10-601: Homework 3

Due: 9 October 2014 11:59pm (Autolab)

TAs: Henry Gifford, Jin Sun

Name: ARAVIND SELVAN

Andrew ID: ASELVAN

Please answer to the point, and do not spend time/space giving irrelevant details. You should not require more space than is provided for each question. If you do, please think whether you can make your argument more pithy, an exercise that can often lead to more insight into the problem. Please state any additional assumptions you make while answering the questions. You need to submit a single PDF file on autolab. Please make sure you write legibly for grading.

You can work in groups. However, no written notes can be shared, or taken during group discussions. You may ask clarifying questions on Piazza. However, under no circumstances should you reveal any part of the answer publicly on Piazza or any other public website. The intention of this policy is to facilitate learning, not circumvent it. Any incidents of plagiarism will be handled in accordance with CMU's Policy on Academic Integrity.

⋆: Code of Conduct Declaration

- Did you receive any help whatsoever from anyone in solving this assignment? Yes / No.
- If you answered *yes*, give full details: ____________ (e.g. *Jane explained to me what is asked in Question 3.4*)
- Did you give any help whatsoever to anyone in solving this assignment? Yes / No.
- If you answered *yes*, give full details: ____________ (e.g. *I pointed Joe to section 2.3 to help him with Question 2*).

⋆: Notifications

This is the handout for theoretical questions in homework 3, you need to download the handout for programming part as well. If you have any questions, please post it on Piazza or email:

Henry Gifford: hgifford@andrew.cmu.edu

Jin Sun: jins@andrew.cmu.edu

1: Decision Boundaries and Complexity (TA:- Jin Sun)

(a) Figure 1 in appendix shows three decision boundaries. Please list **all possible** decision boundaries for the following classifiers. Please write down the picture labels. No explanations required.

Decision Tree: (C)

Logistic Regression for binary classification: (A)

Perceptrons (Single-layer Neural Networks): (A)

Multi-layer Neural Networks (Single Hidden Layer): (A), (B), (C)

[*8 points*]

(b) For the four classifiers mentioned in part(a), analyse the separability and complexity on several datasets. For separability, you need to state whether the classifier is able to perfectly separate the data points. For complexity, you only need to state whether the decision tree need to be a full tree (at each leaf node there is no attribute to split) to achieve best performance. Please refer to the appendix for detailed explanation on these datasets.

- Logic OR
- Logic XOR
- Majority
- Parity

[Refer to the attached pages in the Same pdf].

[*12 points*]

2: Activation Function (TA:- Jin Sun)

In lectures we use the logistic sigmoid function as the activation function for logistic regression and neural networks. However, there are many other activation functions such as linear function, hyperbolic tangent function and Gaussian function. In this homework, you need to derive the gradient on **one sample** for logistic regression using hyperbolic tangent function as activation function.

The hyperbolic function is defined as follows:

$$\tanh(z) = \frac{\sinh(z)}{\cosh(z)} = \frac{e^z - e^{-z}}{e^z + e^{-z}} \tag{1}$$

and you should calculate the following term:

$$\frac{\partial Loss(\mathbf{w})}{\partial(\mathbf{w})} \tag{2}$$

Let's start with writing down the loss function on one sample for logistic regression:

$$Loss(\mathbf{w}) = -\ln P(Y = y | X = \mathbf{x}, \mathbf{w}) = -y \ln p - (1-y)\ln(1-p) \tag{3}$$

where $p = \tanh(z)$ and $z = \mathbf{w}^T\mathbf{x}$

And then you should derive the derivative and use the chain rule to get the final answer.

[*20 points*]

[Refer to the attached pages in the same pdf].

Total: 40

NAME: ARAVIND SELVAN
ANDREW ID: ASELVAN.

Question 1. (b)

1) Decision Tree:

Separability:

Dataset Name		Can perfectly separate?
* Logic OR	—	Yes
* Logic XOR	—	Yes
* Majority	—	Yes
* Parity	—	Yes.

Complexity:

Dataset Name		Decision tree need to be a full tree?
* Logic OR	—	No
* Logic XOR	—	Yes
* Majority	—	No
* Parity	—	Yes.

2) Logistic Regression for binary classification : (Separability)

Data set Name	Can perfectly separate?
* Logic OR	— Yes
* Logic XOR	— No
* Majority	— No
* Parity	— No.

