

CS 615 - Assignment 4

Marcus Sullivan - MS5262

25 July 2021

Part 1 Answers

1. answers:

(a)

$$J = -\ln(\hat{y})$$
$$\frac{\partial J}{\partial \hat{y}} = -\frac{1}{\hat{y}}$$

(b)

$$\frac{\partial J}{\partial x} = W_d^T \bullet \left(\frac{1}{1 + e^{-x}} \circ \left(1 - \frac{1}{1 + e^{-x}} \right) \right) \circ -\frac{1}{\hat{y}}$$

(c)

$$\frac{\partial J}{\partial k} = k \geq 0$$

(d)

$$\frac{\partial J}{\partial W} = z^T$$

(e)

$$\frac{\partial J}{\partial W} = z^T \bullet k \geq 0 \circ W_d^T \bullet \left(\frac{1}{1 + e^{-x}} \circ \left(1 - \frac{1}{1 + e^{-x}} \right) \right) \circ -\frac{1}{\hat{y}}$$

Part 2

During training, I noticed that although my images were converging, the fake observations were often blurry and contained a fair amount of "noise" in what should be the negative space. This made the images easily identifiable (by a human) as synthetic. In order to attempt to address this issue, I took a somewhat irregular approach. I modified the Forward Propagation method of the ReLu layer in order to mimic that of a binary step equation. The altered equation for Forward Propagation in the ReLu layer is as follows:

$$g(x) = \begin{cases} 0 & x \leq 0 \\ 3 & x > 0 \end{cases}$$

Note: The value of 3 was chosen based on empirically observed performance.

The alteration ensures that all pixel values are either fully opaque or black. The gradient for the altered ReLu layer should just be 0, but it was necessary to retain the original ReLu gradient function in order to allow for learning in the Generator's Fully Connected layer. In this way, I am technically breaking the "rules" that we have learned for backwards propagation. The generated fake data from this architecture, however, is far more convincing than that of the original. As such, I have included both architectures in my submission. The videos linked on the next page will show the learning process for both architectures.

Hyperparameters (Standard):

- η : 0.0001
- Batch Size: 100
- No Adam
- Stop after 30 epochs

Hyperparameters (with modified ReLu Layer):

- $\eta_{discriminator}$: 0.0002
- $\eta_{generator}$: 1
- Batch Size: 100
- No Adam
- Stop after 30 epochs
- ReLu output when $i > 0$: 3

Videos:

- [All Digits](#)
- [0](#)
- [1](#)
- [2](#)
- [3](#)
- [4](#)
- [5](#)
- [6](#)
- [7](#)
- [8](#)
- [9](#)