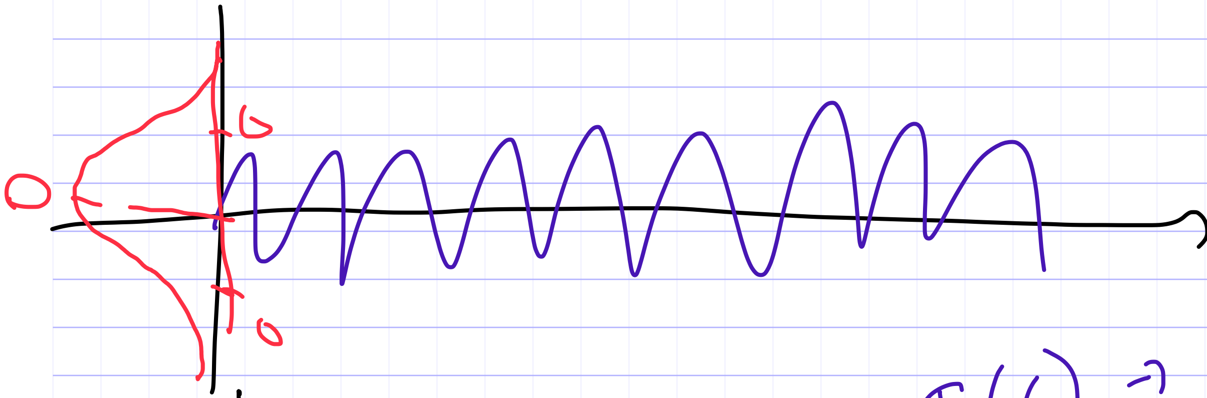
$$\log\left(p(\epsilon)\right)$$

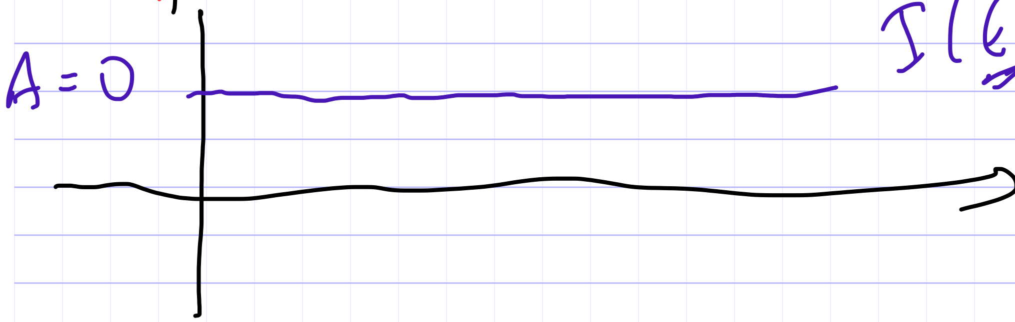
$$I(\epsilon) = \log\left(\frac{1}{p(\epsilon)}\right) = -\log\left(p(\epsilon)\right)$$

$$-I(\epsilon) = +\log\left(p(\epsilon)\right) = \log\left(p(y \mid \phi(x)w, \sigma^2)\right)$$

$$w^* = \arg\max \log\left(p(y \mid \phi(x), w, \sigma^2)\right) \underset{(-1)}{=} +\log\left(p(\epsilon)\right)$$

$$w^* = \arg\min_{w} -\log\left(p(\epsilon)\right) = I(\epsilon)$$





$I(\epsilon) \to 0$

NO HAY INFORMACIÓN

$\|y - \hat{y}\| = \|\epsilon\|$

$p(y \mid \phi(X) w, \sigma^2)$

---

1. Estimación frecuentista → máx verosimilitud.
   (incertidumbre está en $y = f(x,w) + \epsilon$)
   PUNTUAL $w = ?$

2. Estimación Bayesiana.

$y = f(x,w) + \epsilon$; $\epsilon \sim p(\epsilon)$

$w \sim p(w)$

Aleatorio

Conocimiento del proceso.

