FALL 2019: CSCI 6522 , AML-II HOMEWORK #1

NAME : PRIYANKA VELAMALA and ID : 2547756

PART (A)

TASK #1:

Draw the graphs of a sigmoid function 1/(1+ e^{-x})and a hyperbolic tangent function, (e^{x} - e^{-x})/(e^{x} + e^{-x}) overlapping each other, with x-axis ranging from -10 to +10. Describe the differences between them.

Solution:

Differences between sigmoid function and a hyperbolic tangent function

Sigmoid function is a an s-shaped curve with output range from 0 to 1 . where hyperbolic tangent is a rescaling of the sigmoid logistic function with output ranging from -1 to 1. Hyperbolic tangent function is better than sigmoid function

sigmoid function: 1/(1+ e^{-x})

hyperbolic function: (e^{x} - e^{-x})/(e^{x} + e^{-x})

In tanh, the input value negative are strongly mapped to negative where as for zero inputs are mapped to zero.
In sigmoid function, towards either end of the sigmoid function, the y values tend to respond very less to changes in x values. This means the gradient at the region is going to be small. Sigmoid function is still very popular in classification problems.
But the gradient is stronger for tanh function than sigmoid function . Tanh is also a very popular and widely used activation

functions. Deciding between the sigmoid function or tanh function will depend on your requirement of gradient strength.

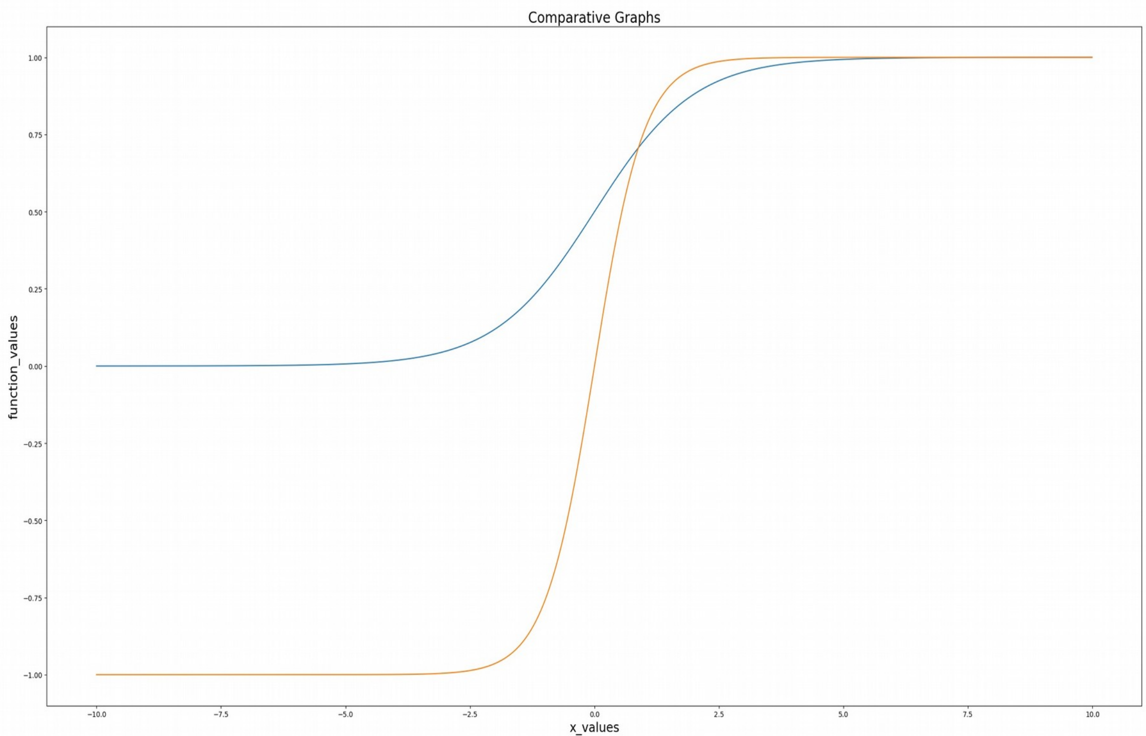
The program for the sigmoid function and hyperbolic tangent function is given below

```
import numpy as np
import matplotlib.pyplot as plt
import math

def sigmoid_funtion(x):
	a=[]
	for item in x:
		a.append(1/(1+math.exp(-item)))
	return a


def hyperbolic_function(x):
	a=[]
	for item in x:

	a.append((math.exp(item)-math.exp(-item))/(math.exp(item)
+math.exp(-item)))
	return a

x=np.arange(-10.,10.,0.01)
curve_1=sigmoid_funtion(x)
curve_2=hyperbolic_function(x)

plt.plot(x,curve_1,label='sigmoid funtion')
plt.plot(x,curve_2,label='hyperbolic function')
plt.xlabel('x_values',fontsize=18)
plt.ylabel('function_values',fontsize=18)
plt.title('Comparative Graphs',fontsize=20)

plt.show()
```

