CS391 Assignment 4

Chris-Emio (chrisr98@bu.edu)

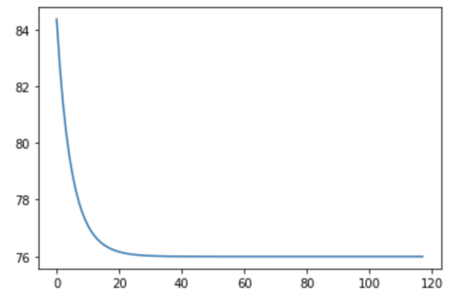
May 1, 2020

Collaborators: Zhan Hao Xu

Exercise 1

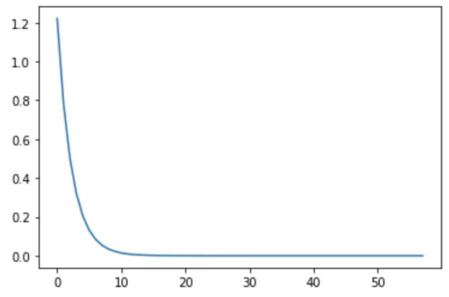
a) $eta : float = 0.001$ gave me the fastest convergence for this equation

```
Exercise 1 part 1

1e-06
Original X:  [2.12553301]
X value found:  [1.7039684]
Number of iterations before convergence:  117
```



b)$eta : float = 0.1$ gave me a convergence almost twice as fast as the part 1.

```
Exercise 1 part 2

1e-06
Original X:  [0.83404401 2.44064899 2.00022875]
X value found:  [0.9999995  2.00000132 2.99999701]
Number of iterations before convergence:  57
```



(Citations for this question)
Slide 18: Gradient Descent

Exercise 2

a) f given by $f(x) = x^2$ is Lipschitz continuous because there exist no number c, s.t. $|f(x) - f(y)| \leq c|x - y|$.

Proof by contradiction: Let's say it is Lipschitz continuous, given that $0 < x, y < c$

$|f(x) - f(y)| \leq c|x-y| \longrightarrow |x^2 - y^2| \leq c|x-y| \longrightarrow |x+y| \times |x-y| \leq c|x-y|$

$|x+y| \leq c$ Hence the contradiction because there is no real number c for which this property holds $|x+y| \leq c$.

b) I don't know

Exercise 3

$\hat{\mathbf{y}} = \text{softmax}(\mathbf{x}) = \frac{e^{\mathbf{x}}}{\|e^{\mathbf{x}}\|_1} = \frac{1}{\sum_{z=1}^{n} e^{x_z}}[e^{x_1}, e^{x_2}, ..., e^{x_n}]$

$$\frac{\delta\hat{\mathbf{y}}}{\delta\mathbf{x}} = \begin{bmatrix} \frac{\delta\hat{y}_1}{\delta x_1} & ... & \frac{\delta\hat{y}_1}{\delta x_n} \\ ... & ... & ... \\ \frac{\delta\hat{y}_n}{\delta x_1} & ... & \frac{\delta\hat{y}_n}{\delta x_n} \end{bmatrix}$$

We want to obtain $\frac{\delta\hat{\mathbf{y}}_i}{\delta\mathbf{x}_i}$. We will be using the quotient rule

If we take for example $f(x) = \frac{g(x)}{h(x)} \longrightarrow f'(x) = \frac{g'(x)h(x) - h'(x)g(x)}{h(x)^2}$

For our softmax case, we have $g(x) = e^{x_i}$ and $h(x) = \sum_{z=1}^{n} e^{x_z}$

We will always get a derivative of e^{x_i} for h_j. But for h_j it would only have a derivative of e^{x_i} if $i = j$

So first we want to compute $\frac{\delta}{\delta\mathbf{x}}\frac{e^{\mathbf{x}}}{\|e^{\mathbf{x}}\|_1}$ for the $i = j$ case.

Using the quotient rule we get

$\frac{\delta\hat{\mathbf{y}}}{\delta\mathbf{x}_i} = \frac{e^{x_i}(\sum_{z=1}^{n} e^{x_z}) - e^{x_i}e^{x_j}}{(\sum_{z=1}^{n} e^{x_z})^2} \longrightarrow \frac{e^{x_j}}{(\sum_{z=1}^{n} e^{x_z})}\frac{(\sum_{z=1}^{n} e^{x_z}) - e^{x_i}}{(\sum_{z=1}^{n} e^{x_z})} = \mathbf{y}_j(1 - \mathbf{y}_i)$

For the $i \neq j$ case:

$\frac{\delta\hat{\mathbf{y}}}{\delta\mathbf{x}_i} = \frac{0 - e^{x_i}e^{x_j}}{(\sum_{z=1}^{n} e^{x_z})^2} \longrightarrow -\frac{e^{x_j}}{(\sum_{z=1}^{n} e^{x_z})}\frac{e^{x_i}}{(\sum_{z=1}^{n} e^{x_z})} = \textbf{-}\mathbf{y}_i\mathbf{y}_j$