PRIOR

$$p(y \mid f(x,w)) \to p(y \mid w)$$

$$p(y,w) = p(w,y)$$

$$p(y \mid w)\,p(w) = p(w \mid y)\,p(y)$$

Si y ind w = ? NO HAGO NADA.

$$p(w \mid y) = \frac{p(y \mid w)\,p(w)}{p(y)}$$

PRIOR

POSTERIOR

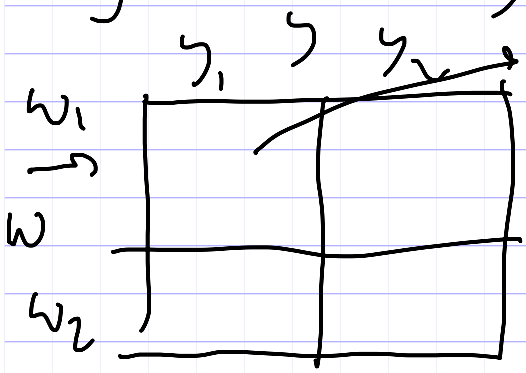
VEROSIMILITUD

EVIDENCIA

NORMALIZAR

$$\int p(y)\,dy = 1$$

$$\int p(y,w)\,dw = \int p(y \mid w)\,p(w)\,dw = p(y)$$



$$p(w,y) = p(y,w)$$

$$p(w_1,y_1) + p(w_1,y_2) = p(w_1)$$

$$\sum_j p(w,y_j) = p(w)$$

$$\int p(w,y)\,dy = p(y)$$

Inferencia $\longrightarrow$ Optimizar.

$$p(w|y) = \frac{p(y|w)\,p(w)}{p(y)}$$

Suposiciones en Bayes.

1. Max A-posteriori $\quad p(w|y) \propto p(y|w)\,p(w)$

2. Prior Conjugador $\quad p(y|w)\,p(w)$

↳ combinación de pdf con solución analítica.

$$p(y) = \underbrace{\int p(y|w)\,p(w)\,dw}$$

3. Aproximación o muestreo $\quad$ iterativo

4. Variacional → cota del posterior

↳ gradiente estocástico.

$y = f(x,w) + \epsilon$, $\quad \epsilon \sim p(\epsilon) = N(\epsilon|0,\sigma_\epsilon^2)$

$w \sim p(w) = N(w|0,\sigma_w^2)$

MAP → BAYES MAXIMUM A-POSTERIORI

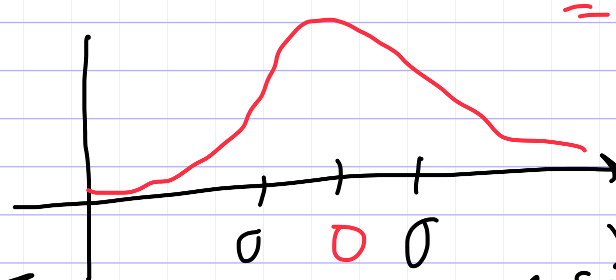
$$p(w|y) \propto p(y|w)\,p(w)$$

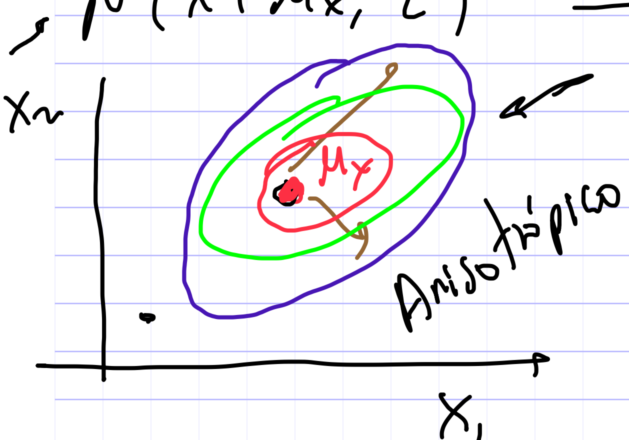
$$\epsilon = y - f(x,w) \sim N(y|f(x,w),\sigma_\epsilon^2)$$

