Nathaniel Xu (徐星成)　2019280359

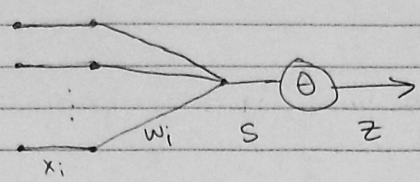
HW1 — Logistic Regression

I have implemented a basic 2-class logistic regression model with the following hyper-parameters

- learning rate = 0.01
- batch size = 32
- max_epoch = 50
- no regularization

The model's computation graph and gradients are calculated:

COMPUTATION GRAPH



$$\begin{cases} E = -\frac{1}{N}\sum_n \left[t^{(n)} \ln z^{(n)} + (1-t^{(n)}) \ln(1-z^{(n)}) \right] \\ z^{(n)} = \frac{1}{1+e^{-s^{(n)}}} \\ s^{(n)} = \underline{w}^T \underline{x}^{(n)} = \sum_i w_i x_i^{(n)} \end{cases}$$

BACKPROP

$$\frac{\partial E}{\partial w_i} = \sum_n \frac{\partial E}{\partial z^{(n)}} \cdot \frac{\partial z^{(n)}}{\partial s^{(n)}} \cdot \frac{\partial s^{(n)}}{\partial w_i}$$

$$\begin{cases} \frac{\partial E}{\partial z^{(n)}} = -\frac{1}{N}\left[\frac{t^{(n)}}{z^{(n)}} - \frac{1-t^{(n)}}{1-z^{(n)}} \right] \\ \frac{\partial z^{(n)}}{\partial s^{(n)}} = \frac{e^{-s}}{(1-e^{-s})^2} = z^{(n)}(1-z^{(n)}) \\ \frac{\partial s^{(n)}}{\partial w_i} = x_i^{(n)} \end{cases}$$

$$\Rightarrow \frac{\partial E}{\partial w_i} = \sum_n \left[-\frac{1}{N}\left(\frac{t^{(n)}}{z^{(n)}} - \frac{1-t^{(n)}}{1-z^{(n)}} \right) z^{(n)}(1-z^{(n)})\, x_i^{(n)} \right]$$

$$= -\frac{1}{N}\sum_n \left[x_i^{(n)} (t^{(n)} - z^{(n)}) \right]$$

$$\Rightarrow \boxed{\frac{\partial E}{\partial \underline{w}} = \frac{1}{N}\sum_n \left[\underline{x}^{(n)} (z^{(n)} - t^{(n)}) \right]}$$

The resulting loss and accuracy graphs are depicted below: