Q1-1: Are these statements true or false for the binary logistic regression?

(A) When the linear log odds $\log \frac{p(Y=1|X)}{p(Y=2|X)} \rightarrow +\infty$, the predicted probability $p(Y = 1|X)$ tends to zero.

(B) When $p(X|Y = i)$ are Gaussian $\mathcal{N}(X|\mu_i, I)$, we can derive that the corresponding log odds is linear with respect to X.

1. True, True
2. True, False
3. False, True
4. False, False

Q1-1: Are these statements true or false for the binary logistic regression?

(A) When the linear log odds $\log\frac{p(Y=1|X)}{p(Y=2|X)} \to +\infty$, the predicted probability $p(Y=1|X)$ tends to zero.

(B) When $p(X|Y=i)$ are Gaussian $\mathcal{N}(X|\mu_i, I)$, we can derive that the corresponding log odds is linear with respect to X.

1. True, True
2. True, False
3. False, True ←
4. False, False

(A) When the log odds $a = \log\frac{p(Y=1|X)}{p(Y=2|X)} \to \infty$, we have $p(Y=1|X) = \frac{1}{1+\exp(-a)} \to 1$.

(B) Just as is shown in the lecture.

Q1-2: Please calculate the w and b of log odds as $p(x|y=1) = \mathcal{N}(x|\mu_1, I)$ and $p(x|y=2) = \mathcal{N}(x|\mu_2, I)$,
where $x \in \mathbb{R}^3$, $\mu_1 = [1, 0, 1]^T$, $\mu_2 = [-1, 1, 0\ ]^T$, $p(y=1) = \frac{1}{2}$, and $p(y=2) = \frac{1}{2}$.

1. $w = [0, 1, 1]^T, b = -2 \log 2$
2. $w = [2, -1, 1]^T, b = 0$
3. $w = [-2, 1, -1]^T, b = 1$
4. $w = [2, -1, 1]^T, b = 1$

Q1-2: Please calculate the w and b of log odds as $p(x|y=1) = \mathcal{N}(x|\mu_1, I)$ and $p(x|y=2) = \mathcal{N}(x|\mu_2, I)$,
where $x \in \mathbb{R}^3$, $\mu_1 = [1, 0, 1]^T$, $\mu_2 = [-1, 1, 0\,]^T$, $p(y=1) = \frac{1}{2}$, and $p(y=2) = \frac{1}{2}$.

log odd is

$$a = \ln \frac{p(x|y=1)p(y=1)}{p(x|y=2)p(y=2)} = w^T x + b$$

where

$$w = \mu_1 - \mu_2, \qquad b = -\frac{1}{2}\mu_1^T \mu_1 + \frac{1}{2}\mu_2^T \mu_2 + \ln \frac{p(y=1)}{p(y=2)}$$

1. $w = [0, 1, 1]^T, b = -2\log 2$
2. $w = [2, -1, 1]^T, b = 0$ ⬅
3. $w = [-2, 1, -1]^T, b = 1$
4. $w = [2, -1, 1]^T, b = 1$

By the lecture, we have
$w = \mu_1 - \mu_2 = [2, -1, 1]^T$,
$b = -\frac{1}{2}\mu_1{}^T \mu_1 + \frac{1}{2}\mu_2{}^T \mu_2 + \log \frac{p(y=1)}{p(y=2)} = 0.$

Q2-1: Are these statements true or false for the multiclass logistic regression?

(A) We model $p(y=i|x) = \frac{\exp(a_i)}{\sum_j \exp(a_j)}$ with $a_j = \log \frac{p(x|y=j)p(y=j)}{p(x|y=i)p(y=i)}$.

(B) When $p(x|y=i)$ are Gaussian $\mathcal{N}(x|\mu_i, I)$ and we model $p(y=i|x) = \frac{\exp(a_i)}{\sum_j \exp(a_j)}$, a_i can NOT be linear because $a_i = -\frac{1}{2}x^T x + w_i{}^{\mathrm{T}} x + b_i$ and there's a $-\frac{1}{2}x^T x$ term.

1. True, True
2. True, False
3. False, True
4. False, False

Q2-1: Are these statements true or false for the multiclass logistic regression?

(A) We model $p(y = i|x) = \frac{\exp(a_i)}{\sum_j \exp(a_j)}$ with $a_j = \log \frac{p(x|y=j)p(y=j)}{p(x|y=i)p(y=i)}$.

(B) When $p(x|y = i)$ are Gaussian $\mathcal{N}(x|\mu_i, I)$ and we model $p(y = i|x) = \frac{\exp(a_i)}{\sum_j \exp(a_j)}$, a_i can NOT be linear because $a_i = -\frac{1}{2}x^T x + w_i{}^{\mathrm{T}} x + b_i$ and there's a $-\frac{1}{2}x^T x$ term.

1. True, True
2. True, False
3. False, True
4. False, False ⬅

(A) By the lecture, we have $a_j = \log[p(x|y = j)p(y = j)]$.

(B) Since we have $\exp\left(-\frac{1}{2}x^T x\right)$ both on the numerator and denominator, we can cancel it out and get a linear expression for a_i.

