Nadav Yakar

036241818

Introduction to Machine Learning – exercise 2:

1.

a. Let w_i be the i'th weight vector of size d out of the columns of a weight matrix w

$$P(Y = \text{i}|X = x_t) = \frac{e^{w_i x_t + b_i}}{\sum_{j=1}^{k} e^{w_j x_t + b_j}}$$

b. Let y_t be a number between 1 to k

$$L(w;b) = -ln[\prod_{t=1}^{m} P(Y = y_t|X = x_t)] = -\sum_{t=1}^{m} ln\left[\frac{e^{w_{y_t} x_t + b_{y_t}}}{\sum_{j=1}^{k} e^{w_j x_t + b_j}}\right] =$$

$$\sum_{t=1}^{m} [ln(\sum_{j=1}^{k} e^{w_j x_t + b_j}) - \ln(e^{w_{y_t} x_t + b_{y_t}})] =$$

$$\sum_{t=1}^{m} [ln(\sum_{j=1}^{k} e^{w_j x_t + b_j}) - w_{y_t} x_t - b_{y_t}]$$

hence the expression we seek is (where w is a matrix):

$$argmin_w(\sum_{t=1}^{m} [ln(\sum_{j=1}^{k} e^{w_j x_t + b_j}) - w_{y_t} x_t - b_{y_t}])$$

c. The update rule of the weight vector of class l at time s would be $w_l{}^s = w_l{}^{s-1} + \eta \frac{\partial L}{\partial w_l}$ and for the bias weight vector of class l at time s would be $b_l{}^s = b_l{}^{s-1} + \eta \frac{\partial L}{\partial b_l}$ for a certain learning factor η that can be determined empirically.

following are the calculations of the various derivatives:

i. l=y_t

$$\frac{\partial L}{\partial w_{y_t}} = \frac{\partial}{\partial w_{y_t}} ln(\sum_{j=1}^{k} e^{w_j x_t + b_j}) - \frac{\partial}{\partial w_{y_t}}(w_{y_t} x_t - b_{y_t}) =$$

$$\frac{e^{w_{y_t} x_t + b_{y_t}}}{\sum_{j=1}^{k} e^{w_j x_t + b_j}} x_t - x_t$$

$$\frac{\partial L}{\partial b_{y_t}} = \frac{\partial}{\partial b_{y_t}} ln(\sum_{j=1}^{k} e^{w_j x_t + b_j}) - \frac{\partial}{\partial b_{y_t}} (w_{y_t} x_t - b_{y_t}) =$$
$$\frac{e^{w_{y_t} x_t + b_{y_t}}}{\sum_{j=1}^{k} e^{w_j x_t + b_j}} - 1$$

ii. l=i, where i represent any class different from y_t:

$$\frac{\partial L}{\partial w_i} = \frac{\partial}{\partial w_i} ln\left(\sum_{j=1}^{k} e^{w_j x_t + b_j}\right) - \frac{\partial}{\partial w_i} (w_{y_t} x_t - b_{y_t}) = \frac{e^{w_i x_t + b_i}}{\sum_{j=1}^{k} e^{w_j x_t + b_j}} x_t$$

$$\frac{\partial L}{\partial b_i} = \frac{\partial}{\partial b_i} ln(\sum_{j=1}^{k} e^{w_j x_t + b_j}) - \frac{\partial}{\partial b_i} (w_{y_t} x_t - b_{y_t}) = \frac{e^{w_i x_t + b_i}}{\sum_{j=1}^{k} e^{w_j x_t + b_j}}$$

2. following please find the practicle sections output graph, where the blue and red lines represents the actual and logistic regression trained probability functions, respectively, over 1000 samples between 0 to 10: