COMPSCI 589 Machine Learning
Assignment 5 Report

Task 1

- A typeset proof showing that the multi-class Softmax cost reduces to the two-class Softmax cost when $C = 2$ and $y_p \in \{1, -1\}$

1.) $g(w_0, \ldots w_{C-1}) = \frac{1}{P} \sum_{p=1}^{P} \left[\log \left(\sum_{j=0}^{C-1} e^{\dot{x}_p^T w_j} \right) - \dot{x}_p^T w_{y_p} \right]$

when $c = 2$, $y_p \in \{-1, 1\}$

$$= \frac{1}{P} \sum_{p=1}^{P} \left[\log \left(e^{\dot{x}_p^T w_0} + e^{\dot{x}_p^T w_1} \right) - \log e^{\dot{x}_p^T w_{y_p}} \right]$$

using $\log a - \log b = \log a/b$

$$= \frac{1}{P} \sum_{p=1}^{P} \left[\log \left(\frac{e^{\dot{x}_p^T w_0} + e^{\dot{x}_p^T w_1}}{e^{\dot{x}_p^T w_{y_p}}} \right) \right]$$

$$= \frac{1}{P} \sum_{p=1}^{P} \left[\log \left(e^{\dot{x}_p^T w_0 - \dot{x}_p^T w_{y_p}} + e^{\dot{x}_p^T w_1 - \dot{x}_p^T w_{y_p}} \right) \right]$$

when $y_p = 1$

$w_{y_p} = w_1$

$$= \frac{1}{P} \sum_{p=1}^{P} \left[\log \left(1 + e^{\dot{x}_p^T w_0 - \dot{x}_p^T w_1} \right) \right]$$

$$= \frac{1}{P} \sum_{p=1}^{P} \left[\log \left(1 + e^{-(-1) \dot{x}_p^T w_0 - (1) \dot{x}_p^T w_1} \right) \right]$$

let $w' = w_1 - w_0$

$$= \frac{1}{P} \sum_{p=1}^{P} \left[\log \left(1 + e^{-\dot{x}_p^T w'} \right) \right]$$

when $y_p = -1$

$w_{y_p} = w_0$

$$= \frac{1}{P} \sum_{p=1}^{P} \left[\log \left(1 + e^{\dot{x}_p^T w_1 - \dot{x}_p^T w_0} \right) \right]$$

$$= \frac{1}{P} \sum_{p=1}^{P} \left[\log\left(1 + e^{\mathring{x}_p^T w'}\right) \right]$$

$$= \frac{1}{P} \sum_{p=1}^{P} \left[\log\left(1 + e^{\mathring{x}_p^T w'}\right) \right]$$

$\therefore$ Using $y_p = +1$ & -1

$$g(w) = \frac{1}{P} \sum_{p=1}^{P} \left[\log\left(1 + e^{-y_p \mathring{x}_p^T w}\right) \right]$$

Task 2

- A typeset proof showing that the multi-class Softmax cost is equivalent to two-class Cross Entropy cost when $C = 2$ and $y_p \in \{0, 1\}$

a) $$g(w_0, \ldots w_{C-1}) = \frac{1}{P} \sum_{p=1}^{P} \left[\log\left(\sum_{j=0}^{C-1} e^{\mathring{x}_p^T w_j} \right) - \mathring{x}_p^T w_{y_p} \right]$$

Taking $C = 2$,

$$= \frac{1}{P} \sum_{p=1}^{P} \left[\log\left(e^{\mathring{x}_p^T w_0} + e^{\mathring{x}_p^T w_1}\right) - \mathring{x}_p^T w_{y_p} \right]$$

where, $\mathring{x}_p^T w_{y_p} = \log e^{\mathring{x}_p^T w_{y_p}}$

$$= \frac{1}{P} \sum_{p=1}^{P} \left[\log\left(e^{\mathring{x}_p^T w_0} + e^{\mathring{x}_p^T w_1}\right) - \log e^{\mathring{x}_p^T w_{y_p}} \right]$$

using, $\log a - \log b = \log \frac{a}{b}$

$$= \frac{1}{P} \sum_{p=1}^{P} \log\left(\frac{e^{\mathring{x}_p^T w_0} + e^{\mathring{x}_p^T w_1}}{e^{\mathring{x}_p^T w_{y_p}}} \right)$$

Taking $y_p = 0$

$$= \frac{1}{P} \sum_{p=1}^{P} \log\left(\frac{e^{\mathring{x}_p^T w_0} + e^{\mathring{x}_p^T w_1}}{e^{\mathring{x}_p^T w_0}} \right)$$

$$= \frac{1}{P} \sum_{p=1}^{P} \log\left(1 + e^{\mathring{x}_p^T (w_1 - w_0)}\right)$$

using, $w' = w_1 - w_0$

$$= -\frac{1}{P} \sum_{p=1}^{P} \log\left(\frac{1}{1 + e^{\mathring{x}_p^T w'}} \right)$$

considering $\alpha = \mathring{x}_p^T w'$

$$= -\frac{1}{P} \sum_{p=1}^{P} \log\left(\frac{1}{1 + e^{\alpha}} \right)$$

$$= -\frac{1}{P} \sum_{p=1}^{P} \log\left(\frac{e^{-\alpha}}{1+e^{-\alpha}}\right)$$

