Comp 642: Assignment #2 (80 points)

Rice University — Due Date: Thursday, 03/09/2023

Submission Instructions: For coding questions, please submit python notebook along with all the plots and 2-3 paragraphs explaining what you observe and what are your conclusions. Please do familiarize yourself with python notebooks (like Jupyter or Google Collab), they are very convenient, you can run them in your browser without any installations.

1 Forward and Back-propagation in Neural Network - 20 Points

Answer:

STEP1: we create a numpy matrix of 3*2 for X with the values of X as given [[1,-2], [5,-9],[-3,2]] and Y matrix of (3*) with values [[01], [1], [0.8]]

STEP2 : we assign W1 initial values to 6 variables and use these variables to build a numpy array.
W1(11) = 1 , W1(12) =0, W1(13) =1, W1(21) = -1 ,W1(22) = 0 , W1(23) =-1

STEP3: we create a numpy matrix of (2*3) [[W1(11), W1(12), W1(13)], [W1(21), W1(22), W1(23)]]

STEP4 : Similarly create a W2 matrix with elements
[W2(1), W2(2), W2(3)] WHERE W2(1) = 1, W2(2) = 0 AND W2(3) =-1

STEP5: Forward propagation:
- Step A) Apply matrix multiplication on X * W1 which will result in a Matrix temp1 of size (3*3)
- Step B) Vectorize RELU function and apply over the whole matrix to result in a temp2 (3*3) matrix
- Step C) Apply matrix multiplication of temp2 * W2(transpose) which will yield a matrix of (3*1) temp3
- Step D) vectorize sigmoid function and apply it over temp3 matrix which will yield yhat matrix of (3*1)

STEP6: Calculate LOSS function:
- Step A) apply matrix subtraction Y- yhat – call it temp4
- Step B) Apply numpy.square function and apply 1/n*numpy.sum on temp4 to yield Loss (L)
  Note: Numpy.sum and numpy.square are vectorized operations so they are very efficient

STEP7 : Calculating the gradiants:
- Step A) Apply vectorized functions (LOSS, [W2(1), W2(2), W2(3)]) which will yield a matrix of
  W2Grad1 = [$\frac{\partial L}{\partial W2(1)}, \frac{\partial L}{\partial W2(2)}, \frac{\partial L}{\partial W2(3)},$]
  Here W2Gradi is as given below

- Step B) Apply vectorized functions (LOSS, [W1(11), W1(12), W1(13), W1(21), W1(22), W1(23)]) which will yield a matrix of Grad2 =[[$\frac{\partial L}{\partial W1(12)}, \frac{\partial L}{\partial W1(12)}, \frac{\partial L}{\partial W1(13)},$][$\frac{\partial L}{\partial W1(12)}, \frac{\partial L}{\partial W1(22)}, \frac{\partial L}{\partial W1(23)}$]]
  Here W1Gradi,j is given as below
- Step C) Now calculate new W2 matrix as W2 – α *Grad1
- Step D) Now calculate new W1 matrix as W1 – α *np.reshape(Grad1,(3,2))

W2 Grad$_j$

$$\Rightarrow \frac{\partial L}{\partial W^2_j} = -\frac{2}{3}\left(\sum_{i=1}^{3} (Y_i - \hat{Y}_i)(\hat{Y}_i)^2 * e^{-z_i} * \text{TEMP2}[i, j]\right)$$

W1 Grad$_{jk}$

$$\Rightarrow \frac{\partial L}{\partial W^1_{jk}} = -\frac{2}{3}\left(\sum_{i=1}^{3} (Y_i - \hat{Y}_i)(\hat{Y}_i) * e^{-z_i} * \frac{\partial z_i}{\partial W_{jk}}\right).$$

HERE
$$\frac{\partial z_i}{\partial W^1_{jk}} = W^2_k * (X_{ij}) \rightarrow \text{if } (\text{Temp2}[i, k]) > 0$$
$$= 0 \rightarrow \text{if} \leq 0$$

STEP 8: we jump back to Step2 with new W1 and W2 Matrixes and calculate new LOSS values and keep a track of the loss values. Will continue the loop till the time the LOSS value stops reducing and those will be the final W1 and W2 matrixes

Calculations in following pages:

1) ①

$$X = \begin{bmatrix} 1 & -2 \\ 5 & -9 \\ -3 & 2 \end{bmatrix} \quad W^1 = \begin{bmatrix} \underset{1}{w^1_{11}} & \underset{0}{w^1_{12}} & \underset{+1}{w^1_{13}} \\ \underset{-1}{w^1_{21}} & \underset{0}{w^1_{22}} & \underset{+1}{w^1_{23}} \end{bmatrix} \quad W^2 = \begin{bmatrix} \underset{w^2_1}{1} & \underset{w^2_2}{0} & \underset{w^2_3}{-1} \end{bmatrix}$$

FORWARD PROPAGATION:-

$$X * W^1 = \begin{bmatrix} w^1_{11} - 2w^1_{21} & w^1_{12} - 2w^1_{22} & w^1_{13} - 2w^1_{23} \\ 5w^1_{11} - 9w^1_{21} & 5w^1_{12} - 9w^1_{22} & 5w^1_{13} - 9w^1_{23} \\ -3w^1_{11} + 2w^1_{21} & -3w^1_{12} + 2w^1_{22} & -3w^1_{13} + 2w^1_{23} \end{bmatrix}$$

(TEMP₁)

$$(TEMP_2) = RELU(TEMP_1)$$

$$(TEMP_3) = (TEMP_2) * \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \rightsquigarrow \begin{bmatrix} w^2_1 \\ w^2_2 \\ w^2_3 \end{bmatrix}$$

$$Z = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} w^2_1 * RELU(w^1_{11} - 2w^1_{21}) + w^2_2 * RELU(w^1_{12} - 2w^1_{22}) + w^2_3 * RELU(w^1_{13} - 2w^1_{23}) \\ w^2_1 * RELU(5w^1_{11} - 9w^1_{21}) + w^2_2 * RELU(5w^1_{12} - 9w^1_{22}) + w^2_3 * RELU(5w^1_{13} - 9w^1_{23}) \\ w^2_1 * RELU(-3w^1_{11} + 2w^1_{21}) + w^2_2 * RELU(-3w^1_{12} + 2w^1_{22}) + w^2_3 * RELU(-3w^1_{13} + 2w^1_{23}) \end{bmatrix}$$

$$= \begin{bmatrix} 1 * (1+2) + 0 + -1 * RELU(1-2)^{\to 0} \\ 1 * (5+9) + 0 + -1 * RELU(5-9)^{\to 0} \\ 1 * (RELU(-3-2)^{\to 0}) + 0 + -1 * RELU(-3+2)^{\to 0} \end{bmatrix} = \begin{bmatrix} 3 + 0 - 0 \\ 14 + 0 - 0 \\ 0 + 0 + 0 \end{bmatrix}$$

$$Z = \begin{bmatrix} 3 \\ 14 \\ 0 \end{bmatrix}$$

②

$$\hat{Y} = \begin{bmatrix} \frac{1}{1+e^{-z_1}} \\ \frac{1}{1+e^{-z_2}} \\ \frac{1}{1+e^{-z_3}} \end{bmatrix} = \begin{bmatrix} \frac{1}{1+e^{-3}} \\ \frac{1}{1+e^{-15}} \\ \frac{1}{1+e^{-0}} \end{bmatrix} = \begin{bmatrix} 0.9525 \\ 0.9999 \\ 0.5 \end{bmatrix}$$

$$\text{LOSS} = \frac{1}{3} * \begin{bmatrix} (0.1 - 0.9525)^2 \\ + \\ (1 - 0.9999)^2 \\ + \\ (0.8 - 0.5)^2 \end{bmatrix} \approx \frac{1}{3} * 0.8168 \approx 0.2723$$

