Amir Hossein Eyvazkhani
1747696
Ex6

Task I)

1. $P(h=1|v) = \dfrac{\alpha_h + N_h}{\alpha_h + \alpha_t + N_h + N_t} = \dfrac{1+4}{1+1+4+3} = \dfrac{5}{9}$

2. $P(b=1|v) = \dfrac{\alpha_h + N_h}{\alpha_h + \alpha_t + N_h + N_t} = \dfrac{1+4}{1+1+4+3} = \dfrac{5}{9}$

3. $P(c=1|f=1, b=1, v) = \dfrac{\alpha_h + N_h}{\alpha_h + \alpha_T + N_h + N_t} = \dfrac{1+2}{1+1+2+0} = \dfrac{3}{4}$

4. $P(c=1|f=1, b=0, v) = \dfrac{\alpha_h + N_h}{\alpha_h + \alpha_T + N_h + N_T} = \dfrac{1+1}{1+1+1+1} = \dfrac{2}{4}$

5. $P(c=1|f=0, b=1, v) = \dfrac{\alpha_h + N_h}{\alpha_h + \alpha_T + N_h + N_T} = \dfrac{1+1}{1+1+1+1} = \dfrac{2}{4}$

6. $P(c=1|f=0, b=0, v) = \dfrac{\alpha_h + N_h}{\alpha_h + \alpha_T + N_h + N_t} = \dfrac{1+0}{1+1+0+1} = \dfrac{1}{3}$

Task II)

a) Loss function for probabilistic regression is the likelihood which we try to maximize as:

$$\theta^* = \arg\max_\theta P(y|X, \theta) = \arg\max_\theta \prod_{n=1}^{N} P(y_n|x_n, \theta)$$

then we apply log which is strictly monotone and multiply by -1 : $\theta^* = \arg\min_\theta -\log P(y|X, \theta)$ (I)

And in our case : $P(y|X, \theta) = N(y|x^T\theta, \sigma^2) = \dfrac{1}{\sqrt{2\pi}\,\sigma^2} e^{\frac{-(y - x^T\theta)^2}{2\sigma^2}}$ (II)

$$\text{I}, \text{II} \Longrightarrow L(\theta) = \sum_{n=1}^{N} \log p(y_n \mid x_n, \theta)$$

$$= \sum_{n=1}^{N} \log \left[\frac{1}{\delta^2 \sqrt{2\pi}} e^{\frac{-(y_n - x_n^T\theta)^2}{2\delta^2}} \right]$$

$$= \sum_{n=1}^{N} \log \left(\frac{1}{\delta^2 \sqrt{2\pi}} \right) + \sum_{n=1}^{N} \log \left[e^{\frac{-(y_n - x_n^T\theta)^2}{2\delta^2}} \right]$$

$$= \underbrace{-N \log (\delta^2 \sqrt{2\pi})}_{\text{constant}} - \underbrace{\underbrace{\frac{1}{2\delta^2}}_{\text{Constant}} \sum_{i=1}^{N} (y_i - x_n^T\theta)^2}_{\text{MSE}}$$

Hence it is the same as minimizing the MSE.

ⓑ $p(\theta) = N(\theta \mid 0, \delta^2 I)$, $p(y \mid X, \theta) = N(y \mid x^T\theta, \sigma^2)$

$$p(\theta \mid y, X) = \frac{1}{Z} p(y \mid X, \theta)\, p(\theta)$$

by bayes rule we have

$$\Rightarrow \theta^* = \underset{\theta}{\operatorname{argmax}}\ N(y \mid x^T\theta, \sigma^2)\, N(\theta \mid 0, \delta^2 I)$$

we apply log and minus and the likelihood function will be

$$\Rightarrow L(\theta) = \sum_{n=1}^{N} \log \left[\frac{1}{\delta^2 \sqrt{2\pi}} e^{\frac{-(y_n - x_n^T\theta)^2}{2\delta^2}} \right] + \sum_{n=1}^{D} \log \left[\frac{1}{\delta^2 \sqrt{2\pi}} e^{\frac{-\theta_n^2}{2\delta^2}} \right]$$

$$= -N \log \delta^2 \sqrt{2\pi} - \frac{1}{2\delta^2} \sum_{n=1}^{N} (y_n - x_n^T\theta)^2 - D \log (\delta^2 \sqrt{2\pi}) - \frac{1}{2\delta^2} \sum_{n=1}^{D} \theta_n^2$$

$$= -\underbrace{\sum_{n=1}^{N} (y_i - x_i^T\theta)^2}_{\text{MSE}} - \underbrace{\frac{1}{\delta^2} \|\theta\|_2^2}_{\substack{\text{Regularization} \\ \text{term with} \\ \lambda = \frac{1}{\delta^2}}} - \underbrace{D \log (\delta^2 \sqrt{2\pi}) - N \log \delta^2 \sqrt{2\pi}}_{\text{Constant}}$$

Hence it is equivalent to minimizing with MSE and regularization term of type L2.

CamScanner

c) $P(\theta) = \prod_{m=1}^{M} p(\theta_m) = \prod_{m=1}^{M} \frac{1}{2b} e^{\frac{-|\theta_m|}{b}}$ $(\mu = 0)$

Like the previous Question, we have:

$$p(\theta | x, y) = \frac{1}{Z} p(y | x, \theta) p(\theta) = \frac{1}{Z} N(y | \theta' x, \delta^2) L(\theta | 0, b)$$

$$\Rightarrow L(\theta) = \sum_{i=1}^{N} \log \frac{1}{\delta^2 \sqrt{2\pi}} e^{\frac{-(y_i - x_i\theta)^2}{2\delta^2}} + \sum_{j=1}^{m} \log \frac{1}{2b} e^{\frac{-|\theta_m|}{b}}$$

$$= \underbrace{\frac{-1}{2\delta^2} \sum_{i=1}^{N} (y_i - \theta x_i)^2}_{\text{MSE}} \underbrace{- \frac{1}{b} \sum_{j=1}^{m} \|\theta_j\|}_{\text{regularization}} \underbrace{- N \log \delta^2 \sqrt{\pi} - \log 2b}_{\text{Constant}}$$

Hence it's like minimizing with MSE with L1 regularization

with $\lambda = \frac{2\delta^2}{b}$