Intruduction to Machine Learning

Dr S.Amini

department: Electrical Engineering

Amirreza Velae 400102222
github repository

Homework 3

May 28, 2023

Intruduction to Machine Learning

Homework 3

Amirreza Velae	400102222
github	repository

Design simple neural network

Design a simple neural network with one hidden-layer that implements the following function:

$$(A \vee \bar{B}) \oplus (\bar{C} \vee \bar{D})$$

Draw the network and determine all its weights.

Soloution

$$(A \vee \bar{B}) \oplus (\bar{C} \vee \bar{D}) = ((A \vee \bar{B}) \wedge (C \wedge D)) \vee ((\bar{A} \wedge B) \wedge (\bar{C} \vee \bar{D}))$$

If you plot the karnaugh map for this special function, it looks like this:

CD \ AB	00	01	11	10
00	0	1	0	0
01	0	1	0	0
11	1	0	1	1
10	0	1	0	0

so the function can be simplified to:

$$\begin{gathered}(\bar{A} \wedge B) \vee (C \wedge D) - (\bar{A} \wedge B \wedge C \wedge D) \\ = (\bar{A} \wedge B \wedge \bar{C}) \vee (\bar{A} \wedge B \wedge C \wedge \bar{D}) \vee (A \wedge C \wedge D) \vee (\bar{A} \wedge \bar{B} \wedge C \wedge D)\end{gathered}$$

we can say:

$$h_1 = \bar{A} \wedge B \wedge \bar{C} = w_1^T x + b_1, \begin{cases} w_1 = \begin{bmatrix} -1 & 1 & 1 & 0 \end{bmatrix} \\ b_1 = -1.5 \end{cases}$$

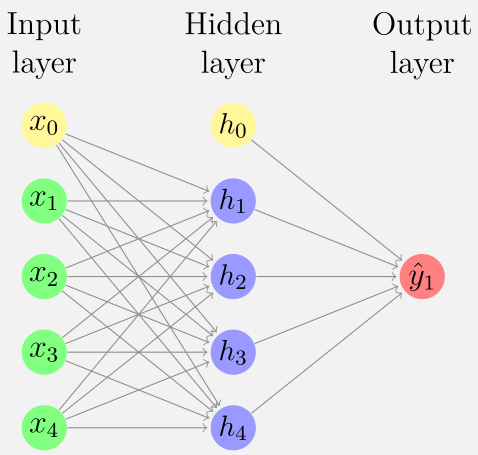
$$A \wedge C \wedge D = w_2^T x + b_2, \begin{cases} w_2 = \begin{bmatrix} 1 & 0 & 1 & 1 \end{bmatrix} \\ b_2 = -2.5 \end{cases}$$

$$\bar{A} \wedge B \wedge C \wedge \bar{D} = w_3^T x + b_3, \begin{cases} w_3 = \begin{bmatrix} -1 & 1 & 1 & -1 \end{bmatrix} \\ b_3 = -1.5 \end{cases}$$

$$\bar{A} \wedge \bar{B} \wedge C \wedge D = w_4^T x + b_4, \begin{cases} w_4 = \begin{bmatrix} -1 & -1 & 1 & 1 \end{bmatrix} \\ b_4 = -1.5 \end{cases}$$

$$y = h_1 \vee h_2 \vee h_3 \vee h_4 = w_5^T x + b_5, \begin{cases} w_5 = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \\ b_5 = -3.5 \end{cases}$$

So the MLP model for $(A \vee \bar{B}) \oplus (\bar{C} \vee \bar{D})$ looks look like this:



Where every link is weighted via w_i or b_i. Also a good activation function is heaviside function.

Vector Derivative

Consider following function:

$$\mathbf{f_1}(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}) = \begin{bmatrix} \frac{1}{\pi} sin(\pi x_2) \\ e^{x_1 - 1} x_2^2 \\ x_1 x_2 \end{bmatrix}, \quad \mathbf{f_2}(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}) = \begin{bmatrix} x_1 + x_2 + x_3 \\ {x_1}^2 + {x_2}^2 + {x_3}^2 \end{bmatrix}$$

$$\mathbf{f}(\mathbf{x}) = (\mathbf{f_2} \circ \mathbf{f_1})(\mathbf{x})$$

Determine $\frac{\partial \mathbf{f}}{\partial \mathbf{x}}$ at point $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

*Note that your solution must follow the methods mentioned in the course slides.

