1. MSE loss function is

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

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

Now for XOR, we have,

$X =$ x_1	x_2	Y
0	0	0
1	0	1
0	1	1
1	1	0

Taking gradient of eq^n wrt w, we get

$$\frac{\partial J}{\partial w_1} = \frac{1}{2} \sum_{x \in X} (-x_1)(y - x_1 w_1 - x_2 w_2 - b) = 0 \quad \cdots (1)$$

$$\frac{\partial J}{\partial w_2} = \frac{1}{2} \sum_{x \in X} (-x_2)(y - x_1 w_1 - x_2 w_2 - b) = 0 \quad \cdots (2)$$

substituting values in 1^{st} eq^n,

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

$$\therefore \frac{1}{2}(-1+w_1+b+w_1+w_2+b)=0$$

$$\therefore -1+2w_1+w_2+2b=0 \quad \ldots③$$

substituting values in 2^{nd} eq^n we get,

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

$$\therefore -1+w_2+b+w_1+w_2+b=0$$
$$\therefore -1+w_1+2w_2+2b=0 \quad \ldots④$$

equating eq^n ③ & ④ we get

$$-1+2w_1+w_2+2b=-1+w_1+2w_2+2b$$
$$\boxed{w_1=w_2}$$

Now, we take gradient of eq^n wrt b,

$$\frac{\partial J}{\partial b} = -\frac{1}{2} \sum_{x \in X} (Y - x_1 w_1 - x_2 w_2 - b)$$

Substituting values of x_1, x_2 we get,

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

$$= \frac{-1}{2} \left[(-b + 1 - w_1 - b + 1 - w_2 - b - w_1 - w_2 - b) \right]$$

$$= \frac{-1}{2} \left[2 - 2w_1 - 2w_2 - 4b \right]$$

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

equating this eq^n to zero we get,

$$b = \frac{1}{2} - \frac{w_1}{2} - \frac{w_2}{2}$$

but $w_1 = w_2$

$$\therefore \boxed{b = \frac{1}{2} - w_1}$$

Substituting value of b in equation 3

$$-1 + 2w_1 + w_2 + 2\left(\frac{1}{2} - w_1\right) = 0$$

$$\therefore 0 + w_2 = 0 \quad \therefore \boxed{w_2 = 0}$$

as $w_1 = w_2 \quad \therefore \boxed{w_1 = 0}$

$b = -w_1 + 1/2 \quad \therefore \boxed{b = 1/2}$

Q3 Ans

consider the below given equation

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

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

Here we want to find S such that $w_1 = w_2$ to make sure that the image does not become too asymmetric after applying weights

$$\therefore \frac{\alpha}{2} w^T S w = \frac{\alpha}{2} [w_1 \ w_2] \begin{bmatrix} s_1 & s_2 \\ s_3 & s_4 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

Assuming $S = \begin{bmatrix} s_1 & s_2 \\ s_3 & s_4 \end{bmatrix}$

$$= \frac{\alpha}{2} [w_1 \ w_2] \begin{bmatrix} w_1 s_1 + w_2 s_2 \\ w_1 s_3 + w_2 s_4 \end{bmatrix}$$

$$= \frac{\alpha}{2} \begin{pmatrix} w_1^2 s_1 + w_1 w_2 s_2 + \\ w_1 w_2 s_3 + w_2^2 s_4 \end{pmatrix}$$

$$= \frac{\alpha}{2} (w_1^2 s_1 + w_1 w_2 (s_2 + s_3) + w_2^2 s_4)$$

Choosing the values of s_1, s_2, s_3 and s_4 such that we get, $(w_1 - w_2)^2$

Let $s_1 = 1 \quad s_4 = 1 \quad s_2 = s_3 = -1$

$\therefore$ we get

$$\frac{\alpha}{2} w^T S w = \frac{\alpha}{2} [w_1^2 - 2 w_1 w_2 + w_2^2]$$

$$= \frac{\alpha}{2} [w_1 - w_2]^2$$

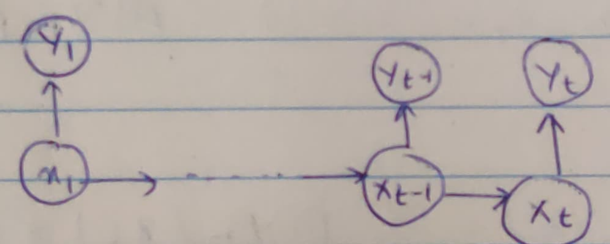
we get the regularisation term as $(w_1 - w_2)^2$
So when we set it to zero

$$(w_1 - w_2)^2 = 0$$

$$\boxed{w_1 = w_2}$$

This enables the weights to be equal which makes sure that the regularisation induces symmetry in the image due to equal weights. and makes sure that the image does not become too asymmetric.

