Problem 2

1. (a) 1. Maximize the conditional log-likelihood ~~#~~ is easier. Why?

0/5

2. maximize the log of a function and the function itself is the same

Not a reason.

5/5 (b) Because log function is a monotonically increasing function

35/140

2. (a) $l(w) = \log\left[\prod_{j=1}^{n}\left(\sum_{k=1}^{C} I(y_j=k)\frac{e^{w_k^T x_j}}{\sum_{c'=1}^{C} e^{w_{c'}^T x_j}}\right)\right]$

10/10

$$= \sum_{j=1}^{n}\sum_{k=1}^{C}\left[I(y_j=k)\, w_k^T x_j - \log \sum_{c'=1}^{C} e^{w_{c'}^T x_j}\right]$$

10/10 (b). $\nabla_{w_c} l(w) = \sum_{j=1}^{n}\left(I(y_j=c)\nabla_{w_c} w_c^T x_j - \nabla_{w_c} \ln \sum_{c'=1}^{C} e^{w_{c'}^T x_j}\right)$

$$= \sum_{j=1}^{n}\left[I(y_j=c)\, x_j - \frac{e^{w_c^T x_j}\cdot x_j}{\sum_{c'}^{C} e^{w_{c'}^T x_j}}\right]$$

$$= \sum_{j=1}^{n} x_j\left[I(y_j=c) - P(y_j=c\mid x^j, w)\right]$$

(c).

10/10

When $c'=c$,

$$\frac{\partial l(w)}{\partial w_c^2} = \left\{\sum_{j=1}^{n} x_j\left[I(y_j=c) - P(y_j=c\mid x_j, w)\right]\right\}'$$

$$= \sum_{j=1}^{n} x_j\left[I(y_j=c) - \frac{e^{w_c^T x_j}}{\sum_{c'=1}^{C} e^{w_{c'}^T x_j}}\right]'$$

$$= \sum_{j=1}^{n} x_j (x_j)^T\left[\left(\frac{e^{w_c^T\cdot x_j}}{\sum_{c'=1}^{C} e^{w_{c'}^T x_j}}\right)^2 - \left(\frac{e^{w_c^T\cdot x_j}}{\sum_{c'=1}^{C} e^{w_{c'}^T x_j}}\right)\right]$$

$$= \sum_{j=1}^{n} P(y_j=c\mid x_j, w)\left[P(y_j=c\mid x_j, w) - 1\right] x_j\cdot (x_j)^T$$

when $c' \neq c$

$$\frac{\partial^2 l(w)}{\partial w_c \partial w_{c'}} = \left\{ \sum_{j=1}^{n} x_j \left[ I(y_j = c) - P(y_j = c \mid x_j, w) \right] \right\}'$$

$$= \sum_{j=1}^{n} x_j \left[ I(y_j = c) - \frac{e^{w_c^T x_j}}{\sum_{c'=1}^{} e^{w_{c'}^T x_j}} \right]'$$

$$= \sum_{j=1}^{n} x_j \cdot (x_j)^T \cdot \frac{e^{w_c^T x_j} \cdot e^{w_{c'}^T x_j}}{\left( \sum_{c'=1}^{C} e^{w_{c'}^T x_j} \right)^2}$$

$$= \sum_{j=1}^{n} P(y_j = c \mid x_j, w) \cdot P(y_j = c' \mid x_j, w) \cdot x_j (x_j)^T$$