Soloution

In this specific example, n=m. Thus it is possible to calculate the Jacobian matrix of the function f(x) as follows:

$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

First thing to do is to, we have to get v_1 and v_2 at point $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$:

$$v_1 = e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad v_2 = e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

For $\mathbf{f_1}$:

$$J_{\mathbf{f_1}}(\mathbf{x}) = \begin{bmatrix} 0 & cos(\pi x_2) \\ e^{x_1 - 1} x_2^2 & 2e^{x_1-1} x_2 \\ x_2 & x_1 \end{bmatrix}$$

and for $\mathbf{f_2}$:

$$J_{\mathbf{f_2}}(\mathbf{x}) = \begin{bmatrix} 1 & 1 & 1 \\ 2x_1 & 2x_2 & 2x_3 \end{bmatrix}$$

at point $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$:

$$x_2 = \mathbf{f_1}(\mathbf{x_1}) = \begin{bmatrix} \frac{1}{\pi} sin(2\pi) \\ e^{1-1} 2^2 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 4 \\ 2 \end{bmatrix} \qquad J_{\mathbf{f_1}}(\mathbf{x}) = \begin{bmatrix} 0 & cos(2\pi) \\ e^{1-1} 2^2 & 2e^{1-1} 2 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 4 & 4 \\ 2 & 1 \end{bmatrix}$$

then for new v_1 and v_2:

$$v_1 = J_{\mathbf{f_1}}(\mathbf{x}) v_1 = \begin{bmatrix} 0 & 1 \\ 4 & 4 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 4 \\ 2 \end{bmatrix} \qquad v_2 = J_{\mathbf{f_1}}(\mathbf{x}) v_2 = \begin{bmatrix} 0 & 1 \\ 4 & 4 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix}$$

For $\mathbf{f_2}$:

$$x_2 = \mathbf{f_2}(\mathbf{x_1}) = \begin{bmatrix} 0+4+2 \\ 0+16+4 \end{bmatrix} = \begin{bmatrix} 6 \\ 20 \end{bmatrix} \qquad J_{\mathbf{f_2}}(\mathbf{x}) = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 8 & 4 \end{bmatrix}$$

then for new v_1 and v_2:

$$v_1 = J_{\mathbf{f_2}}(x_2) v_1 = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 8 & 4 \end{bmatrix} \begin{bmatrix} 0 \\ 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 \\ 40 \end{bmatrix} \qquad v_2 = J_{\mathbf{f_2}}(x_2) v_2 = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 8 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 36 \end{bmatrix}$$

Thus:

$$J_f(\mathbf{x}) = \begin{bmatrix} v_1 & v_2 \end{bmatrix} = \begin{bmatrix} 6 & 6 \\ 40 & 36 \end{bmatrix}$$

One Convolutional Layer

Consider we have one convolutional layer with following equation:

$$\begin{bmatrix} z_1 \\ \vdots \\ z_m \end{bmatrix} = \begin{bmatrix} k_1 & \cdots & k_d & & & & \\ & k_1 & \cdots & k_d & & & \\ & & & \ddots & & & \\ & & & k_1 & \cdots & k_d \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} b \\ \vdots \\ b \end{bmatrix}$$

and also we know:

$$\frac{\partial L}{\partial z_i} = \alpha_i$$

calculate $\frac{\partial L}{\partial k_j}$ and $\frac{\partial L}{\partial b}$ in terms of x_i and z_i.

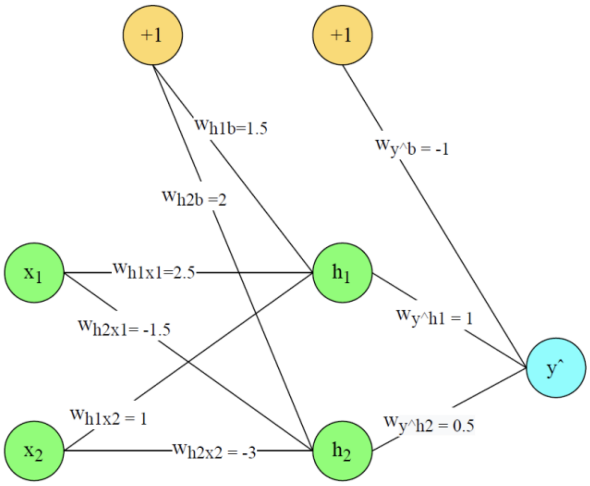
Soloution

$$\frac{\partial L}{\partial k_j} = \sum_{i=1}^{m} \frac{\partial L}{\partial z_i} \frac{\partial z_i}{\partial k_j} = \sum_{i=1}^{m} \alpha_i x_{i+j-1}$$
$$\frac{\partial L}{\partial b} = \sum_{i=1}^{m} \frac{\partial L}{\partial z_i} \frac{\partial z_i}{\partial b} = \sum_{i=1}^{m} \alpha_i$$

