1

Linear Regression Mean Square Error is :

$$J(\theta) = \frac{1}{4}\sum_{x\in X}\left(f^*(x) - f(x;\theta)\right)^2$$

for Linear Regression , $f(x;\theta) = x^T w + b$

$$\therefore J(\theta) = \frac{1}{4}\sum_{x\in X}\left(f^*(w) - (x^T w + b)\right)^2 \quad \text{—— (I)}$$

Given $w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$ and $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$$\therefore x^T w = x_1 w_1 + x_2 w_2$$

$$\therefore J(\theta) = \frac{1}{4}\sum_{x\in X}\left(f^*(x) - (x_1 w_1 + x_2 w_2 + b)\right)^2$$

Taking gradient of $J(\theta)$ wrt w_1 :

$$\nabla w_1 (J(\theta)) = \nabla w_1\left(\frac{1}{4}\sum_{x\in X}\left(f^*(x) - (x_1 w_1 + x_2 w_2 + b)\right)^2\right)$$

$$\nabla w_1 (J(\theta)) = \frac{-1}{2}\sum_{x\in X}(x_1)\left(f^*(x) - x_1 w_1 - x_2 w_2 - b\right) \quad \text{—— (II)}$$

similarly taking gradient of $J(\theta)$ wrt w_2 :

$$\nabla w_2 (J(\theta)) = -\frac{1}{2}\sum_{x\in X}(x_2)\left(f^*(x) - x_1 w_1 - x_2 w_2 - b\right) \quad \text{—— (III)}$$

~~Hqfocling~~ Taking gradient of $J(\theta)$ wrt b :

$$\nabla b (J(\theta)) = -\frac{1}{2}\sum_{x\in X}\left(f^*(x) - x_1 w_1 - x_2 w_2 - b\right) \quad \text{—— (IV)}$$

Equating equations (II), (III) & (IV) to 0, yields:

$$0 = \frac{-1}{2}\sum_{x \in X} (x_1)\left(f^*(x) - x_1 w_1 - x_2 w_2 - b\right) \quad \text{--- (V)}$$

$$0 = \frac{-1}{2}\sum_{x \in X} (x_2)\left(f^*(x) - x_1 w_1 - x_2 w_2 - b\right) \quad \text{--- (VI)}$$

$$0 = \frac{-1}{2}\sum_{x \in X} \left(f^*(x) - x_1 w_1 - x_2 w_2 - b\right) \quad \text{--- (VII)}$$

for X-OR,

x_1	x_2	$f^*(x)$
0	0	0
0	1	1
1	0	1
1	1	0

Putting values of x_1 & x_2 in eq^n (V)

$$0 = \frac{-1}{2}\Big[(0)(0-0\times w_1 - 0\times w_2 - b) + (0)(1 - 0\times w_1 - 1\times w_2 - b) + (1)(1 - 1\times w_1 - 0\times w_2 - b) + (1)(0 - 1\times w_1 - 1\times w_2 - b)\Big]$$

$$0 = 0 + 0 + (1 - w_1 - b - w_1 - w_2 - b) = 1 - 2w_1 - w_2 - 2b$$

$$1 = 2w_1 + w_2 + 2b \quad \text{--- (VIII)}$$

Putting values of x_1 & x_2 in eq^n (VI)

$$0 = -\frac{1}{2}\Big[(0)(0 - 0\times w_1 - 0 w_2 - b) + (1)(1 - 0\times w_1 - 1\times w_2 - b) + (0)(1 - 1\times w_1 - 0\times w_2 - b) + (1)(0 - 1\times w_1 - 1\times w_2 - b)\Big]$$

$$0 = [0 + (1 - w_2 - b) + 0 + (-w_1 - w_2 - b)]$$

$$0 = 1 - w_1 - 2w_2 - 2b$$

$$1 = w_1 + 2w_2 + 2b \quad \text{—} \quad (IX)$$

Putting values of x_1 & x_2 in eq^n (VII)

$$0 = -\frac{1}{2}\Big([0 - 0\times w_1 - 0\times w_2 - b] + [1 - 0\times w_1 - 1\times w_2 - b] + [1 - 1\times w_1 - 0\times w_2 - b] + [0 - 1\times w_1 - 1\times w_2 - b]\Big)$$

$$0 = [2 - 2w_1 - 2w_2 - 4b]$$

$$1 = w_1 + w_2 + 2b \quad \text{—} \quad (X)$$

Solving eq^n (VIII), (IX), (X) yields,

$$\boxed{w_1 = 0, \quad w_2 = 0, \quad b = 0.5}$$

Hence Proved.

