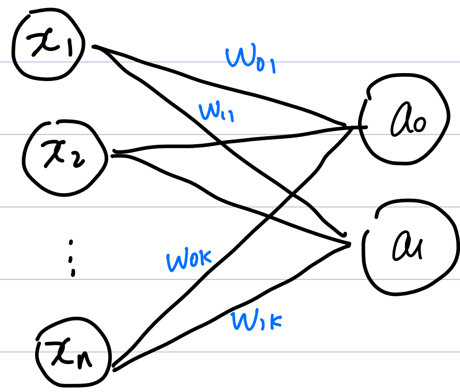
2. Show the following:

(a) For binary outcomes, an ANN model without any hidden layer and with a softmax output layer is equivalent to logistic regression.



W.L.O.G., assume there is only one sample for the following. The loss function and the log-likelihood of multiple samples will be the sum of individual loss or log(L) and the same statement still holds.

An ANN model without any hidden layer and with a softmax activation output layer can be described by:

$$a_0 = \sigma\left(\sum_{i=1}^{n} w_{0i} x_i + b_0\right) \equiv \sigma(z_0) = \frac{e^{z_0}}{e^{z_0} + e^{z_1}} = \frac{1}{1 + e^{z_1 - z_0}}$$

$$a_1 = \sigma\left(\sum_{i=1}^{n} w_{1i} x_i + b_1\right) \equiv \sigma(z_1) = \frac{e^{z_1}}{e^{z_0} + e^{z_1}} = \frac{e^{z_1 - z_0}}{1 + e^{z_1 - z_0}}$$

$$= 1 - a_0$$

If we use the cross-entropy loss, then the objective of the ANN model is to

$$\text{minimize} \quad C = -\left[y_0 \log(a_0) + y_1 \log(a_1) \right]$$

where y_0 and y_1 are the truth labels.

MLE

A logistic regression has the following likelihood:

$$L(y;\pi) = \pi^{y}(1-\pi)^{1-y} \quad \text{where } \pi = \frac{1}{1+e^{-\beta^T x}}$$

where y is the observed label.

The log-likelihood is:

$$\log L(y;\pi) = y\log(\pi) + (1-y)\log(1-\pi)$$

Note that both

$$z_1 - z_0 = \left(\Sigma_{i=1}^{n} w_{1i} x_i + b_1\right) - \left(\Sigma_{i=1}^{n} w_{0i} x_i + b_0\right)$$

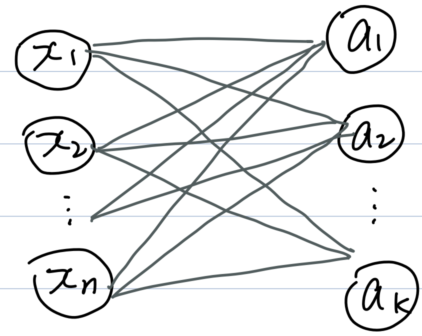
$$= \Sigma_{i=1}^{n} (w_{1i} - w_{0i}) x_i + (b_1 - b_0)$$

and $-\beta^T x = -\left(\Sigma_{i=1}^{n} \beta_i x_i + \beta_0\right)$
are linear combinations of $(x_1, x_2, \dots, x_n, 1)$

Therefore, minimizing C is the same as maximazing $\log L$.
This shows that the above statement is true.

(b) For categorical outcomes with more than two outcome categories, an ANN model without any hidden layer and with a softmax output layer is equivalent to multinomial logistic regression.

ANN



W.L.O.G., assume there is only one sample for the following. The loss function and the log-likelihood of multiple samples will be the sum of individual loss or log(L) and the same statement still holds.

An ANN model without any hidden layer and with a softmax activation output layer can be described by:

$$a_j = \sigma\left(\sum_{i=1}^{n} w_{ji} x_i + b_j\right) \equiv \sigma(z_j) = \frac{e^{z_j}}{\sum_{j'=1}^{K} e^{z_{j'}}} \propto e^{z_j}$$

If we use the cross-entropy loss, then the objective of the ANN model is to

$$\text{minimize} \quad C = -\sum_{j=1}^{K} y_j \log(a_j)$$

where y_j is the truth label for the j^{th} category.

MLE

A multinomial logistic regression has the following likelihood:

$$L(y;\pi) = \prod_{j=1}^{K} \pi_j^{y_j} \quad \text{and} \quad \sum_{j=1}^{K} \pi_j = 1$$

where $\pi_j = \dfrac{e^{-\beta_j^T x}}{\sum_{j=1}^{K} e^{-\beta_j^T x}} \propto e^{-\beta_j^T x}$ in the logistic regression.

The log-likelihood is:

$$\log L(y;\pi) = \sum_{j=1}^{K} y_j \log(\pi_j)$$

where y_j is the observed label for the j^{th} category.

Now we can see that $C = -\log L$. Therefore, minimizing C is the equivalent to maximizing $\log L$. This shows that the statement is true.