Backpropagation Algorithm

The following image shows a two-layer neural network with two nodes in the hidden layer and one node in the output. x_1 and x_2 are two inputs to the network. Each node has a bias with value of 1.
Assume that the value of the learning rate is 0.1 and the activation function is sigmoid in both hidden-layer and output-layer.

- Calculate the value at nodes $\hat{y}, h_1, h_2$ for input $\{x_1 = 0, x_2 = 1\}$.
- Execute one step of backpropagation algorithm for the previous input in part (a) and output $y = 1$.
- Calculate the updated weights for the hidden-layer and output-layer (a total of 9 weights) by executing one step of the gradient descent algorithm.



Part I

$$\hat{y} = \sigma(\mathbf{w_{y^{h_0}} h_0} + \mathbf{w_{y^{h_1}} h_1} + \mathbf{w_{y^{h_2}} h_2})$$
$$h_1 = \sigma(\mathbf{w_{h_1^{x_1}} x_1} + \mathbf{w_{h_1^{x_2}} x_2} + \mathbf{w_{h_1^b} b})$$
$$h_2 = \sigma(\mathbf{w_{h_2^{x_1}} x_1} + \mathbf{w_{h_2^{x_2}} x_2} + \mathbf{w_{h_2^b} b})$$

Given values for weights:

$$\mathbf{w_{y^{h_0}}} = -1 \qquad \mathbf{w_{y^{h_1}}} = 1 \qquad \mathbf{w_{y^{h_2}}} = 0.5$$
$$\mathbf{w_{h_1^{x_1}}} = 2.5 \qquad \mathbf{w_{h_1^{x_2}}} = 1 \qquad \mathbf{w_{h_1^b}} = 1.5$$
$$\mathbf{w_{h_2^{x_1}}} = -1.5 \qquad \mathbf{w_{h_2^{x_2}}} = -3 \qquad \mathbf{w_{h_2^b}} = 2$$

also $\mathbf{h_0} = +1$ and $\mathbf{b} = +1$.
So:

$$h_1 = \sigma(2.5 \times 0 + 1 \times 1 + 1.5 \times 1) = \sigma(1 + 1.5) = \sigma(2.5) = 0.924$$
$$h_2 = \sigma(-1.5 \times 0 + -3 \times 1 + 2 \times 1) = \sigma(-3 + 2) = \sigma(-1) = 0.268$$
$$\hat{y} = \sigma(-1 \times 1 + 1 \times 0.993 + 0.5 \times 0.268) = \sigma(-1 + 0.924 + 0.134) = \sigma(0.058) = 0.514$$

Part II

To execute back propagation algorithm for one step, we need to calculate $\frac{\partial L}{\partial \mathbf{w_{y^{h_0}}}}$, $\frac{\partial L}{\partial \mathbf{w_{y^{h_1}}}}$, $\frac{\partial L}{\partial \mathbf{w_{y^{h_2}}}}$, $\frac{\partial L}{\partial \mathbf{w_{h_1^{x_1}}}}$, $\frac{\partial L}{\partial \mathbf{w_{h_1^{x_2}}}}$, $\frac{\partial L}{\partial \mathbf{w_{h_1^{b}}}}$, $\frac{\partial L}{\partial \mathbf{w_{h_2^{x_1}}}}$, $\frac{\partial L}{\partial \mathbf{w_{h_2^{x_2}}}}$, $\frac{\partial L}{\partial \mathbf{w_{h_2^{b}}}}$.