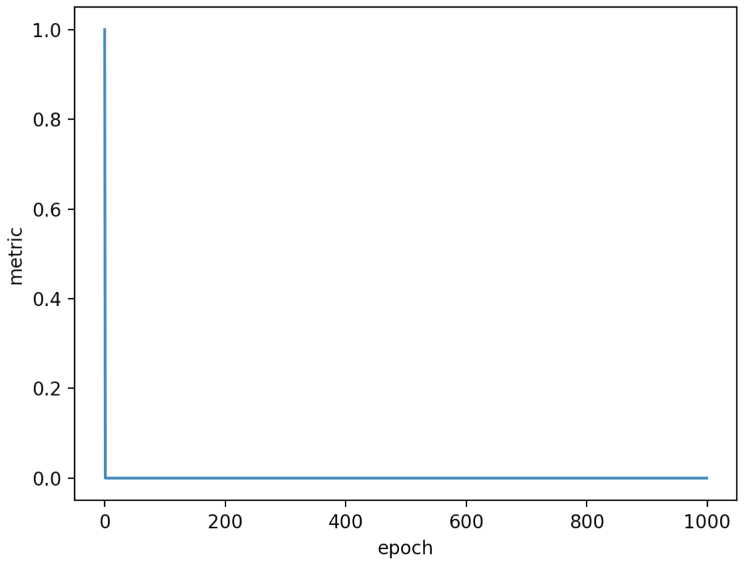
$$\delta_{i,j} = \begin{cases} 1, & \text{if } i = j. \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

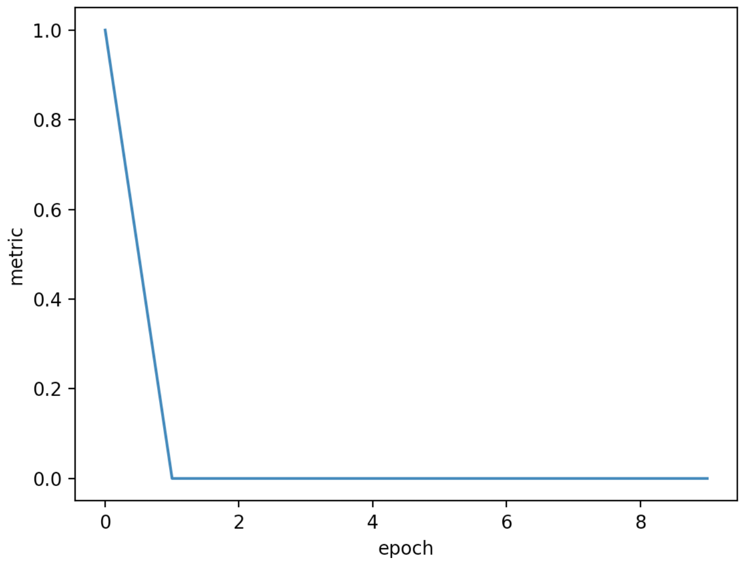
(Citations for this question)
The softmax function and its derivative $\longrightarrow$ *https : //eli.thegreenplace.net/*2016*/the−softmax−function−and−its−derivative*

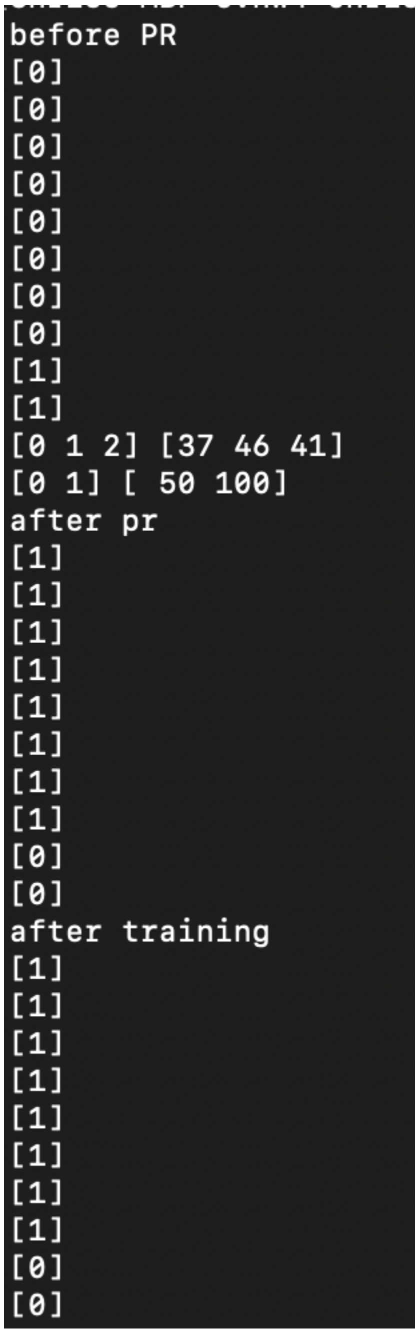
Exercise 4

I created everything and included it in the LR code for exercise 4. After talking to Andrew he explained why and accuracy might not be best for evaluating a model. Especially if the data is imbalanced. In the case of the Iris data, it is imbalanced. So I chose to do precision instead of accuracy. Precision is calculated $\frac{TruePositives}{(TruePositives + FalsePositives)}$ Precision as a metric to show us the performance on your validation data will more meticulous when demonstrating how well the model is able to True positives/negatives.

When displaying the metric results, it looks a bit weird because precision quickly drops to zero. That isn't because the model is inaccurate but rather an problem when converting to the class labels. The data gets inverted. Which is why I shows as 0, but the precision is actually is 100%. Andrew said I didn't have to worry about fixing the conversion so I left it as is. As shown in the pictures below, the precision drops to 0 rather quickly and I also show the inverted data of class labels.







```
before PR
[0]
[0]
[0]
[0]
[0]
[0]
[0]
[0]
[1]
[1]
[0 1 2] [37 46 41]
[0 1] [ 50 100]
after pr
[1]
[1]
[1]
[1]
[1]
[1]
[1]
[1]
[0]
[0]
after training
[1]
[1]
[1]
[1]
[1]
[1]
[1]
[1]
[0]
[0]
```