BACK PROPAGATION:-

$$\text{LOSS}(L) = \frac{1}{n} * \left(\sum_{i=1}^{3} \left(Y_i - \frac{1}{1+e^{-z_i}} \right)^2 \right)$$

$$\frac{\partial L}{\partial w_1^2} = \frac{1}{n} \left(\sum_{i=1}^{3} \left(2 * \left(Y_i - \frac{1}{1+e^{-z_i}} \right) * \frac{\partial \left(\frac{-1}{1+e^{-z_i}} \right)}{\partial w_1^2} \right) \right)$$

$$\Rightarrow \frac{\partial \left(-\frac{1}{1+e^{-z_i}} \right)}{\partial w_1^2} = -\frac{*(-1)}{(1+e^{-z_i})^2} * \frac{\partial (1+e^{-z_i})}{\partial w_1^2}$$

$$= \frac{1}{(1+e^{-z_i})^2} * -e^{-z_i} * \left(\frac{-\partial z_i}{\partial w_1^2} \right)$$

HERE:- $z_i =$ ~~$w_1^2 * (x_{i1} w_{11}^1 + x_{i2} w_{21}^1) + w_2^2 * (x_{i1}$~~ $w_1^2 * \text{RELU}(x_{i1} w_{11}^1 + x_{i2} w_{21}^1) + w_2^2 * \text{RELU}(x_{i1} w_{12}^1 + x_{i2} w_{22}^1) +$ $w_3^2 * \text{RELU}(x_{i1} w_{13}^1 + x_{i2} w_{23}^1)$

$$\Rightarrow \frac{\partial \left(\frac{-1}{1+e^{-z_i}} \right)}{\partial w_1^2} = \frac{-e^{-z_i}}{(1+e^{-z_i})^2} * \text{RELU}(x_{i1} w_{11}^1 + x_{i2} w_{21}^1).$$

$$\Rightarrow \text{HENCE } \frac{\partial L}{\partial w_1^2} = \frac{1}{n} \sum_{i=1}^{3} \left(2 * \left(y_i - \frac{1}{1+e^{-z_i}}\right) * \left(\frac{-e^{-z_i}}{(1+e^{-z_i})^2}\right) * \text{RELU}(\text{TEMP2}_{i1}) \right)$$

$$= \frac{2}{n} * \Bigg(\left((0.1 - 0.9525) * \left(\frac{-e^{-3}}{(1+e^{-3})^2}\right) * \text{RELU}(\underbrace{w_{11}^1 - 2w_{21}^1}_{3}) \right) +$$

$$\left((1 - 0.9999) * \left(\frac{-e^{-14}}{1+e^{-14}}\right) * \text{RELU}(\underbrace{5w_{11}^1 - 9w_{21}^1}_{14}) \right) +$$

$$\left((0.8 - 0.5) * \left(\frac{-e^{0}}{1+e^{-0}}\right) * \text{RELU}(\underbrace{-3w_{11}^1 + 2w_{21}^1}_{0}) \right) \Bigg).$$

$$= \frac{2}{3} * \Big(\left(-0.8525 * (0.9525)^2 * (-0.0498) * 3 \right) +$$

$$\left((0.0001) * (0.9999)^{\to 0} * (0) * 14 \right) +$$

$$\left((0 - 0.3) * (0.5)^{\to 0} * (-1) * 0 \right) \Big)$$

$$= 0.080878 = 0.077034$$

Generic derivation of W_0^2 matrix:

$$\frac{\partial L}{\partial W_j^2} = -\frac{2}{3}\left(\sum_{i=1}^{3}\left(\left(y_i - \frac{1}{1+e^{-z_i}}\right) * \frac{e^{-z_i}}{(1+e^{-z_i})} * \text{RELU}(\text{TEMP1}_{ij})\right)\right)$$

$$\frac{\partial L}{\partial W_2^2} = \frac{-2}{3}\left(0 + 0 + 0 \right) = 0$$

∂

(4)