$$w_{t+1} = w_t - \eta \frac{\partial L}{\partial w_t}$$

$$\frac{\partial L}{\partial w_t} = \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial w_t} = \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial h_0}\frac{\partial h_0}{\partial w_t} + \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial h_1}\frac{\partial h_1}{\partial w_t} + \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial h_2}\frac{\partial h_2}{\partial w_t}$$

$$\frac{\partial L}{\partial \hat{y}} = \frac{2}{2}(\hat{y} - y) = (\hat{y} - y) = (0.514 - 1) = -0.486$$

$$\frac{\partial \hat{y}}{\partial h_0} = \sigma(h_0)(1 - \sigma(h_0)) = \sigma(1)(1 - \sigma(1)) = 0.731 \times 0.269 = 0.196$$

$$\frac{\partial \hat{y}}{\partial h_1} = \sigma(h_1)(1 - \sigma(h_1)) = \sigma(0.924)(1 - \sigma(0.924)) = 0.715 \times 0.075 = 0.054$$

$$\frac{\partial \hat{y}}{\partial h_2} = \sigma(h_2)(1 - \sigma(h_2)) = \sigma(0.268)(1 - \sigma(0.268)) = 0.567 \times 0.732 = 0.415$$

$$\frac{\partial h_0}{\partial w_{y^{h_0}}} = 1 \qquad \frac{\partial h_1}{\partial w_{y^{h_0}}} = 0 \qquad \frac{\partial h_2}{\partial w_{y^{h_0}}} = 0$$

$$\frac{\partial h_0}{\partial w_{y^{h_1}}} = 0 \qquad \frac{\partial h_1}{\partial w_{y^{h_1}}} = 1 \qquad \frac{\partial h_2}{\partial w_{y^{h_1}}} = 0$$

$$\frac{\partial h_0}{\partial w_{y^{h_2}}} = 0 \qquad \frac{\partial h_1}{\partial w_{y^{h_2}}} = 0 \qquad \frac{\partial h_2}{\partial w_{y^{h_2}}} = 1$$

$$\frac{\partial h_0}{\partial w_{h_1^{x_1}}} = 0 \qquad \frac{\partial h_1}{\partial w_{h_1^{x_1}}} = \sigma(h_1)(1 - \sigma(h_1))x_1 = 0.924 \times 0.076 \times 1 = 0.070$$

$$\frac{\partial h_0}{\partial w_{h_1^{x_2}}} = 0 \qquad \frac{\partial h_1}{\partial w_{h_1^{x_2}}} = \sigma(h_1)(1 - \sigma(h_1))x_2 = 0.924 \times 0.076 \times 1 = 0.070$$

$$\frac{\partial h_0}{\partial w_{h_1^{b}}} = 0 \qquad \frac{\partial h_1}{\partial w_{h_1^{b}}} = \sigma(h_1)(1 - \sigma(h_1)) = 0.924 \times 0.076 = 0.070$$

$$\frac{\partial h_0}{\partial w_{h_2^{x_1}}} = 0 \qquad \frac{\partial h_2}{\partial w_{h_2^{x_1}}} = \sigma(h_2)(1 - \sigma(h_2))x_1 = 0.268 \times 0.732 \times 1 = 0.196$$

$$\frac{\partial h_0}{\partial w_{h_2^{x_2}}} = 0 \qquad \frac{\partial h_2}{\partial w_{h_2^{x_2}}} = \sigma(h_2)(1 - \sigma(h_2))x_2 = 0.268 \times 0.732 \times 1 = 0.196$$

$$\frac{\partial h_0}{\partial w_{h_2^{b}}} = 0 \qquad \frac{\partial h_2}{\partial w_{h_2^{b}}} = \sigma(h_2)(1 - \sigma(h_2)) = 0.268 \times 0.732 = 0.196$$

$$\frac{\partial L}{\partial w_{y^{h_0}}} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial h_0} \frac{\partial h_0}{\partial w_{y^{h_0}}} = -0.486 \times 0.196 \times 1 = -0.095$$
$$\frac{\partial L}{\partial w_{y^{h_1}}} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial h_1} \frac{\partial h_1}{\partial w_{y^{h_1}}} = -0.486 \times 0.054 \times 1 = -0.026$$
$$\frac{\partial L}{\partial w_{y^{h_2}}} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial h_2} \frac{\partial h_2}{\partial w_{y^{h_2}}} = -0.486 \times 0.415 \times 1 = -0.202$$
$$\frac{\partial L}{\partial w_{h_1^{x_1}}} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial h_0} \frac{\partial h_0}{\partial w_{h_1^{x_1}}} = -0.486 \times 0.196 \times 0.070 = -0.006$$
$$\frac{\partial L}{\partial w_{h_1^{x_2}}} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial h_0} \frac{\partial h_0}{\partial w_{h_1^{x_2}}} = -0.486 \times 0.196 \times 0.070 = -0.006$$
$$\frac{\partial L}{\partial w_{h_1^{b}}} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial h_0} \frac{\partial h_0}{\partial w_{h_1^{b}}} = -0.486 \times 0.196 \times 0.070 = -0.006$$
$$\frac{\partial L}{\partial w_{h_2^{x_1}}} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial h_0} \frac{\partial h_0}{\partial w_{h_2^{x_1}}} = -0.486 \times 0.196 \times 0.196 = -0.019$$
$$\frac{\partial L}{\partial w_{h_2^{x_2}}} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial h_0} \frac{\partial h_0}{\partial w_{h_2^{x_2}}} = -0.486 \times 0.196 \times 0.196 = -0.019$$
$$\frac{\partial L}{\partial w_{h_2^{b}}} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial h_0} \frac{\partial h_0}{\partial w_{h_2^{b}}} = -0.486 \times 0.196 \times 0.196 = -0.019$$