Q4 Ans



According to conditional probability.

$$P(x_t / y_1 \cdots y_t) = \frac{P(x_t, y_1, y_2 \cdots y_t)}{P(y_1, y_2 \cdots y_t)} \cdots$$

$$\therefore P(x_t / y_1 \cdots y_t) \propto P(x_t, y_1 \cdots y_t) - (1)$$

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

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

Now factorising the given term gives us,

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

As property of Markov Model Hidden, using marginal probability,

$$p(x_t / x_1 \ldots x_{t-1}) = p(x_t / x_{t-1})$$
$$p(y_t / x_t, y_1 \ldots y_{t-1}) = p(y_t / x_t)$$

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 Recurssion

$$p(x_t / y_1 \ldots y_t) \propto p(y_t / x_t) * \sum_{x_{t-1}} p(x_t / x_{t-1}) * p(x_{t-1} / y_1 \ldots y_{t-1})$$

Q5 Ans

5

Given that:

$$P(Y|x) = \mathcal{N}(\mu = x^T w, \sigma^2)$$

$$P(Y^{(i)}/x^{(i)}, w) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(Y - x^T w)^2}{2\sigma^2}\right) \quad -(1)$$

$$P(D/w, \sigma^2) = \prod_{i=1}^{n} P(Y^{(i)}/x^{(i)}, w)$$

$$= \log \prod_{i=1}^{n} P(Y^{(i)}/x^{(i)}, w)$$

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

Substituting from above eq^n (1)

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

$$= \sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi}}\right) + \sum_{i=1}^{n} \log \frac{1}{\sqrt{\sigma^2}}$$

$$+ \sum_{i=1}^{n} \log \left[\exp\left(-\frac{(Y - x^T w)^2}{2\sigma^2}\right)\right]$$

$$= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2)$$

$$+ \sum_{i=1}^{n} \cancel{\log} - \frac{(Y - x^T w)^2}{2\sigma^2} \quad -(2)$$

Let $\sigma^2 = a$ in eq^n (2).

$$P(D|w,\sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln a$$

$$- \sum_{i=1}^{n} \frac{(Y - x^T w)^2}{2a} \quad -$$

Taking gradient w.r.t a

$$\nabla a\,[P(D|w,\sigma^2)] = \nabla a\left[-\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln a\right.$$

$$\left.- \sum_{i=1}^{n} \frac{(Y - x^T w)^2}{2a}\right]$$

$$= \left[0 - \frac{n}{2a} - \sum_{i=1}^{n} \frac{(Y - X^T w)^2(-1)}{2a^2}\right]$$

$$= \left[-\frac{n}{2a} + \sum_{i=1}^{n} \frac{(Y - x^T w)^2}{2a^2}\right]$$

$$\therefore \nabla a\,[P(D|w,\sigma^2)] = 0$$

$$\therefore -\frac{n}{2a} + \sum_{i=1}^{n} \frac{(Y - x^T w)^2}{2a^2} = 0$$

$$\sum_{i=1}^{n} \frac{(Y - x^T w)^2}{\cancel{2}a^{\cancel{2}}} = \frac{n}{\cancel{2}\cancel{a}}$$

$$\sum_{i=1}^{n} (Y - x^T w)^2 = na$$

$$\therefore a = \frac{1}{n}\sum_{i=1}^{n} (Y - x^T w)^2$$

Resubstitute $a = \sigma^2$

$$\therefore \boxed{\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} (Y - x^T w)^2}$$

From equation ②, we have

$$P(D|w, \sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) + \sum_{i=1}^{n} -\frac{(Y - x^T w)^2}{2\sigma^2}$$

Taking gradient with respect to w,

$$\nabla_w [P(D|w, \sigma^2)] = \nabla_w \left[-n/2 \ln(2\pi) - n/2 \ln(\sigma^2) + \sum_{i=1}^{n} -\frac{(Y - x^T w)^2}{2\sigma^2} \right]$$

$$= 0 + 0 \cancel{\sum_{i} \frac{x (-2\sigma)}{(-2\sigma^2)} (Y - x^T w)^2}$$

$$= 0 + 0 + \sum_{i=1}^{n} \frac{(-2)}{(-2\sigma^2)} (Y - x^T w)^{\cancel{2}} x$$

Let $\nabla_w (P(D|w, \sigma^2)) = 0$

$$\therefore 0 = \sum_{i=1}^{n} (Y - x^T w) x$$

$$0 = \sum_{i=1}^{n} x^{(i)} y^{(i)} - \sum_{i=1}^{n} x^{(i)} x^{(i)T} w$$

$$\sum_{i=1}^{n} x^{(i)} x^{(i)T} w = \sum_{i=1}^{n} x^{(i)} y^{(i)}$$

$$\therefore w = \frac{\sum_{i=1}^{n} x^{(i)} y^{(i)}}{\sum_{i=1}^{n} x^{(i)} x^{(i)T}}$$

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