Module 1 Project – Bike Sharing.

By:
Abubakar Musa –
Almustapha Wakili –

PART A1: SOLUTIONS

A1-1) (B) b.shape=(1, 4, 5) c.shape=(4, 5)

A1-2) (A) torch.Size([60])

A1-3) (B) torch.Size([1, 3, 4, 5])

A1-4) (C) torch.Size([1, 1, 3, 4, 5])

A1-5) (D) torch.Size([3, 1, 4, 5])

PART A2: SOLUTIONS

A)
$w0 > 0$
$w0 + w1 + w2 > 0$
$w0 + w1 < 0$

B)
Given that the scale is arbitrary, we can begin with $w0 = 1$. And the Third equation says $1 + w1 < 0$,
Thus we set $w1 = -2$.
Similarly, the second equation indicates $1 - 2 + w2 > 0$,
Thus, we select $w2 = 2$.
Therefore, among the feasible solutions could be $w0 = 1$, $w1 = -2$, and $w2 = 2$.

PART A3: SOLUTIONS

A3.

$$\frac{\partial J}{\partial w} = \frac{\partial}{\partial w}\left[L(w) + \frac{\lambda}{2}w^2\right]$$

$$= \frac{\partial}{\partial w}\left[\frac{1}{2}(\sigma(wx+b) - t)^2 + \frac{\lambda}{2}w^2\right]$$

$$= \frac{1}{2}\frac{\partial}{\partial w}\left[(\sigma(wx+b)-t)^2\right] + \frac{\lambda}{2}\frac{\partial}{\partial w}\left[w^2\right]$$

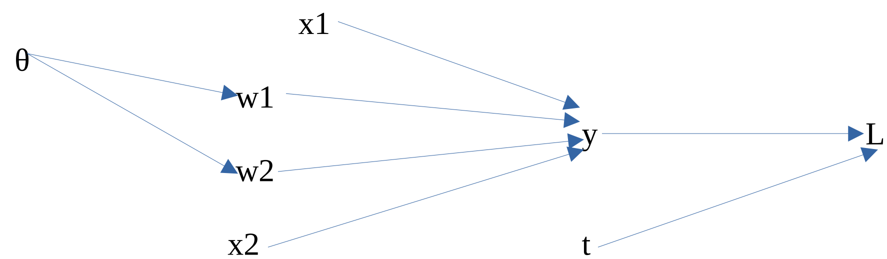
$$= \left[\sigma(wx+b)-t\right]\frac{\partial}{\partial w}\left[\sigma(wx+b)-t\right] + \lambda w$$

$$= (\sigma(wx+b)-t)(\sigma'(wx+b))\frac{\partial}{\partial w}\left[wx+b\right] + \lambda w$$

$$= (\sigma(wx+b)-t)(\sigma'(wx+b))\,x + \lambda w$$

PART A4: SOLUTIONS

A:



B:

$dL/dL = L = 1$
$dL/dy = y = L * (y - t)$
$dL/dw1 = w1 = y * x1$
$dL/dw2 = w2 = y * x2$
$dL/d\theta = \theta = -w1 * \sin\theta + w2 * \cos\theta$

PART A5: SOLUTIONS

Applying the Quotient Rule,
i.e., $f(x)/g(x) = f'(x)g(x)-g'(x)f(x) / g(x)^2$,

$$\sigma'(z) = \partial/\partial z \; (1/ (1 + e^{(-z)}))$$

$$= 0 - (-e^{(-z)}) / (1 + e^{(-z)})^2$$

$$= 1/ (1 + e^{(-z)}) * \; e^{(-z)} / (1 + e^{(-z)})$$

$$= \sigma(z) * (1 - \sigma(z))$$

Part C Follow-up Questions of Part B

C1:

The complete dataset records information from January 1, 2011, and December 31, 2012. For testing, we keep the last 21 days' data. We also keep the last 60 days as a validation set from the remaining data.

Training Set: January 1, 2011 to October 10, 2012
Validation Set: October 11, 2012, to December 10, 2012
Testing Set: 12/31/2012 to 12/11/2012

C2:

A model is said to converge if the if the training loss does not diminish while the training session is on course, or when the training loss varies within a narrow range in the subsequent training rounds.
However, improper hyper-parameter values, flawed back propagation function logics, insufficient training, etc are among the factors leading to non-convergence of models.

C3:

How well does your well-trained model predict the data?
Our model performs well in predicting the data, as indicated by the successful execution of all 5 tests without any errors or failures.

What is the testing MSE under your trained model?
We did use the MSE Function to get the Train and validation losess; however we did'nt call method to print it's value; therefore we can't say the actual MSE value; However Below are our Training and Validation Losses
Training loss: 0.068 ... Validation loss: 0.159

If not, where does it fail? Why does it fail where it does?
Since all tests passed without failures, there are no apparent failures identified in the unit test output.