The graph below is shown below for both the sigmoid functions and hyperbolic tangent function.

INDICATION:

Blue : Sigmoid function
Orange : Hyperbolic tangent function



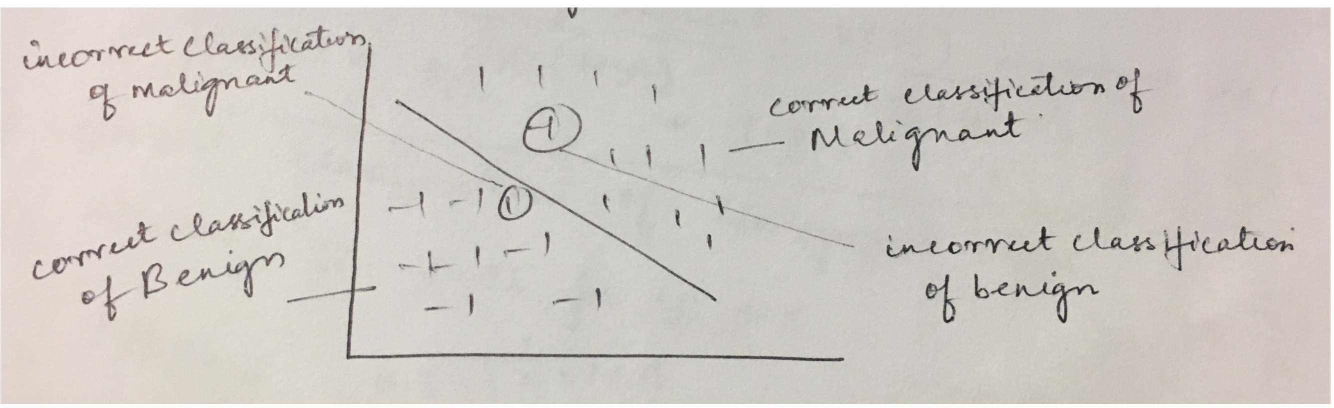
PART(B)

TASK #2:

In chapter 04 (see LECTURE_Chapter_04_Neural_Network_v5) from pages 1-3, and 09-13, we have developed the logistic regression (classification) algorithm using sigmoid function. Similarly, develop the same in this HW#1, however, use the hyperbolic tangent function instead of sigmoid function. Formulate the Bernoulli distribution using hyperbolic tangent function: the formulation should provide higher score for correct classification, and you can relax the perfect formulation of the probability space if needed. You must show by cases, that for correct classification your formulated distribution returns higher value compared to the incorrect classification. Then, for the rest of the part provide your elaborate answer as described in the lecture note for sigmoid function. Assume, Malignant class label is ‘+1’ and Benign class label is ‘-1’.

Solution :

Given Malignant class label is ‘+1’ and Benign class label is ‘-1’.
Lets use the Bernouli distribution:
For correct classification, the outcome is positive numbers(high) whereas for incorrect classification the outcome is negtaive numbers(low).



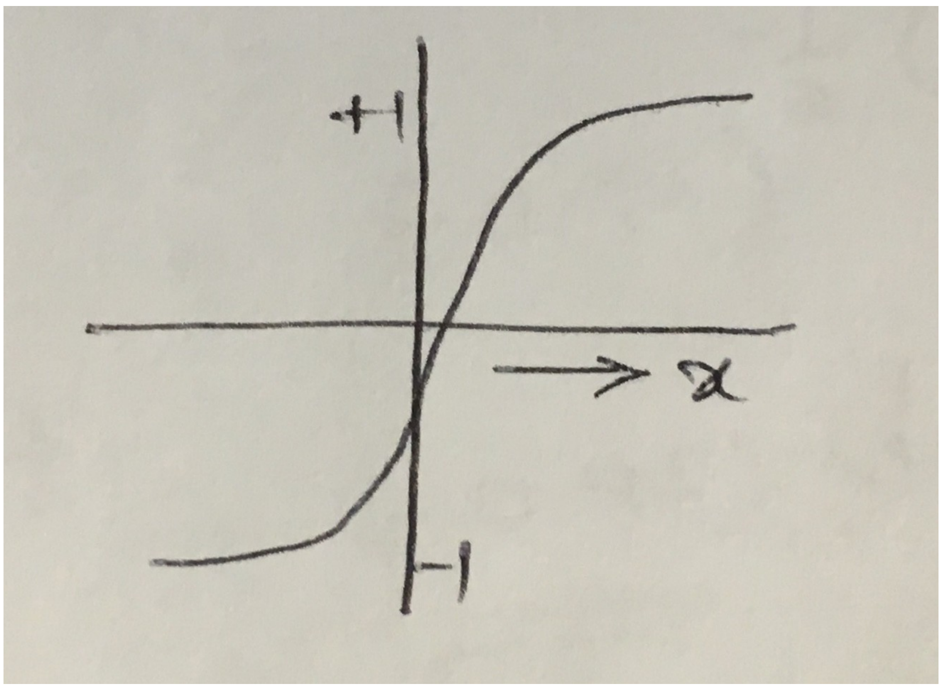
The four different cases in classification :

1. Correct classification of Malignant class
2. Correct classification of Benign class
3. InCorrect classification of Malignant class

4. InCorrect classification of Benign class

The correct classification of the formulated distribuution returns higher value compared to the incorrect clasification is shown below:
For tanh function P(y)= p when y=1 and P(y) = -p when y= -1 where p =f(positive values)=1 and p(negative values) = -1.

where p = (e^{x} - e^{-x})/(e^{x} + e^{-x})



According to the requirement , P(y) is given as :

P(y) = $p^{1/2(1+y)} * -p^{1/2(y-1)}$

1. Correct classification of Malignant class
For y=1 assume p = 0.95
P(y) = $p^{1/2(1+y)} * -p^{1/2(y-1)}$
= $0.95^{1/2(1+1)} * -0.95^{-1/2(1-1)}$
= 0.95(high)

2. Correct classification of Benign class
For y=-1 assume p = -0.95
P(y) = $p^{1/2(1+y)} * -p^{1/2(y-1)}$
= $-0.95^{1/2(1-1)} * -(-0.95)^{-1/2(-1-1)}$

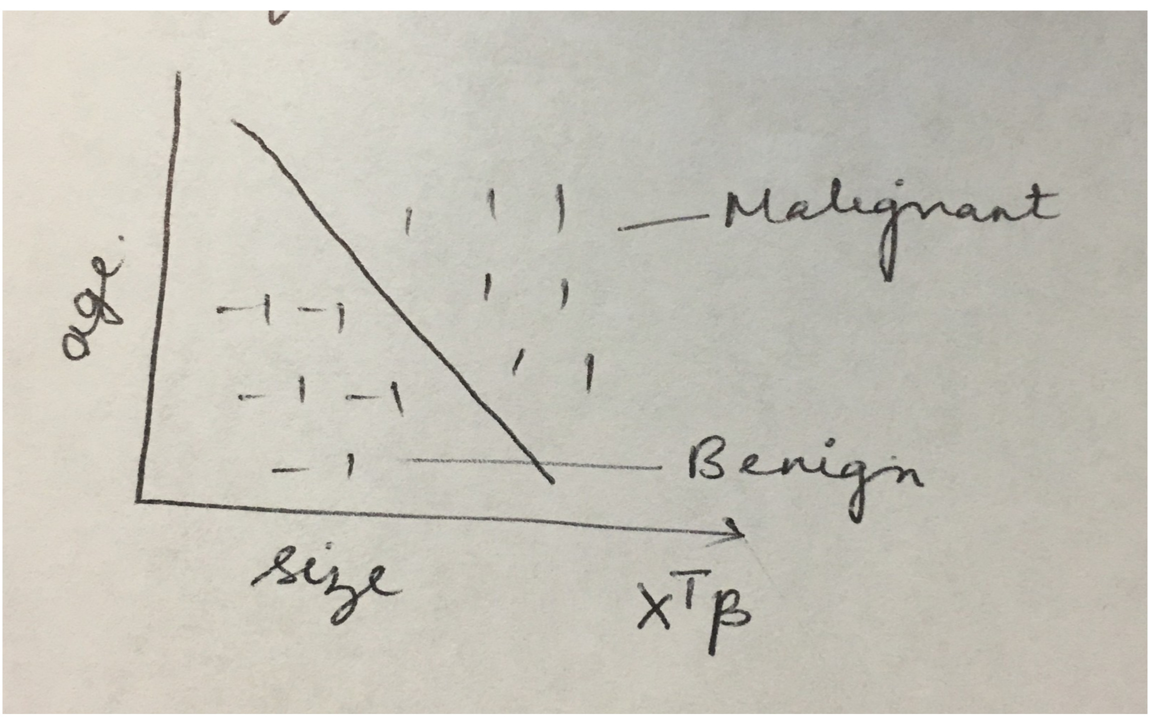
=0.95(high)