Q2-2: Please calculate the w_i and b_i of a_i for some i in our multiclass logistic regression model. Assume $p(x|y=i) = \mathcal{N}(\mu_i|I)$,
where $x \in \mathbb{R}^3$, $\mu_i = [1,0,1]^T$, $p(y=i) = \frac{1}{2}$.

Cancel out $-\frac{1}{2}x^T x$ and $\ln \frac{1}{(2\pi)^{d/2}}$, we have

$$p(y=i|x) = \frac{\exp(a_i)}{\sum_j \exp(a_j)}, \qquad a_i := \left(w^i\right)^T x + b^i$$

where

$$w^i = \mu_i, \qquad b^i = -\frac{1}{2}\mu_i^T \mu_i + \ln p(y=i)$$

1. $w_i = [1,0,1]^T, b_i \sim -1$
2. $w_i = [2,0,2]^T, b_i \sim -\log 2$
3. $w_i = [0.5,0,0.5]^T, b_i \sim -1 - \log 2$
4. $w_i = [1,0,1]^T, b_i \sim -1 - \log 2$

Q2-2: Please calculate the w_i and b_i of a_i for some i in our multiclass logistic regression model. Assume $p(x|y=i) = \mathcal{N}(\mu_i|I)$,
where $x \in \mathbb{R}^3$, $\mu_i = [1,0,1]^T$, $p(y=i) = \frac{1}{2}$.

Cancel out $-\frac{1}{2}x^T x$ and $\ln \frac{1}{(2\pi)^{d/2}}$, we have

$$p(y=i|x) = \frac{\exp(a_i)}{\sum_j \exp(a_j)}, \qquad a_i := \left(w^i\right)^T x + b^i$$

where

$$w^i = \mu_i, \qquad b^i = -\frac{1}{2}\mu_i^T \mu_i + \ln p(y=i)$$

1. $w_i = [1,0,1]^T, b_i \sim -1$
2. $w_i = [2,0,2]^T, b_i \sim -\log 2$
3. $w_i = [0.5,0,0.5]^T, b_i \sim -1-\log 2$
4. $w_i = [1,0,1]^T, b_i \sim -1-\log 2$ ←

By the lecture, we have
$w_i = \mu_i = [1,0,1]^T$,
$b_i = -\frac{1}{2}{\mu_i}^T \mu_i + \log[p(y=i)] \sim -1-\log 2.$

Q3-1: Please calculate the softmax of (1, 2, 3, 4, 5).

1. (0.067, 0.133, 0.2, 0.267, 0.333)
2. (0, 0.145, 0.229, 0.290, 0.336)
3. (0.012, 0.032, 0.086, 0.234, 0.636)
4. (0.636, 0.234, 0.086, 0.032, 0.012)

Q3-1: Please calculate the softmax of (1, 2, 3, 4, 5).

1. (0.067, 0.133, 0.2, 0.267, 0.333)
2. (0, 0.145, 0.229, 0.290, 0.336)
3. (0.012, 0.032, 0.086, 0.234, 0.636) ←
4. (0.636, 0.234, 0.086, 0.032, 0.012)

By the lecture, we have for some $a = (a_i)$,

$$softmax(a)_i = \frac{\exp(a_i)}{\sum_j \exp(a_j)}.$$

Here:

(A) $\frac{a_i}{\sum_j a_j}$

(B) $\frac{\log(a_i)}{\sum_j \log(a_j)}$

(C) $\frac{\exp(a_i)}{\sum_j \exp(a_j)}$

(D) $\frac{\exp(-a_i)}{\sum_j \exp(-a_j)}$

Q3-2: Please calculate the cross entropy for the following data point and corresponding prediction. Given $\log(0.1) \sim -2.3, \log(0.3) \sim -1.2, \log(0.6) \sim -0.51.$

$$-\log p\left(y = y^{(i)} \middle| x^{(i)}\right) = -\sum_{j=1}^{K} q_j^{(i)} \log p\left(y = j \middle| x^{(i)}\right) = H\left(q^{(i)}, p^{(i)}\right)$$

1. ~ 1.20
2. ~ 2.30
3. ~ 4.02
4. ~ 0.51

One-Hot Label	Prediction
0	0.1
0	0.3
0	0
0	0
0	0
1	0.6
0	0
0	0
0	0
0	0

Q3-2: Please calculate the cross entropy for the following data point and corresponding prediction. Given $\log(0.1) \sim -2.3, \log(0.3) \sim -1.2, \log(0.6) \sim -0.51$.

$$-\log p\left(y = y^{(i)} \middle| x^{(i)}\right) = -\sum_{j=1}^{K} q_j^{(i)} \log p\left(y = j \middle| x^{(i)}\right) = H(q^{(i)}, p^{(i)})$$

1. *~ 1.20*
2. *~ 2.30*
3. *~ 4.02*
4. ~ 0.51 ←

One-Hot Label

$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Prediction

$$\begin{bmatrix} 0.1 \\ 0.3 \\ 0 \\ 0 \\ 0 \\ 0.6 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Here $q = [0, 0, 0, 0, 0, 1, 0, 0, 0, 0]^T, p = [0.1, 0.3, 0, 0, 0, 0.6, 0, 0, 0, 0]^T$.
So we have $H(q, p) = -\sum_{j=1}^{10} q_j \log(p_j) = -1 \times \log(0.6) \sim 0.51$.