Homework 5

CpE 4903 Neural Networks and Machine Learning

Logistic Regression Using Gradient Descent

Samiya Zaidi (szaidi4)

September 26, 2023

1. $\boxed{\text{5 points}}$ **Sigmoid Function.**
   Calculate the Sigmoid (logistic) function's output for the following values:
   (a) Input = 0

   Solution: The sigmoid function, $g(x)$, can be written as:

   $$g(x) = \frac{1}{1+e^{-x}}$$

   Hence, $g(0)$ is;

   $$g(0) = \frac{1}{1+e^{-0}} = \frac{1}{1+1}$$
   $$g(0) = 0.5$$

   (b) Input = 1

   Solution: The sigmoid function, $g(x)$, can be written as:

   $$g(x) = \frac{1}{1+e^{-x}}$$

   Hence, $g(1)$ is;

   $$g(1) = \frac{1}{1+e^{-1}}$$
   $$g(1) = 0.7310$$

(c) Input = -1

Solution: The sigmoid function, $g(x)$, can be written as:

$$g(x) = \frac{1}{1+e^{-x}}$$

Hence, $g(-1)$ is;

$$g(-1) = \frac{1}{1+e^{+1}}$$

$$g(-1) = 0.2689$$

2. $\boxed{\text{15 points}}$ **Cost/Loss Function**. You are given a dataset of 4 training samples, each with two features (x_1 and x_2) and a target (y) with binary value (0 or 1).

(a) Assuming the weight vector $A = \{a_0, a_1, a_2\} = \{-2, 1, 1\}$, please calculate the predicted probability, $\hat{y}^{(i)}$, for each of the 4 training samples.

Solution: The sigmoid function can be used to find the probabilities, using the matrix notation:

$$\hat{y} = g(x) = \frac{1}{1+e^{-X \cdot A^T}}$$

Define the matrices first:

$$X = \begin{bmatrix} 1 & 0.5 & 1.5 \\ 1 & 1 & 1 \\ 1 & 3 & 0.5 \\ 1 & 2 & 2 \end{bmatrix}$$

$$A = \begin{bmatrix} -2 & 1 & 1 \end{bmatrix}$$

Now, find $X \cdot A^T$

$$\begin{bmatrix} 1 & 0.5 & 1.5 \\ 1 & 1 & 1 \\ 1 & 3 & 0.5 \\ 1 & 2 & 2 \end{bmatrix} \cdot \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1.5 \\ 2 \end{bmatrix}$$

Now, substitute the matrix in the sigmoid function:

$$\hat{y} = \frac{1}{1+e^{-\begin{bmatrix} 0 \\ 0 \\ 1.5 \\ 2 \end{bmatrix}}} = \begin{bmatrix} 0.5 \\ 0.5 \\ 0.8175 \\ 0.8808 \end{bmatrix}$$

(b) With the predicted probability, $\hat{y}^{(i)}$, please calculate the Log-Loss function, $L(\hat{y}^{(i)}, y^{(i)})$ for each of the 4 training samples. Finally, calculate the cost function, $J(A)$.

Solution: The log loss function can be calculated as;

$$L(\hat{y}^{(i)}, y^{(i)}) = -Y^T \cdot \log(\hat{Y}) - (1 - Y^T) \cdot \log(1 - \hat{Y})$$

$$L(\hat{y}^{(i)}, y^{(i)}) = -\begin{bmatrix} 0 & 0 & 1 & 1 \end{bmatrix} \cdot \log(\begin{bmatrix} 0.5 \\ 0.5 \\ 0.8175 \\ 0.8808 \end{bmatrix}) - (1 - \begin{bmatrix} 0 & 0 & 1 & 1 \end{bmatrix}) \cdot \log(1 - \begin{bmatrix} 0.5 \\ 0.5 \\ 0.8175 \\ 0.8808 \end{bmatrix})$$

The log loss for each training sample;

$$L(\hat{y}^{(i)}, y^{(i)}) = \begin{bmatrix} 0.6931 \\ 0.6931 \\ 0.2015 \\ 0.1269 \end{bmatrix}$$

And the total log loss is:

$$L(\hat{y}^{(i)}, y^{(i)}) = 0.3284 + 1.3862 = 1.7146$$

To calculate the cost function:

$$J(A) = \frac{1}{4} \times 1.7146 = 0.42865$$

3. $\boxed{\text{10 points}}$ **Gradient Calculation**.
With the same weight vector, $A = \{a_0, a_1, a_2\} = \{-2, 1, 1\}$, please calculate the gradient for each of the weight, $\frac{\partial J(A)}{\partial a_i}$.

Solution: The gradient can be calculated with the following equation:

$$\frac{\partial J(A)}{\partial A} = \frac{1}{m} X^T \cdot (\hat{Y} - Y)$$

$$\frac{\partial J(A)}{\partial A} = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0.5 & 1 & 3 & 2 \\ 1.5 & 1 & 0.5 & 2 \end{bmatrix} \cdot (\begin{bmatrix} 0.5 \\ 0.5 \\ 0.8175 \\ 0.8808 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix})$$

$$\frac{\partial J(A)}{\partial A} = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0.5 & 1 & 3 & 2 \\ 1.5 & 1 & 0.5 & 2 \end{bmatrix} \cdot \begin{bmatrix} 0.5 \\ 0.5 \\ -0.1825 \\ -0.1192 \end{bmatrix}$$

$$\frac{\partial J(A)}{\partial A} = \frac{1}{4} \begin{bmatrix} 0.6983 \\ 0.0359 \\ 0.92035 \end{bmatrix}$$

The gradient for each weight is:

$$\frac{\partial J(A)}{\partial A} = \begin{bmatrix} 0.1745 \\ 0.00448 \\ 0.2300 \end{bmatrix} = \begin{bmatrix} \frac{\partial J(A)}{\partial a_0} \\ \frac{\partial J(A)}{\partial a_1} \\ \frac{\partial J(A)}{\partial a_2} \end{bmatrix}$$