Part III

to calculate updated weights, we use the following formula:

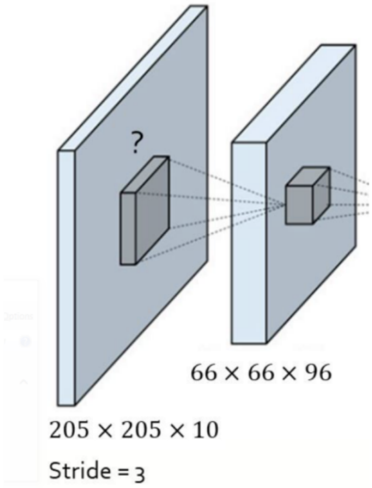
$$w_{new} = w_{old} - \eta \frac{\partial L}{\partial w_{old}}$$

where $\eta = 0.1$ is the learning rate and w_{old} is the weight before the update.

$$w_{y^{h_0}} = -1 - 0.1 \times 0.095 = -1.0095$$
$$w_{y^{h_1}} = 1 - 0.1 \times 0.026 = 0.9974$$
$$w_{y^{h_2}} = 0.5 - 0.1 \times 0.202 = 0.4798$$
$$w_{h_1^{x_1}} = 2.5 - 0.1 \times -0.006 = 2.5006$$
$$w_{h_1^{x_2}} = 1 - 0.1 \times -0.006 = 1.0006$$
$$w_{h_1^{b}} = 1.5 - 0.1 \times -0.006 = 1.5006$$
$$w_{h_2^{x_1}} = -1.5 - 0.1 \times -0.019 = -1.4981$$
$$w_{h_2^{x_2}} = -3 - 0.1 \times -0.019 = -2.9981$$
$$w_{h_2^{b}} = 2 - 0.1 \times -0.019 = 2.0019$$

Model Parameters

Consider the following two-layer convolutional network.



- Based on the input and output dimensions shown in the figure, determine the size of the kernel used for this operation.
- Determine the number of trainable parameters in this layer.
- Calculate the number of multiplication operations required to obtain the output.

Part I

Via CNN we have the following formula:

$$\frac{N - F + 2P}{S} + 1 = M$$

The input dimension is $205 \times 205 \times 10$ and the output dimension is $66 \times 66 \times 96$ and the stride is 3.
The kernel size is $k \times k \times 10$ and the output dimension is $\frac{205-k}{3} + 1 \times \frac{205-k}{3} + 1 \times 96$

$$\frac{205 - k}{3} + 1 = 66$$
$$\frac{205 - k}{3} = 65$$
$$205 - k = 195$$
$$k = 10$$

Part II

Total trainable parameters of this CNN is equal to $(h_w \times w_w \times C + 1) \times D$.
So in this specific problem, we have $((10 \times 10 \times 10) + 1) \times 96 = 96096$ trainable parameters.

Part III

Total number of equations regarding to convolution formula:

$$z_{pqd} = \phi(b_c + \sum_{c=0}^{C-1} \sum_{i=0}^{h_w-1} \sum_{j=0}^{w_w-1} w_{ijkc}\ \hat{x}_{(h_s \times p+i)(w_s \times q+j)c})$$

is 2 for calculating indices of $\mathbf{x}$ and one for multiplying w and x. if we ignore the calculation of indices of x, then we have $(66 \times 66 \times 96) \times (10 \times 10 \times 10) = 418176000$ multiplying operations.

End of Homework 3