$$= -\frac{1}{P} \sum_{p=1}^{P} \log\left(\frac{\cancel{e^{-\alpha}}}{1+e^{-\alpha}} \times \frac{e^{-\alpha}}{\cancel{e^{-\alpha}}}\right)$$

$$= -\frac{1}{P} \sum_{p=1}^{P} \log\left(\frac{1+e^{-\alpha}-1}{1+e^{-\alpha}}\right)$$

$$= -\frac{1}{P} \sum_{p=1}^{P} \log\left(1 - \frac{1}{1+e^{-\alpha}}\right)$$

$$= -\frac{1}{P} \sum_{p=1}^{P} \log\left(1 - \sigma(\alpha)\right)$$

$$= -\frac{1}{P} \sum_{p=1}^{P} \log\left(1 - \sigma(\mathring{x}_p^T w')\right)$$

when $y_p = 0$

$$= -\frac{1}{P} (1 - y_p) \sum_{p=1}^{P} \log\left(1 - \sigma(\mathring{x}_p^T w)\right)$$

when $y_p = 1$

$$= \frac{1}{P} \sum_{p=1}^{P} \log\left(1 + e^{\mathring{x}_p^T (w_0 - w_1)}\right) \quad , w' = w_1 - w_0$$

$$= -\frac{1}{P} \sum_{p=1}^{P} \log\left(\frac{1}{1+e^{-\mathring{x}_p^T w'}}\right)$$

$$= -\frac{1}{P} \sum_{p=1}^{P} \log\left(\frac{1}{1+e^{-\alpha}}\right) \quad \text{using } \alpha = \mathring{x}_p^T w'$$

$$= -\frac{1}{P} \sum_{p=1}^{P} \log \sigma(\alpha)$$

$$= -\frac{1}{P} \sum_{p=1}^{P} \log \sigma(x_p^{\circ T} w')$$

$\therefore$ when $y_p = 1$

$$= -\frac{1}{P} \sum_{p=1}^{P} y_p \log \sigma(x_p^{\circ T} w')$$

$\therefore$ when $C = 2$ & $y_p \in \{0, 1\}$

$$g(w) = -\frac{1}{P} \sum_{p=1}^{P} \left[y_p \log \sigma(x_p^{\circ T} w') + (1 - y_p) \log (1 - \sigma(x_p^{\circ T} w')) \right]$$

Task 3

- A description of your multi-class classification model solution including local optimization method, the initial values you choose for weights, values of parameters (alpha, max iterations, lambda), and any additional techniques applied. Provide the rationale behind your selection.

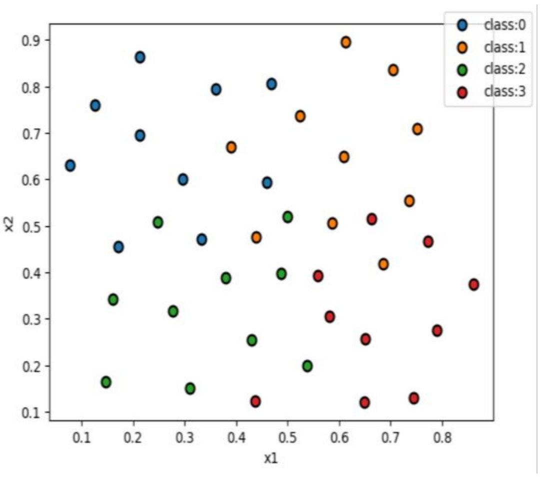
I've chosen the following values for the parameters and weights. I did so since I observed the least misclassifications of 10 in this combination of alpha and max_iteratons. For weights, I chose random values since I was not seeing any significant changes in results with different weights.

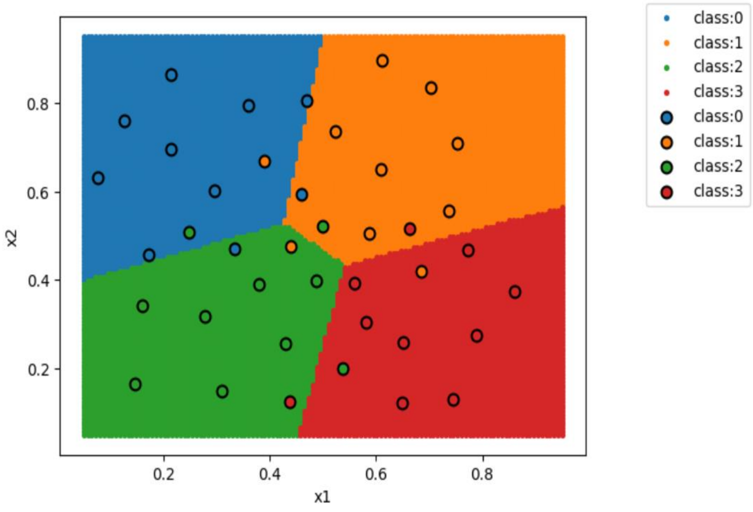
```
# TODO: fill in your code
w_initial = np.random.randn(3,4)
alpha = 1.0
max_its = 2000
```

```
lam = 10 **(-5)
```

I've used gradient decent to do the local optimization for the task as it is one of the most widely used optimization methods.

- A figure showing the original data and regions of your model solution





- The final cost and accuracy of your solution to the 4-class classification task

```
Misclassifications: 10
Accuracy: 75.00%
Final cost:  0.4614795
```