INT 305 Assignment 1

(The deadline is 31st of Oct.)

1. Please write down the whole derivation process to obtain the gradient for logistic regression. (30%)

$z = w^T x$

$y = \sigma(z) = \frac{1}{1+e^{-z}}$

$L_{CE} = -t \log y - (1-t) \log(1-y)$

Therefore

$$\frac{\partial L_{CE}}{\partial w_j} = \frac{\partial L_{CE}}{\partial y} \cdot \frac{\partial y}{\partial z} \cdot \frac{\partial z}{\partial w_j}$$

$$= \left(-\frac{t}{y} + \frac{1-t}{1-y}\right) \cdot y(1-y) \cdot x_j$$

$$= [-t(1-y) + y(1-t)] \cdot x_j$$

$$= [-t + ty + y - ty] \cdot x_j$$

$$= (y-t) x_j$$

$$J = \sum_{i=1}^{N} L_{CE}$$

$$w_j \leftarrow w_j - \frac{\partial J}{\partial w_j}$$

$$= w_j - \frac{\alpha}{N} \sum_{i=1}^{N} (y^{(i)} - t^{(i)}) x_j^{(i)}$$

2. Please write down the whole derivation process to obtain the gradient for multiclass classification with softmax. (40%)

① $$\frac{\partial L_a}{\partial y_T} = -\frac{1}{y_T} (t_T = 1)_{t_{m \neq T} = 0}$$

② $T = k$:

$$\frac{\partial y_T}{\partial z_k} = \frac{\partial y_k}{\partial z_k} = \frac{\partial}{\partial z_k}\left(\frac{e^{z_k}}{\sum_{k'} e^{z_{k'}}}\right)$$

$$= \frac{e^{z_k} \cdot \sum_{k'} e^{z_{k'}} - e^{z_k} \cdot e^{z_k}}{(\sum_{k'} e^{z_{k'}})^2}$$

$$= \frac{e^{z_k}(\sum_{k'} e^{z_{k'}} - e^{z_k})}{\sum_{k'} e^{z_{k'}} \cdot \sum_{k'} e^{z_{k'}}} = y_k(1-y_k)$$

$T \neq k$:

$$\frac{\partial y_T}{\partial z_k} = \frac{\partial}{\partial z_k}\left(\frac{e^{z_T}}{\sum_{k'} e^{z_{k'}}}\right)$$

$$= \frac{-e^{z_T} \cdot e^{z_k}}{(\sum_{k'} e^{z_{k'}})^2}$$

$$= -y_T \left(\frac{e^{z_k}}{\sum_{k'} e^{z_{k'}}}\right)$$

$$= -y_T \cdot y_k$$

③ if $t_k = 1$, $(T = k)$

$$\frac{\partial L_{CE}}{\partial z_k} = \frac{\partial L_{CE}}{\partial y_T} \cdot \frac{\partial y_T}{\partial z_k}$$

$$= -\frac{1}{y_k} \cdot y_k(1-y_k)$$

$$= (y_k - 1)$$

$$= y_k - t_k$$

if $t_k \neq 1$ $(T \neq k)$

$$\frac{\partial L_{CE}}{\partial z_k} = \frac{\partial L_{CE}}{\partial y_T} \cdot \frac{\partial y_T}{\partial z_k} = -\frac{1}{y_T}(-y_T \cdot y_k)$$

$$= y_k = y_k - 0 = y_k - t_k$$

④ $$\frac{\partial z_k}{\partial w} = x$$

⑤ $$\frac{\partial L_{CE}}{\partial w} = \frac{\partial L_{CE}}{\partial z_k} \cdot \frac{\partial z_k}{\partial w}$$

$$= (y_k - t_k) \cdot x$$

3. Please compare the SVM loss and Softmax loss for multiclass classification, please explain which one is better? (30%)

SVM is better than Softmax.

SoftMax versus classification SVMs simply vary in that their goals are objectifiles by many connection weight.
SVM give you values for the classification based on picture. while SoftMax determine the probability of the signaling pathway. SVM to provide stable results and trains faster while softmax might be bogged down by all the calculations if have complex training data.
An SVM on the other hand is a classifier with a higher-loss cost function that result in a maximum margin hyperplace. This can be extend to non-linearly separable problems using kernel approaches by mapping the data into higher dimensional space.