3) Perceptrons (single-layer Neural Networks) - Separability

Data set Name	Can perfectly separate?
* Logic OR	— Yes.
* Logic XOR	— No.
* Majority	— Yes.
* Parity	— No.

4) Multi-layer Neural Networks (Single Hidden Layer) : - Separability

Dataset Name	Can perfectly separate?
* Logic OR	— Yes.
* Logic XOR	— Yes
* Majority	— Yes.
* Parity	— Yes.

ANDREW ID: ASELVAN.

Question (2)

$$\frac{\partial Loss(w)}{\partial w} = \frac{\partial Loss(w)}{\partial p} \times \frac{\partial p}{\partial z} \times \frac{\partial z}{\partial w}.$$

$$\frac{\partial Loss(w)}{\partial p} \Rightarrow \partial/\partial p \left(-y \ln p - (1-y) \ln(1-p)\right)$$

$$= \left(-y/p + \frac{(1-y)}{1-p}\right)$$

$$= \left(\frac{-y + \cancel{yp} + p - \cancel{py}}{p(1-p)}\right).$$

$$\frac{\partial Loss(w)}{\partial p.} = \frac{(p-y)}{p(1-p)} \quad \text{———} \quad (1)$$

$$\frac{\partial p}{\partial z} = \frac{\partial(\tanh(z))}{\partial z}$$

$$= \frac{\partial}{\partial z}\left(\frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}\right)$$

$$= \partial/\partial z\left(\frac{e^{2z} - 1}{e^{2z} + 1}\right)$$

$$\frac{\partial}{\partial z} = \frac{(e^{2z}+1)\cdot\frac{\partial}{\partial z}(e^{2z}-1) - \frac{\partial}{\partial z}(e^{2z}+1)\cdot(e^{2z}-1)}{(e^{2z}+1)^2}$$

$$= \frac{e^{2z}\cdot\frac{\partial}{\partial z}(2z)\cdot(e^{2z}+1) - e^{2z}\cdot\frac{\partial}{\partial z}(2z)(e^{2z}-1)}{(e^{2z}+1)^2}$$

$$= \frac{2e^{2z}(e^{2z}+1) - 2(e^{2z}-1)\cdot e^{2z}}{(e^{2z}+1)^2}$$

$$= \frac{4e^{2z}}{(e^{2z}+1)^2}$$

$$= (1 - \tanh^2 z)$$

$$\frac{\partial}{\partial z} = (1-p^2) \quad \text{——} \quad (2)$$

$$\frac{\partial z}{\partial w} = \frac{\partial}{\partial w}(w^T x) = x \quad \text{——} \quad (3)$$

Using (1), (2), (3).

$$\frac{\partial\, Loss(w)}{\partial (w)} = \frac{p-y}{p(1-p)} * (1-p^2) * x.$$

ANDREW ID: ASELVAN.

$$\frac{\partial Loss(\omega)}{\partial(\omega)} = \frac{(p-y)}{p} * (1+p) * x.$$

$$= \frac{(p-y)}{\left[\frac{e^z - e^{-z}}{e^z + e^{-z}}\right]} * \left[1 + \left(\frac{e^z - e^{-z}}{e^z + e^{-z}}\right)\right] * x$$

$$= \frac{(p-y)}{\left[\frac{e^z - e^{-z}}{e^z + e^{-z}}\right]} * \left[\frac{e^z + \cancel{e^{-z}} + e^z - \cancel{e^{-z}}}{e^z + e^{-z}}\right] * x$$

$$= \left[\frac{(p-y) * \cancel{(e^z + e^{-z})}}{(e^z - e^{-z})}\right] * \left[\frac{e^z + e^{+z}}{\cancel{(e^z + e^{-z})}}\right] * x$$

$$= \frac{(p-y)}{(e^z - e^{-z})} * 2e^z * x.$$

Now, substituting the values for 'p' and 'z'

$$\frac{\partial loss(\omega)}{\partial(\omega)} = \left[\frac{\left[\left(\frac{e^{w^Tx} - e^{w^Tx}}{e^{w^Tx} + e^{w^Tx}}\right) - y\right] * 2e^{w^Tx} * x}{e^{w^Tx} - e^{-w^Tx}}\right]$$