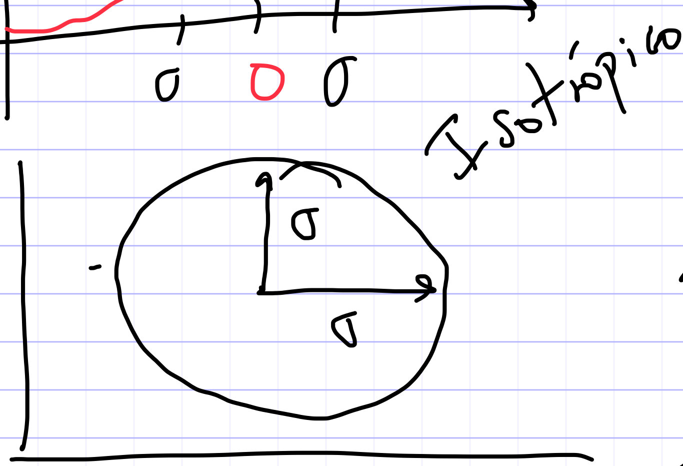
$$\hat{p}(w|y) = N(y|f(x,w),\sigma_\epsilon^2)\,N(w|0,\sigma_w^2)$$

$N(x|0,\sigma^2)$

$N(x|\mu_x,\Sigma)$







$x = [x_1, x_2]$

$$\Sigma = \sigma^2 I$$

$$N(x|\mu_x,\Sigma) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left(-\frac{\overbrace{(x-\mu_x)\Sigma^{-1}(x-\mu_x)^T}^{\text{Distancia Mahalanobis}}}{2}\right)$$

$$\{x_i, y_i\}_{i=1}^{N} \; ; \quad \text{i.i.d.}, \qquad w \sim \mathcal{N}(w \mid 0, \sigma_w^2)$$

$$\Sigma_w = \sigma_w^2 I$$

$$p(w \mid y, f(x, w)) \propto \underbrace{p(y \mid w)}\,\underbrace{p(w)}$$

$$\log\left(\hat{p}(w \mid y)\right) = \log\left(\prod_{i=1}^{N} p(y_i \mid w) \prod_{j=1}^{Q} p(w_j)\right) \; ; \quad X \in \mathbb{R}^{N \times P}$$

$$\phi(X) \in \mathbb{R}^{N \times Q}$$

$$\underset{1\times 1}{\hat{y}} = \underset{1 \times Q}{\phi(x)}\, \underset{Q \times 1}{w} \qquad w \in \mathbb{R}^{Q}$$

$$\log\left(\hat{p}(w \mid y)\right) = \log\left(\prod_{i=1}^{N} p(y_i \mid w)\right) + \log\left(\prod_{j=1}^{Q} p(w_j)\right).$$

$$= \sum_{i=1}^{N} \log\left(p(y_i \mid w)\right) + \sum_{j=1}^{Q} \log p(w_j) \qquad \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left(-\frac{\|w_j\|_2^2}{2\sigma_w^2}\right)$$

$$= -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2}\|y - \phi(X)w\|_2^2$$

$$- \frac{Q}{2}\log(2\pi) - \frac{Q}{2}\log(\sigma_w^2) - \frac{1}{2\sigma_w^2}\|w\|_2^2 .$$

$$- \frac{1}{2\sigma_\epsilon^2}\sigma_\epsilon^2 \|y - \phi(X)w\|_2^2 - \frac{1}{2\sigma_w^2}\sigma_\epsilon^2\|w\|_2^2$$

$$\underbrace{\|y - \hat{y}\|_2^2 + \lambda\|w\|_2^2}_{\text{regularized}}$$

$$- \frac{1}{2}\|y - \phi(X)w\|_2^2 - \frac{1}{2}\frac{\sigma_\epsilon^2}{\sigma_w^2}\|w\|_2^2 . \qquad \lambda = \frac{\sigma_\epsilon^2}{\sigma_w^2},$$