3. InCorrect classification of Malignant class
For y=1 assume p = -0.95
P(y) = $p^{1/2(1+y)} * -p^{1/2(y-1)}$
= $-0.95^{1/2(1+1)} * -(-0.95)^{-1/2(1-1)}$
= -0.95(low)

4. InCorrect classification of Benign class
For y=-1 assume p = 0.95
P(y) = $p^{1/2(1+y)} * -p^{1/2(y-1)}$
= $0.95^{1/2(1-1)} * -0.95^{-1/2(-1-1)}$
= -0.95(low)

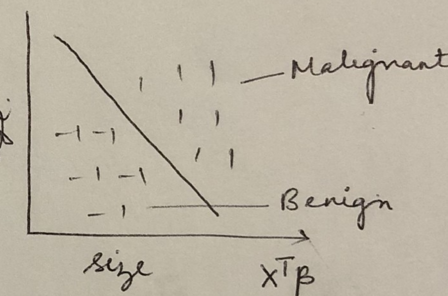
Using the hyperbolic tangent function :

Tumour = { Benign , Malignant }



* Using the hyperbolic tangent function:

Tumour $\in$ {Benign, Malignant}.



$$\eta = P = f_{tanh}(x^T\beta) = \frac{e^{x^T\beta} - e^{-x^T\beta}}{e^{x^T\beta} + e^{-x^T\beta}}$$

$$P(Y) = \begin{cases} P & \text{when } y = 1 \\ -P & \text{when } y = -1 \end{cases}$$

$$L(\beta)_{max} = \prod_{i=1}^{N} P(y_i) = \prod_{i=1}^{N} P^{\frac{1}{2}(1+y_i)} * (-P)^{-\frac{1}{2}(y_i-1)}$$

$$l(\beta) = \log L(\beta) = \log\left(\prod_{i=1}^{N} P^{\frac{1}{2}(1+y_i)} * (-P)^{-\frac{1}{2}(y_i-1)}\right)$$

$$= \sum_{i=1}^{N}\left[\frac{1}{2}(1+y_i)\log(P) - \frac{1}{2}(y_i-1)\log(-P)\right]$$

$$l(\beta) = \sum_{i=1}^{N}\left[\underbrace{\frac{1}{2}(1+y_i)\log\left(\frac{e^{x^T\beta} - e^{-x^T\beta}}{e^{x^T\beta} + e^{-x^T\beta}}\right)}_{l(1)} - \underbrace{\frac{1}{2}(y_i-1)\log\left(\frac{e^{-x^T\beta} - e^{x^T\beta}}{e^{x^T\beta} + e^{-x^T\beta}}\right)}_{l(2)}\right] \rightarrow (1)$$

First solving $l(1)$ and $l(2)$.

$$l(1) = \log\left(\frac{e^{x^T\beta} - e^{-x^T\beta}}{e^{x^T\beta} + e^{-x^T\beta}}\right).$$

$$\frac{\partial}{\partial\beta_j} l(1) = \frac{\partial}{\partial\beta_j}\left(\log(e^{x^T\beta} - e^{-x^T\beta}\right) - \frac{\partial}{\partial\beta_j}\log\left(e^{x^T\beta} + e^{-x^T\beta}\right).$$

$$= \frac{1}{e^{x^T\beta} - e^{-x^T\beta}}\left(x_j e^{x^T\beta} + x_j e^{-x^T\beta}\right) - \frac{1}{e^{x^T\beta} + e^{-x^T\beta}}\left(x e^{x^T\beta} - x e^{-x^T\beta}\right).$$

