<u>שאלה 3</u>

סעיף 1) כיוון ראשון: פונק' הPMF כאשר k=2 תהיה בהינתן סט אימון $S = \{(x_i, y_i)\}_{i=1}^{m}$, כאשר

$y_i \in \{1,2\}$: $P_w(Y_i = 1|x_i) = \frac{e^{w_1^T x_i}}{e^{w_1^T x_i} + e^{w_2^T x_i}} = p_1$, $P_w(Y_i = 2|x_i) = \frac{e^{w_2^T x_i}}{e^{w_1^T x_i} + e^{w_2^T x_i}} = p_2$.

נגדיר $w_1 = w,\ w_2 = 0$ ואז מתקיים: $p_1 = \frac{e^{w^T x_i}}{e^{w^T x_i} + e^{0}} = \frac{e^{w^T x_i}}{e^{w^T x_i} + 1} := p$,

$p_2 = \frac{e^{0}}{e^{w^T x_i} + e^{0}} = \frac{1}{e^{w^T x_i} + 1} := 1 - p$.

כיוון שני: ניקח $w = w_1 - w_2$, ואז פונק' הPMF במקרה הבינארי היא:

$$P_w(Y_i = 1|x_i) = \frac{e^{w^T x_i}}{e^{w^T x_i} + 1} = \frac{e^{w_1^T x_i - w_2{}^T x_i}}{e^{\left(w_1^T x_i - w_2^T x_i\right)} + 1} = \frac{\frac{e^{w_1^T x_i}}{e^{w_2^T x_i}}}{\frac{e^{w_1^T x_i}}{e^{w_2^T x_i}} + \frac{e^{w_2^T x_i}}{e^{w_2^T x_i}}} =$$

$$= \frac{\frac{e^{w_1^T x_i}}{e^{w_2^T x_i}}}{\frac{e^{w_1^T x_i} + e^{w_2^T x_i}}{e^{w_2^T x_i}}} =^{*} \frac{e^{w_1^T x_i}}{e^{w_1^T x_i} + e^{w_2^T x_i}} = p_1$$

$$1 - p_1 = 1 - \frac{e^{w_1^T x_i}}{e^{w_1^T x_i} + e^{w_2^T x_i}} = \frac{e^{w_1^T x_i} + e^{w_2^T x_i} - e^{w_1^T x_i}}{e^{w_1^T x_i} + e^{w_2^T x_i}} = \frac{e^{w_2^T x_i}}{e^{w_1^T x_i} + e^{w_2^T x_i}}$$

*המכנה של המונה והמכנה של המכנה מצטמצמים

המעבר האחרון נובע מההגדרה של p_i, סימון לפונק' הPMF במקרה הבדיד.

סעיף 2) נמצא את ה-w שממקסם את הפונקציה:

$$\sum_{i=1}^{m} log(P_w(Y_i = y_i|x_i)) = \sum_{i=1}^{m} log\left(\frac{e^{w_{y_i}^T x_i}}{\sum_{j=1}^{k} e^{w_{y_j}^T x_i}}\right) =$$

$$= \sum_{i=1}^{m}\left(log\left(e^{w_{y_i}^T x_i}\right) - log\left(\sum_{j=1}^{k} e^{w_{y_j}^T x_i}\right)\right) = \sum_{i=1}^{m}\left(w_{y_i}^T x_i - log\left(\sum_{j=1}^{k} e^{w_{y_j}^T x_i}\right)\right) = \ell_s(w)$$

נעשה זאת באמצעות גזירה והשוואה לאפס:

$$\frac{\partial}{\partial w_k}\ell_s(w) = {}^{*}\sum_{i=1}^{m}\left(I\{y_i = k\}*x_i - \frac{x_i e^{w_k^T x_i}}{\sum_{j=1}^{k} e^{w_j^T x_i}}\right) =$$

$$^{*} = \sum_{i=1}^{m} I\{y_i = k\}*x_i - \sum_{i=1}^{m}\frac{e^{w_k^T x_i}}{\sum_{j=1}^{k} e^{w_j^T x_i}}*x_i = 0$$

בהרצאה ראינו כי אם נגזור את פונקצית הלוג ביחס ל-w ונשווה לאפס נמצא מקסימום גלובלי. בתרגיל זה יש k איברי w, ולכן נגזור בנגזרות חלקיות ונגדיר אינדיקטור עבור הנגזרת של ה-w ה-k-י. נשים לב שבביטוי הימני בשורה שסומנה ב-*, מאחר ואנחנו גוזרים לפי נגזרת חלקית, כל הביטויים התלויים ב-w_i כאשר $i \neq k$ נופלים בגזירה ואנו נשארים רק עם הביטויים הגזורים שהיו תלויים לפני כן ב-w_k.

$$\Longrightarrow \sum_{i=1}^{m} I\{y_i = k\}*x_i = \sum_{i=1}^{m}\left(\frac{e^{w_k^T x_i}}{\sum_{j=1}^{k} e^{w_j^T x_i}}\right)x_i$$

וזו ה-moment matching.

