Ömer Yılmaz
2017721048

1) a) 1st Update:

$$\|[1\ 1\ 0\ 0]-[0.97\ 0.49\ 0.14\ 0.92]\| = \sqrt{0.03^2+0.51^2+(-0.14)^2+(-0.92)^2} = 1.0616$$

$$\|[1\ 1\ 0\ 0]-[0.96\ 0.04\ 0.93\ 0.76]\| = \sqrt{0.04^2+0.96^2+(-0.93)^2+(-0.76)^2} = 1.5381$$

w1 wins: $w_1(2) = w_1(1) + 0.6\times[0.03\ 0.51\ -0.14\ -0.92] = [0.9880\ 0.7960\ 0.0560\ 0.3680]$

2nd Update

$$\|[0\ 0\ 1\ 1]-[0.9880\ 0.7960\ 0.0560\ 0.3680] = \sqrt{(-0.9880)^2+(-0.7960)^2+0.9440^2+0.6320^2} = 1.7030$$

$$\|[0\ 0\ 1\ 1]-[0.96\ 0.04\ 0.93\ 0.76] = \sqrt{(-0.96)^2+(-0.04)^2+0.07^2+0.24^2} = 0.9928$$

w2 wins: $w_2(3) = w_2(2) + 0.6\times[-0.96\ -0.04\ 0.07\ 0.24] = [0.3840\ 0.0160\ 0.9720\ 0.9040]$

3rd Update:

$$\|[1\ 0\ 1\ 0]-[0.9880\ 0.7960\ 0.0560\ 0.3680] = \sqrt{0.0120^2+(-0.7960)^2+0.9440^2+(-0.3680)^2} = 1.2885$$

$$\|[1\ 0\ 1\ 0]-[0.3840\ 0.0160\ 0.9720\ 0.9040] = \sqrt{0.6160^2+(-0.0160)^2+0.0280^2+(-0.9040)^2} = 1.0944$$

w2 wins: $w_2(4) = w_2(3) + 0.6\times[0.6160\ -0.0160\ 0.0280\ -0.9040] = [0.7536\ 0.0064\ 0.9888\ 0.3616]$

4th Update:

$$\|[1\ 0\ 0\ 0]-[0.9880\ 0.7960\ 0.0560\ 0.3680]\| = \sqrt{0.0120^2+(-0.7960)^2+(-0.0560)^2+(-0.3680)^2} = 0.8788$$

$$\|[1\ 0\ 0\ 0]-[0.7536\ 0.0064\ 0.9888\ 0.3616]\| = \sqrt{0.2464^2+(-0.0064)^2+(-0.9888)^2+(-0.3616)^2} = 1.0813$$

w1 wins: $w_1(5) = w_1(4) + 0.6\times[0.0120\ -0.7960\ -0.0560\ -0.3680] = [0.9952\ 0.3184\ 0.224\ 0.1472]$

5th Update:

$$\|[0\ 0\ 0\ 1]-[0.9952\ 0.3184\ 0.0224\ 0.1472]\| = \sqrt{(-0.9952)^2+(-0.3184)^2+(-0.0224)^2+0.8528^2} = 1.3489$$

$$\|[0\ 0\ 0\ 1]-[0.7536\ 0.0064\ 0.9888\ 0.3616]\| = \sqrt{(-0.7536)^2+(-0.0064)^2+(-0.9888)^2+0.6384^2} = 1.3976$$

w1 wins: $w_1(6) = w_1(5) + 0.6\times[-0.9952\ -0.3184\ -0.0224\ 0.8528] = [0.3981\ 0.1274\ 0.0090\ 0.6589]$

Ömer Yılmaz
2017721048

① 6th Update:

$\|[1\ 1\ 0\ 0]-[0.3981\ 0.1274\ 0.0090\ 0.6589]\| = \sqrt{0.6019^2+0.8726^2+(-0.0090)^2+(-0.6589)^2} = 1.2482$

$\|[1\ 1\ 0\ 0]-[0.7536\ 0.0064\ 0.9888\ 0.3616]\| = \sqrt{0.2464^2+0.9936^2+(-0.9888)^2+(-0.3616)^2} = 1.4685$

$w_1$ wins: $w_1(7) = w_1(6) + 0.3\times[0.6019\ 0.8726\ -0.0090\ -0.6589] = [0.5787\ 0.3892\ 0.0063\ 0.4612]$

7th Update:

$\|[0\ 0\ 1\ 1]-[0.5787\ 0.3892\ 0.0063\ 0.4612]\| = \sqrt{(-0.5787)^2+(-0.3892)^2+0.9937^2+0.5388^2} = 1.3232$

$\|[0\ 0\ 1\ 1]-[0.7536\ 0.0064\ 0.9888\ 0.3616]\| = \sqrt{(-0.7536)^2+(-0.0064)^2+0.0112^2+0.6384^2} = 0.9877$

$w_2$ wins: $w_2(8) = w_2(7) + 0.3\times[-0.7536\ -0.0064\ 0.0112\ 0.6384] = [0.5275\ 0.0045\ 0.9922\ 0.5531]$

8th Update:

$\|[1\ 0\ 1\ 0]-[0.5787\ 0.3892\ 0.0063\ 0.4612]\| = \sqrt{0.4213^2+(-0.3892)^2+0.9937^2+(-0.4612)^2} = 1.2366$

$\|[1\ 0\ 1\ 0]-[0.5275\ 0.0045\ 0.9922\ 0.5531]\| = \sqrt{0.4725^2+(-0.0045)^2+0.0078^2+(-0.5531)^2} = 0.7275$

$w_2$ wins: $w_2(9) = w_2(8) + 0.3\times[0.4725\ -0.0045\ 0.0078\ -0.5531] = [0.6693\ 0.0031\ 0.9945\ 0.3872]$

9th Update:

$\|[1\ 0\ 0\ 0]-[0.5787\ 0.3892\ 0.0063\ 0.4612]\| = \sqrt{0.4213^2+(-0.3892)^2+(-0.0063)^2+(-0.4612)^2} = 0.7360$

$\|[1\ 0\ 0\ 0]-[0.6693\ 0.0031\ 0.9945\ 0.3872]\| = \sqrt{0.3307^2+(-0.0031)^2+(-0.9945)^2+(-0.3872)^2} = 1.1173$

$w_1$ wins: $w_1(10) = w_1(9) + 0.3\times[0.4213\ -0.3892\ -0.0063\ -0.4612] = [0.7051\ 0.2724\ 0.0044\ 0.3229]$

10th Update:

$\|[0\ 0\ 0\ 1]-[0.7051\ 0.2724\ 0.0044\ 0.3229]\| = \sqrt{(-0.7051)^2+(-0.2724)^2+(-0.0044)^2+0.6771^2} = 1.0148$

$\|[0\ 0\ 0\ 1]-[0.6693\ 0.0031\ 0.9945\ 0.3872]\| = \sqrt{(-0.6693)^2+(-0.0031)^2+(-0.9945)^2+0.6128^2} = 1.3463$

$w_1$ wins: $w_1(11) = w_1(10) + 0.3\times[-0.6693\ -0.0031\ -0.9945\ 0.6128] = [0.4935\ 0.1907\ 0.0031\ 0.5260]$

Ömer Yılmaz
2017721048

① b) Pattern [1 1 0 0] :

$\|[1\ 1\ 0\ 0]-[0.4935\ 0.1907\ 0.0031\ 0.5260]\| = \sqrt{0.5065^2+0.8093^2+(-0.0031)^2+(-0.5260)^2} = 1.09$

$\|[1\ 1\ 0\ 0]-[0.6693\ 0.0031\ 0.9945\ 0.3872]\| = \sqrt{0.3307^2+0.9969^2+(-0.9945)^2+(-0.3872)^2} = 1.4974$

It is clustered with the 1 output.

Pattern [0 0 1 1] :

$\|[0\ 0\ 1\ 1]-[0.4935\ 0.1907\ 0.0031\ 0.5260]\| = \sqrt{(-0.4935)^2+(-0.1907)^2+0.9969^2+0.4740^2} = 1.2241$

$\|[0\ 0\ 1\ 1]-[0.6693\ 0.0031\ 0.9945\ 0.3872]\| = \sqrt{(-0.6693)^2+(-0.0031)^2+0.0055^2+0.6128^2} = 0.9075$

It is clustered with the 2 output.

Pattern [1 0 1 0] :

$\|[1\ 0\ 1\ 0]-[0.4935\ 0.1907\ 0.0031\ 0.5260]\| = \sqrt{0.5065^2+(-0.1907)^2+0.9969^2+(-0.5260)^2} = 1.2504$

$\|[1\ 0\ 1\ 0]-[0.6693\ 0.0031\ 0.9945\ 0.3872]\| = \sqrt{0.3307^2+(-0.0031)^2+0.0055^2+(-0.3872)^2} = 0.5093$

It is clustered with the 2 output.

