Assignment # 1

Cours:: Introduction to Deep Learning

Submitted to: Sir Nisar Balti

Submitted by: Abdul Hadi Jalil

Roll no: L1F21BSCS0336

Section: H-6

Question # 1

Part 1

(1)

Ans. Bias is a parameter which helps model to predict values, it is the y-intercept which helps in creating the slope of line.

- Firstly bias is initialized as a random number which then changes later on.

(2)

Ans Learning rate has a critical role in gradient descent as it way to change the model

parameters that best fits in.

(3)

Ans

Sigmoid

Sigmoid is an activation function which is used in binary classification

Disadvantages

~~Advantages~~

The learning rate in sigmoid is very slow because of it's nature.

~~Disadvantage~~

Advantage

The binary nature of sigmoid makes it best to classify binary tasks.

ReLU

ReLU is an activation function which is similar to linear function but it rejects negative value.

Advantages

The learning rate is actually fast because of it's linear nature.

Disadvantage

As it neglects negative values so sometimes it is difficult to make predictions.

(4)

Ans. • Loss function represents the error between the predicted and actual value to train the model correctly. If the loss value is greater means the is not trained correctly.

• It is necessary to train the model if not used than we don't know if the model best fits the data.

(Part 2)

(1)

Ans $x_{h_1} = 0.1 * 1 + 0.3 * 0 + 0.1 \Rightarrow 0.2$

$x_{h_2} = 0.3 * 1 + 0.4 * 0 + 0.2 \Rightarrow 0.5$

$a_{h_1} = \sigma(0.2) = 0.5495$ $\quad a_{h_2} = \sigma(0.5) = 0.6225$ ∴

output layer activation :-

$z_o = 0.5 * 0.5495 + 0.6 * 0.6225 + 0.3$

$= 0.9484$

$y_{pred} = \sigma(0.9484) \Rightarrow 0.7207$

(2)

Ans $E = \frac{1}{2}(y_{pred} - y)^2$

$= \frac{1}{2}(0.7207 - 0.8)^2$

$= 0.00314$

(3)

Ans Step 1:. (Output Layer)

$\delta_0 = (y_{pred} - y_{true})\,\sigma'(z_0)$

$= (0.7207 - 0.8)(0.7207 * 0.2793)$

$= -0.01595$

Step 2: (Gradients for W_2)

$\frac{\partial E}{\partial W_{2,1}} = a_{h_1} \cdot \delta_0 \Rightarrow 0.5498 * (-0.01595) \Rightarrow -0.00877$

$\frac{\partial E}{\partial W_{2,2}} = a_{h_2} \cdot \delta_0 \Rightarrow 0.6225 * (-0.01595) \Rightarrow -0.00993$

Step 3:. (Update W_2 and b_{output})

$W_{2,1} = 0.5 - 0.1(-0.00877) \Rightarrow 0.50088$

$W_{2,2} = 0.6 - 0.1(-0.009937) \Rightarrow 0.60099$

$b_o = 0.3 - 0.1(-0.01595) \Rightarrow 0.30160$

Step 4: (Hidden Layer)

$\delta_{h1} = \sigma'(x_{h1})(W_{2,1}\,\delta_o) \Rightarrow 0.5498 * 0.4502 * (0.5 * -0.015$
$= 0.00197$

$\delta_{h2} = \sigma'(x_{h2})(W_{2,2}\,\delta_o) =$
$= 0.6225 * 0.3775\,(0.6 * -0.01595)$
$= -0.00225$

Step 5:- (Gradient for W_1)

$\frac{\partial E}{\partial W_{1,1}} = x_1\,\delta_{h1} = 1(-0.00197) \Rightarrow -0.00197$

$\frac{\partial E}{\partial W_{1,2}} = x_1\,\delta_{h2} = 1(-0.0025) \Rightarrow -0.0025$

$\frac{\partial E}{\partial W_{2,1}} = x_2\,\delta_{h1} = 0$, $\frac{\partial E}{\partial W_{2,2}} = x_2\,\delta_{h2} = 0$

Step 6:- (Update W_1 and b_{hidden})

$$W_1 = \begin{bmatrix} 0.1 - 0.1(-0.00197) & 0.3 - 0.1(-0.0225) \\ 0.2 & 0.4 \end{bmatrix}$$

$$W_1 = \begin{bmatrix} 0.100020 & 0.300023 \\ 0.2 & 0.4 \end{bmatrix}$$

$b_{h1} = 0.1 - 0.1(-0.00197) = 0.100020$

$b_{h2} = 0.2 - 0.1(-0.08225) \Rightarrow 0.200023$

Question 2

Part 1

(1)

Ans It's solely purpose is to extract features from input image using filters (also known as kernels).

(2)

Ans In CNNs padding adds extra zeros around the edges of input, while stride determines how many the convolution kernel (filters) shifts across the input during each step.