סעיף 3)

חלק a:

נשים לב כי בmulticlass logisitc regression בהינתן וקטור פיצ'רים x_i, ההסתברות שהלייבל y_i הוא k היא p_{ik}, כאשר i מתייחס לדוגמה הו וה-k מתייחס למחלקה ה- k כלומר:

$P(Y_i = k|x_i)$=p_{ik}=$\frac{e^{w_k^T x_i}}{\sum_{j=1}^{K} e^{w_j^T x_i}}$, $w_1, \dots, w_k \in \mathbb{R}^d$, לכן יש מספר וקטורי w כמספר המחלקות שהרגרסיה הלוגיסטית מתמודדת איתה - במקרה הזה 3 מחלקות.

חלק b:

מאחר שמשוואת הרגרסיה הלוגיסטית מכילה מכפלה פנימית בין w לבין וקטור הפיצ'רים x ובנוסף איבר bias, נוכל להחליף את הכתיב $w_k^T x_i + b$ בכתיב המקוצר $\widetilde{w}_k^T \widetilde{x}_i$ כאשר $\widetilde{x}_i = (1,\ x_i)$, $\widetilde{w}_k = (b, w_k)$. כלומר נוסיף כניסה שלישית לווקטורים הללו, ל-x נוסיף את הכניסה 1 על מנת שתכפול את הכניסה ב-w ששווה לערכו של איבר ה-bias. בשאלה נתונים וקטורי פיצ'רים עם שתי כניסות ווקטורי w עם 3 כניסות, לכן כל שנותר הוא להוסיף לכל וקטור פיצ'רים את הכניסה המכילה את הערך 1 כאשר נכפיל בינו לבין וקטור w.

חלק c:

עבור הווקטור x_1:

$$\sum_{j=1}^{3} e^{w_j^T x_1} = e^{(8,-2.5,2)^T(1,1,8)} + e^{(2,0.5,-1.5)^T(1,1,8)} + e^{(-10,2,-0.5)^T(1,1,8)} =$$

$$= 21.7*10^8 + 0.000075 + 0.000006 = 21.7*10^8$$

$$\Rightarrow\ P(Y_1 = 0|x_1) = \frac{e^{w_1^T x_1}}{\sum_{j=1}^{3} e^{w_j^T x_1}} = \frac{21.7*10^8}{21.7*10^8} \approx 1$$

$$\Rightarrow\ P(Y_1 = 1|x_1) = \frac{e^{w_2^T x_1}}{\sum_{j=1}^{3} e^{w_j^T x_1}} = \frac{0.000075}{21.7*10^8} \approx 0$$

$$\Rightarrow\ P(Y_1 = 2|x_1) = \frac{e^{w_3^T x_1}}{\sum_{j=1}^{3} e^{w_j^T x_1}} = \frac{0.000006}{21.7*10^8} \approx 0$$

נבחר את הלייבל 0 להיות הפרדיקציה של x_1.

עבור הווקטור x_2:

$$\sum_{j=1}^{3} e^{w_j^T x_2} = e^{(8,-2.5,2)^T(1,6,-2)} + e^{(2,0.5,-1.5)^T(1,6,-2)} + e^{(-10,2,-0.5)^T(1,6,-2)} =$$

$$= 0.000017 + 2980.958 + 20.0855 = 3001.04$$

$$\Rightarrow\ P(Y_2=0|x_2)=\frac{e^{w_1^T x_1}}{\sum_{j=1}^{3} e^{w_j^T x_1}}=\frac{0.000017}{3001.04}=5.6*10^{-9}\approx 0$$

$$\Rightarrow\ P(Y_2=1|x_2)=\frac{e^{w_2^T x_1}}{\sum_{j=1}^{3} e^{w_j^T x_1}}=\frac{2980.958}{3001.04}=0.9933$$

$$\Rightarrow\ P(Y_2=2|x_2)=\frac{e^{w_3^T x_1}}{\sum_{j=1}^{3} e^{w_j^T x_1}}=\frac{20.0855}{3001.04}=0.0067$$

נבחר את הלייבל 1 להיות הפרדיקציה של x_2.

עבור הווקטור x_3:

$$P(Y_3=0|x_3)=p_{31}=\frac{e^{w_1^T x_3}}{\sum_{j=1}^{3} e^{w_k^T x_i}}=$$

$$=\frac{e^{8*1+(-2.5*(12))+2*4}}{e^{8*1+(-2.5*(12))+2*4}+e^{2*1+(0.5*(12))+(-1.5)*4}+e^{-10*1+(2*(12))+(-0.5)*4}}$$

$$=\frac{e^{-14}}{e^{-14}+e^{2}+e^{12}}=5.108*10^{-12}\approx 0$$

$$P(Y_3=1|x_3)=p_{32}=\frac{e^{w_2^T x_3}}{\sum_{j=1}^{3} e^{w_k^T x_i}}=\frac{e^{2}}{e^{-14}+e^{2}+e^{12}}=4.539*10^{-5}$$

$$P(Y_3=2|x_3)=p_{33}=\frac{e^{w_3^T x_3}}{\sum_{j=1}^{3} e^{w_k^T x_i}}=\frac{e^{12}}{e^{-14}+e^{2}+e^{12}}\approx 1$$

לסיכום עבור x_3 נבחר את הלייבל 2 להיות הפרדיקציה של x_3.