Amir Hossein Eyvazkhani
1747696
Exercise 3

Task 1)

a) $f_\theta(x) = \delta(\theta_0 + x_1\theta_1 + x_2\theta_2)$

On the decision boundary lines, $f_\theta(x) = 0.5 \Rightarrow \theta_1 x_1 + \theta_2 x_2 + \theta_0 = 0$

$$\Rightarrow \begin{cases} \overset{(x_1, x_2)}{(0,1)} : \theta_2 + \theta_0 = 0 \\ (4,5) : 4\theta_1 + 5\theta_2 + \theta_0 = 0 \\ (1,2) : \theta_1 + 2\theta_2 + \theta_0 = 0 \end{cases} \Rightarrow \theta^* = (\theta_0^*, \theta_1^*, \theta_2^*) = (-1, -1, 1)$$

checking if it correctly classifies : $f_{\theta^*} = \delta(-x_1 + x_2 - 1)$

x_1	x_2	f_{θ^*}	y	correct?
1	3	$\delta(1) = 0.73$	1	✓
2	4	$\delta(1) = 0.73$	1	✓
2	2	$\delta(-1) = 0.26$	0	✓
3	3	$\delta(-1) = 0.26$	0	✓

$$\log P(Y|X\theta^*) = \sum_{n=1}^{4} y_n x_n^T \theta^* - \log(1 + e^{x_n^T \theta^*})$$

$$= y_1 x_1 \theta_1^* - \log(1 + e^{x_1\theta^*}) + y_2 x_2 \theta^* - \log(1 + e^{x_2\theta^*}) + y_3 x_3 \theta^* - \log(1 + e^{x_3\theta^*}) + y_4 x_4 \theta^* - \log(1 + e^{x_4\theta^*})$$

$$= 1 - \log(1+e) + 1 - \log(1+e) + 0 - \log(1+e^{-1}) + 0 - \log(1+e^{-1})$$

$$= 2 - \log(1+e)^2(1+e^{-1})^2 = 2 - 2\log(1+e)(1+e^{-1}) = 2 - 2 \times 1.62 = -1.25$$

b) Because we want to result in the same decision boundary in the space given, we still will be on the same decision boundary line. Hence, $\theta' = k\theta^*$, $\theta^* = (-1, -1, 1)$

Likelyhood = $\log p(y \mid x, \theta) = \underbrace{\sum_{n=1}^{N} y_n x_n^T \theta - \log\left(1 + e^{x_n^T \theta}\right)}_{H}$

$$\begin{cases} \text{for } \theta_+^* : H > -1.25 \Rightarrow 2\left(k_1 - \log(1+e)\right) > -1.25 \\ \quad \theta_+^* = k_1 \theta^* \\ \text{for } \theta_-^* : H < -1.25 \Rightarrow 2\left(k_2 - \log(1+e)\right) < -1.25 \\ \quad \theta_-^* = k_2 \theta^* \end{cases}$$

$$\Rightarrow \begin{cases} \text{for } \theta_+^* : k_1 - 1.31 > -0.625 \Rightarrow k_1 > 0.685 \\ \text{for } \theta_-^* : k_2 - 1.31 < -0.625 \Rightarrow k_2 < 0.685 \end{cases}$$

Same boundary: $0 < k_2$

$0 < k_2 < 0.68$

Task 2)

a) for the binary logestic regression model we have:

$$P(y=1 \mid x\theta) = \delta(x^T \theta^*) \quad \text{with} \quad \theta^* = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \quad \text{I}$$

for multi-class logestic regression we have:

$$P(y=t \mid x^T\theta) = f_\theta(X)_t \Rightarrow P(y=1 \mid x^T\theta) = \underbrace{f_\theta(X)_1}_{\text{second element of output vector}} \quad \text{II}$$

no. of class (t)

I, II $\Rightarrow \delta(x^T\theta^*) = f_\theta(B\theta)_2 = \text{Softmax}(BX)$

$$\Rightarrow \frac{1}{1+e^{x^T\theta}} = \frac{e^{x^T B_2}}{e^{x^T B_1} + e^{x^T B_2}} \Rightarrow \frac{1}{1+e^{x\theta}} = \frac{e^{x^T\theta_2}}{e^{x^T\theta_1} + e^{x^T\theta_2}}$$

$$\Rightarrow e^{x^T\theta_2} + e^{x^T(\theta+\theta_2)} = e^{x^T\theta_2} + e^{x^T\theta_1} \Rightarrow x^T(\theta+\theta_2) = x^T\theta_1$$

$$\Rightarrow \theta = \theta_1 - \theta_2 \Rightarrow B = \begin{pmatrix} \theta+\theta_2 \\ \theta_2 \end{pmatrix}$$

b) $\theta^* = (-1,-1,1)^T$

from (a) we have $B = \begin{pmatrix} \theta+\theta_2 \\ \theta_2 \end{pmatrix}$, $\theta_1 - \theta_2 = \theta^* = (-1,-1,1)^T$

let $\theta_2 = (1,2,0) \Rightarrow \theta_1 = (0,1,1) \Rightarrow B = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 2 & 0 \end{pmatrix}$

x	Bx	Softmax	Is Second class derived?	y-predict	y	Correct?
(1,1,3)	(4,3)	(0.73, 0.27)	✗	0	0	✓
(1,2,4)	(6,5)	(0.73, 0.27)	✗	0	0	✓
(1,2,2)	(4,5)	(0.27, 0.73)	✓	1	1	✓
(1,3,3)	(6,7)	(0.27, 0.73)	✓	1	1	✓

Since our assuption was prediction of second class in part a)