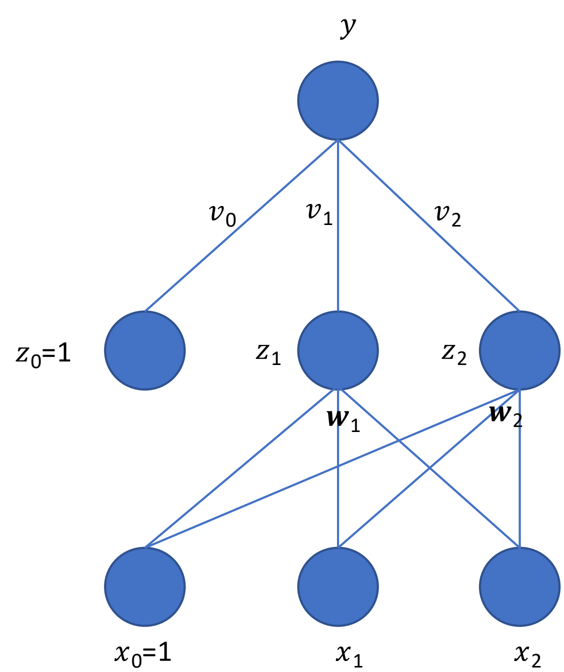
CSCI 5521: Introduction to Machine Learning (Spring 2020)[1]

Homework 4 Solution

1. (**30 points**) Consider the following Multilayer Perceptron (MLP) for binary classification,



we have the following error function:

$$E(\boldsymbol{w_1}, \boldsymbol{w_2}, \boldsymbol{v}|\boldsymbol{X}) = -\sum_t r^t \log y^t + (1 - r^t)\log(1 - y^t),$$

where $y^t = \text{sigmoid}(v_2 z_2 + v_1 z_1 + v_0)$, $z_1^t = \text{ReLU}(w_{1,2}x_2^t + w_{1,1}x_1^t + w_{1,0})$ and $z_2^t = \tanh(w_{2,2}x_2^t + w_{2,1}x_1^t + w_{2,0})$, the rectified linear unit $\text{ReLU}(x)$ is defined as follows

[1]Instructor: Rui Kuang (kuang@cs.umn.edu). TAs: Tianci Song (song0309@umn.edu) and Ruyuan (wanxx199@umn.edu).

$$\text{ReLU}(x) = \begin{cases} 0, & \text{for } x < 0 \\ x, & \text{otherwise} \end{cases}$$

(a) Derive the equations for updating $\{\boldsymbol{w_1}, \boldsymbol{w_2}, \boldsymbol{v}\}$ of the above MLP.

(b) Now, consider shared weights $\boldsymbol{w} = \boldsymbol{w_1} = \boldsymbol{w_2}$. Derive the equations for updating $\{\boldsymbol{w}, \boldsymbol{v}\}$, i.e. to minimize

$$E(\boldsymbol{w}, \boldsymbol{v}|\boldsymbol{X}) = -\sum_t r^t \log y^t + (1 - r^t)\log(1 - y^t),$$

where $y^t = \text{sigmoid}(v_2 z_2 + v_1 z_1 + v_0)$, $z_1^t = \text{ReLU}(w_2 x_2^t + w_1 x_1^t + w_0)$ and $z_2^t = \tanh(w_2 x_2^t + w_1 x_1^t + w_0)$.

Hint: Read Section 11.7.2 to see how Equations 11.23 and 11.24 are derived from Equation 11.22
Hint 2: $\tanh'(x) = 1 - \tanh^2(x)$.
Hint 3: $\text{ReLU}'(x) = \begin{cases} 0, & \text{for } x < 0 \\ 1, & \text{otherwise} \end{cases}$.

Solution: we apply the chain rule to update the parameters v_h and w_h:

$$\begin{aligned}
\Delta v_h &= \eta \sum_t (r^t - y^t) z_h^t \\
\Delta w_1 &= -\eta \frac{\partial E}{\partial w_1} \\
&= -\eta \frac{\partial E}{\partial y^t} \frac{\partial y^t}{\partial z_1^t} \frac{\partial z_1^t}{\partial w_1} \\
&= \begin{cases} 0, \ for \ w_1^T x^t < 0 \\ \eta \sum_t (r^t - y^t) v_1 x^t, \ otherwise \end{cases} \\
\Delta w_2 &= -\eta \frac{\partial E}{\partial w_2} \\
&= -\eta \frac{\partial E}{\partial y^t} \frac{\partial y^t}{\partial z_2^t} \frac{\partial z_2^t}{\partial w_2} \\
&= \eta \sum_t (r^t - y^t) v_2 (1 - (z_1^t)^2) x^t
\end{aligned}$$

When $w_1 = w_2 = w$:

$$\begin{aligned}
\Delta w &= -\eta \frac{\partial E}{\partial w} \\
&= -\eta \frac{\partial E}{\partial y^t} \left(\frac{\partial y^t}{\partial z_1^t} \frac{\partial z_1^t}{\partial w} + \frac{\partial y^t}{\partial z_2^t} \frac{\partial z_2^t}{\partial w} \right) \\
&= \begin{cases} \eta \sum_t (r^t - y^t)(1 - (z^t)^2) v_2 x^t, \ for \ w^T x^t < 0 \\ \eta \sum_t (r^t - y^t)(v_1 + (1 - (z^t)^2) v_2) x^t, \ otherwise \end{cases}
\end{aligned}$$

Updating: $v_h = v_h + \Delta v_h$, $w_h = w_h + \Delta w_h$