(3) Considering regularization term as

$$\frac{\alpha}{2} w^T S w$$

where S is a symmetric matrix $\begin{bmatrix} S_1 & S_2 \\ S_3 & S_4 \end{bmatrix}$

$\therefore$ Regularization $\Rightarrow \frac{\alpha}{2} w^T \begin{bmatrix} S_1 & S_2 \\ S_3 & S_4 \end{bmatrix} w$

for a small 1×2 image, weight matrix is

$$w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$= \frac{\alpha}{2} \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} S_1 & S_2 \\ S_3 & S_4 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$= \frac{\alpha}{2} \begin{bmatrix} w_1 S_1 + w_2 S_3 & w_1 S_2 + w_2 S_4 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$= \frac{\alpha}{2} \left(w_1^2 S_1 + w_1 w_2 S_3 + w_1 w_2 S_2 + w_2^2 S_4 \right)$$

for S to be symmetric, $S_2 = S_3$

$\therefore$ Regularization $= \frac{\alpha}{2} \left(w_1^2 S_1 + 2 S_2 w_1 w_2 + w_2^2 S_4 \right)$

Considering $S_1 = S_4 = 1$

and $S_2 = S_3 = -1$

i.e. $S = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$

gives us,

$$\frac{\alpha}{2}\left(\omega_1^2 - 2\omega_1\omega_2 + \omega_2^2\right)$$

$$= \frac{\alpha}{2}\left(\omega_1 - \omega_2\right)^2$$

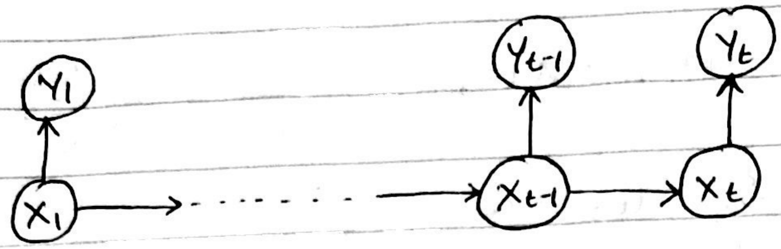
Minimizing Regularization term to 0, yields

$$0 = \frac{\alpha}{2}\left(\omega_1 - \omega_2\right)^2$$

$$\therefore \omega_1 = \omega_2$$

Thus, it is clear from above weight symmetry that with weights being equal to each other preserves symmetry in regularization term. It prevents the image from not being too asymmetric.

4



$$P(x_t \mid y_1 \dots y_t) = \frac{P(x_t, y_1, y_2 \dots y_t)}{P(y_1, y_2, \dots y_t)}$$ $\Rightarrow$ According to Conditional prob.

$$\therefore P(x_t \mid y_1 \dots y_t) \propto P(x_t, y_1, \dots y_t) \quad \text{——— (1)}$$

As per the Hidden Markov Model process diagram, we can expand the probability distribution using Marginal Probability as:

$$P(x_t, y_1 \dots y_t) = \sum_{x_{t-1}} P(x_t, x_{t-1}, y_1 \dots y_t)$$

Now factorizing this term gives us:

$$P(x_t, y_1 \dots y_t) = \sum_{x_{t-1}} P(y_t \mid x_t, x_{t-1}, y_1 \dots y_{t-1}) * P(x_t \mid x_{t-1}, y_1 \dots y_{t-1}) * P(x_{t-1}, y_1 \dots y_{t-1})$$

As per property of Markov Model-Hidden :

$$P(x_t \mid x_1 \dots x_{t-1}) = P(x_t \mid x_{t-1})$$

and $$P(y_t \mid x_t, y_1 \dots y_{t-1}) = P(y_t \mid x_t)$$

Thus,

$$P(x_t, y_1 \ldots y_t) = \sum_{x_{t-1}} P(y_t | x_t) * P(x_t | x_{t-1}) * P(x_{t-1}, y_1 \ldots y_{t-1})$$

$$= P(y_t | x_t) \sum_{x_{t-1}} P(x_t | x_{t-1}) \; P(x_{t-1}, y_1 \ldots y_{t-1})$$

As per equation ① of proportionality and the property of Recursion:

$$P(x_t | y_1 \ldots y_t) \propto P(y_t | x_t) \sum_{x_{t-1}} P(x_t | x_{t-1}) \, P(x_{t-1} | y_1 \ldots y_{t-1})$$

Hence proved.

5

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

$$p(D|w, \sigma^2) = \prod_{i=1}^{n} p(y^{(i)} | x^{(i)}, w)$$

taking log.

$$\log p(D|w, \sigma^2) = \log \prod_{i=1}^{n} p(y^{(i)} | x^{(i)}, w)$$

$$= \sum_{i=1}^{n} \log p(y^{(i)} / x^{(i)}, w)$$

$$= \sum_{i=1}^{n} \log \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y_i - x_i^T w)^2}{2\sigma^2} \right) \right]$$

$$\log p(D|w, \sigma^2) = \sum_{i=1}^{n} \left[\log\left(\frac{1}{\sqrt{2\pi}}\right) + \log\left(\frac{1}{\sigma}\right) + \left(\frac{-(y_i - x_i^T w)^2}{2\sigma^2} \right) \right] \text{— (1)}$$

taking gradient wrt σ^2

$$\nabla_{\sigma^2} \log p(D|w, \sigma^2) = \sum_{i=1}^{n} \left[0 + \left(\frac{1}{\frac{1}{\sigma}} (-1) \frac{1}{\sigma^2} \right) + \left(\frac{-1}{2} (y_i - x_i^T w)^2 \frac{(-2)}{\sigma^3} \right) \right]$$

Scanned with CamScanner

$$= \sum_{i=1}^{n} \left[-\frac{1}{\sigma} + \frac{(y_i - x_i^T w)^2}{\sigma^3} \right]$$

setting gradient to 0.

$$0 = \sum_{i=1}^{n} \left[\frac{-1}{\sigma} + \frac{(y_i - x_i^T w)^2}{\sigma^3} \right]$$

$$0 = -\frac{n}{\sigma} \sum_{i=1}^{n} \left(\frac{(y_i - x_i^T w)^2}{\sigma^3} \right)$$

$$\frac{n}{\sigma} = \frac{1}{\sigma^3} \sum_{i=1}^{n} (y_i - x_i^T w)^2$$

$$\boxed{\begin{aligned} \sigma^2 &= \frac{1}{n} \sum_{i=1}^{n} (y_i - x_i^T w)^2 \\ \sigma^2 &= \frac{1}{n} \sum_{i=1}^{n} (x_i^T w - y_i)^2 \end{aligned}}$$

Scanned with CamScanner

Rewritting eq^n ① as follows:

$$\log P(D|w,\sigma^2) = n \log\left(\frac{1}{\sqrt{2\pi}}\right) + n \log\left(\frac{1}{\sigma}\right) + \sum_{i=1}^{n}\left(-\frac{(y_i - x_i^T w)^2}{2\sigma^2}\right)$$

Taking gradient wrt w

$$\nabla_w \left(\log P(D|w,\sigma^2)\right) = 0 + 0 + \sum_{i=1}^{n} \frac{-1}{2\sigma^2} \times 2 \left(y_i - x_i^T w\right) \cdot x_i$$

Setting gradient to zero.

$$0 = -\sum_{i=1}^{n} \frac{1}{\sigma^2} \left(y_i - x_i^T w\right) \cdot x_i$$

$$0 = \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \cdot x_i^T w$$

$$\sum_{i=1}^{n} x_i x_i^T w = \sum_{i=1}^{n} x_i y_i$$

$$w = \sum_{i=1}^{n} \frac{x_i y_i}{\sum_{i=1}^{n} x_i x_i^T}$$

$$\boxed{w = \left(\sum_{i=1}^{n} x_i x_i^T\right)^{-1} \left(\sum_{i=1}^{n} x_i y_i\right)}$$