$$= x_j\left[\frac{e^{x^T\beta} + e^{-x^T\beta}}{e^{x^T\beta} - e^{-x^T\beta}}\right] - x_j\left[\frac{e^{x^T\beta} - e^{-x^T\beta}}{e^{x^T\beta} + e^{x^T\beta}}\right]$$

$$= x_j\left(\frac{1}{p_i}\right) - x_j\left(\frac{1}{p_i}\right).$$

$$l(1) = \frac{x_j}{p} - x_j p. \,//$$

$$l(2) = \log\left(\frac{e^{-x^T\beta} - e^{x^T\beta}}{e^{x^T\beta} + e^{-x^T\beta}}\right).$$

$$\frac{\partial}{\partial\beta_j} l(2) = \frac{\partial}{\partial\beta_j}\left(\log(e^{-x^T\beta} - e^{x^T\beta})\right) - \frac{\partial}{\partial\beta_j}\log\left(e^{x^T\beta} + e^{-x^T\beta}\right).$$

$$= \frac{1}{e^{-x^T\beta} - e^{x^T\beta}}\left(-x_j e^{-x^T\beta} - x_j e^{x^T\beta}\right) - \frac{1}{e^{x^T\beta} + e^{-x^T\beta}}\left(x_j e^{x^T\beta} - x_j e^{-x^T\beta}\right).$$

$$= -x_j\frac{\left(e^{x^T\beta} + e^{-x^T\beta}\right)}{\left(e^{-x^T\beta} - e^{x^T\beta}\right)} - x_j\frac{\left(e^{x^T\beta} - e^{-x^T\beta}\right)}{\left(e^{x^T\beta} + e^{-x^T\beta}\right)}$$

$$= -x_j\left(-\frac{1}{p}\right) - x_j(p)$$

$$= -x_j\left(\frac{1}{p} - p\right) = x_j\left(p - \frac{1}{p}\right).$$

Now substitute $l(1)$ and $l(2)$ in equation 1

$$l(\beta) = \sum_{i=1}^{N}\left[\frac{1}{2}(1+y_i)\,l(1) - \frac{1}{2}(y_i-1)\log l(2)\right]$$

$$= \sum_{i=1}^{N}\left[\frac{1}{2}(1+y_i)\left(\frac{x_j}{p} - x_j p\right) - \frac{1}{2}(y_i-1)\left(x_j p - \frac{x_j}{p}\right)\right]$$

$$= \frac{1}{2}\left[(1+y_i)\left(\frac{x_j}{p} - x_j p\right) - (y_i-1)\left(x_j p - \frac{x_j}{p}\right)\right]$$

$$= \frac{1}{2}\left[\frac{x_j}{p} - x_j p + \frac{x_j y_i}{p} - x_j y_i p - \left(x_j y_i p - \frac{x_j y_i}{p} - x_j p + \frac{x_j}{p}\right)\right]$$

$$= \frac{1}{2}\left[\cancel{\frac{x_j}{p}} - \cancel{x_j p} + \frac{x_j y_i}{p} - x_j y_i p - x_j y_i p + \frac{x_j y_i}{p} + \cancel{x_j p} - \cancel{\frac{x_j}{p}}\right]$$

$$= \frac{1}{2}\left[2\frac{x_j y_i}{p} - 2 x_j y_i p\right]$$

$$= x_j y_i\left(\frac{1}{p} - p\right).$$

$$l(\beta) = \sum_{i=1}^{N} x_j y_i\left(\frac{1}{p} - p\right).$$

We can apply the gradient ascent to get the the values for which log-likelihood is maximized.

$$\beta(t+1) = \beta(t) + \alpha \nabla f(t).$$

But using Newton Raphson Method is more efficient

$$x(t+1) = x(t) - \frac{f'(t)}{f''(t)}.$$

These equations get solved repeatedly, since at each iteration p changes.

Algorithm using hyperbolic tangent function

1. Initilaze β with zero(s)
2. Load X.
3. Load y matrix $\{y = -1$ if Benign
   else
   $y = 1$ if malignant class$\}$.
4. Compute p or η as. $P_i = \dfrac{e^{x_i^T\beta} - e^{-x_i^T\beta}}{e^{x_i^T\beta} + e^{-x_i^T\beta}}$,

   where $i = 1$ to N.

5. Compute next β as $\beta(t+1)$.
6. Check exit condition, else goto step 4.

This algorithm is referred to as iteratively reweighted least squares or IRLS, since each iteration solves the weighted least squares problem.