$$\frac{\partial L}{\partial W_3^2} = \frac{-2}{3}\Big((-0.8525) * (0.9525) * (0.0498) * 0 + 0 + 0\Big) = 0$$

$$\frac{\partial L}{\partial W^2} = \begin{bmatrix} 0.077034 \\ \cancel{0.0080870} \\ 0 \\ 0 \end{bmatrix}$$

$$\Rightarrow new W^2 = \begin{bmatrix} 1 - \alpha\,(0.077034) \\ 0 \\ -1 \end{bmatrix}$$

$$L = \frac{1}{n} * \sum_{i=1}^{3} \left(y_i - \frac{1}{1+e^{-z_i}} \right)^2$$

$$\frac{\partial L}{\partial W_{11}^1} = \frac{1}{n} \sum_{i=1}^{3} \left(2\left(y_i - \frac{1}{1+e^{-z_i}}\right) * \frac{+1}{(1+e^{-z_i})^2} * e^{-z_i} * -\frac{\partial z_i}{\partial W_{11}^1} \right)$$

$$\Rightarrow \frac{\partial z_i}{\partial w_{11}^1} = \frac{\partial}{\partial w_{11}^1}\Big(w_1^2 * RELU(x_{i1} w_{11}^1 + x_{i2} w_{21}^1) + w_2^2 * RELU(x_{i1} w_{12}^1 + x_{i2} w_{22}^1) + w_3^2 * RELU(x_{i1} w_{13}^1 + x_{i2} w_{23}^1) \Big)$$

(5)

$$\Rightarrow \frac{\partial z_1}{\partial w_{11}'} = w_1^2 * (x_{i1}) \rightarrow \text{if } (x_{i1}w_{11}' + x_{i2}w_{21}') \geq 0$$
$$0 \rightarrow \text{if } (x_{i1}w_{11}' + x_{i2}w_{21}') \leq 0$$

Generic form

$$\frac{\partial z_i}{\partial w_{jk}'} = w_k^2 * (x_{ij}) \rightarrow \text{if } (x_{i1}w_{1k}' + x_{i2}w_{2k}') \geq 0$$
$$= 0 \rightarrow \text{if } (x_{i1}w_{1k}' + x_{i2}w_{2k}') < 0$$

$$\Rightarrow \frac{\partial L}{\partial w_{jk}'} = -\frac{2}{3}\left(\sum_{i=1}^{3} \left(y_i - \frac{1}{1+e^{-z_i}}\right) * \frac{1}{(1+e^{-z_i})^2} * e^{-z_i} * \frac{\partial z_i}{\partial w_{jk}'}\right)$$

$$\Rightarrow \frac{\partial L}{\partial w_{11}'} = -\frac{2}{3}\Big(\big((-0.8525) \times (0.9525)^2 * (+0.0498) * (1+1)\big) +$$
$$\big((1-0.9999) * (0) + \dots\big) +$$
$$\big((-0.3) * (0.5)^2 * (1) * 0\big)\Big)$$

$$= \left(\frac{-2}{3}\right) = 0.025678$$

Similary $\frac{\partial L}{\partial w_{12}'} = 0$; $\frac{\partial L}{\partial w_{13}'} = 0$; $\frac{\partial L}{\partial w_{21}'} = -0.0513$

$\frac{\partial L}{\partial w_{22}'} = 0$; $\frac{\partial L}{\partial w_{22}} = 0$

OLD W^1 VALUES:

New W^1 Values

$$\begin{bmatrix} 1 & 0 & 1 \\ -1 & 0 & +1 \end{bmatrix} \rightarrow \begin{bmatrix} 1-\alpha(0.02567) & 0 & 1 \\ -1+\alpha(0.0513) & 0 & +1 \end{bmatrix}$$

OLD W^2 VALUES

NEW W^2 VALUES

$$\begin{bmatrix} 1 & 0 & -1 \end{bmatrix} \rightarrow \begin{bmatrix} 1-\alpha(0.077034) & 0 & -1 \end{bmatrix}$$

Here α is the step size