Pattern [1 0 0 0] :

$\|[1\ 0\ 0\ 0]-[0.4935\ 0.1907\ 0.0031\ 0.5260]\| = \sqrt{0.5065^2+(-0.1907)^2+(-0.0031)^2+(-0.5260)^2} = 0.7547$

$\|[1\ 0\ 0\ 0]-[0.6693\ 0.0031\ 0.9945\ 0.3872]\| = \sqrt{0.3307^2+(-0.0031)^2+(-0.9945)^2+(-0.3872)^2} = 1.1173$

It is clustered with the 1 output.

Pattern [0 0 0 1] :

$\|[0\ 0\ 0\ 1]-[0.4935\ 0.1907\ 0.0031\ 0.5260]\| = \sqrt{(-0.4935)^2+(-0.1907)^2+(-0.0031)^2+0.4740^2} = 0.7104$

$\|[0\ 0\ 0\ 1]-[0.6693\ 0.0031\ 0.9945\ 0.3872]\| = \sqrt{(-0.6693)^2+(-0.0031)^2+(-0.9945)^2+0.6128^2} = 1.3463$

It is clustered with the 1 output.

Ömer Yılmaz
2017721048

① c) $[1\ 1\ 1\ 0]$ input:

$$\|[1\ 1\ 0\ 0]-[0.4935\ 0.1907\ 0.0031\ 0.5260]\| = \sqrt{0.5065^2+0.8093^2+0.9969^2+(-0.5260)^2} = 1.4772$$

$$\|[1\ 1\ 0\ 0]-[0.6693\ 0.0031\ 0.9945\ 0.3872]\| = \sqrt{(-0.6693)^2+(-0.0031)^2+(-0.9945)^2+0.6128^2} = 1.3463$$

It is clustered with the 2 output.

d)

$$w_{12}=w_{21}=\|[1\ 1\ 0\ 0]-[0\ 0\ 1\ 1]\| = \sqrt{1^2+1^2+(-1)^2+(-1)^2} = 2$$

$$w_{13}=w_{31}=\|[1\ 1\ 0\ 0]-[1\ 0\ 1\ 0]\| = \sqrt{0^2+1^2+(-1)^2+0^2} = \sqrt{2}$$

$$w_{14}=w_{41}=\|[1\ 1\ 0\ 0]-[1\ 0\ 0\ 0]\| = \sqrt{0^2+1^2+0^2+0^2} = 1$$

$$w_{15}=w_{51}=\|[1\ 1\ 0\ 0]-[0\ 0\ 0\ 1]\| = \sqrt{1^2+1^2+0^2+(-1)^2} = \sqrt{3}$$

$$w_{16}=w_{61}=\|[1\ 1\ 0\ 0]-[1\ 1\ 1\ 0]\| = \sqrt{0^2+0^2+(-1)^2+0^2} = 1$$

$$w_{23}=w_{32}=\|[0\ 0\ 1\ 1]-[1\ 0\ 1\ 0]\| = \sqrt{(-1)^2+0^2+0^2+1^2} = \sqrt{2}$$

$$w_{24}=w_{42}=\|[0\ 0\ 1\ 1]-[1\ 0\ 0\ 0]\| = \sqrt{(-1)^2+0^2+1^2+1^2} = \sqrt{3}$$

$$w_{25}=w_{52}=\|[0\ 0\ 1\ 1]-[0\ 0\ 0\ 1]\| = \sqrt{0^2+0^2+1^2+0^2} = 1$$

$$w_{26}=w_{62}=\|[0\ 0\ 1\ 1]-[1\ 1\ 1\ 0]\| = \sqrt{(-1)^2+(-1)^2+0^2+1^2} = \sqrt{3}$$

$$w_{34}=w_{43}=\|[1\ 0\ 1\ 0]-[1\ 0\ 0\ 0]\| = \sqrt{0^2+0^2+1^2+0^2} = 1$$

$$w_{35}=w_{53}=\|[1\ 0\ 1\ 0]-[0\ 0\ 0\ 1]\| = \sqrt{1^2+0^2+1^2+(-1)^2} = \sqrt{3}$$

$$w_{36}=w_{63}=\|[1\ 0\ 1\ 0]-[1\ 1\ 1\ 0]\| = \sqrt{0^2+(-1)^2+0^2+0^2} = 1$$

Ömer Yılmaz
201772104B

① d)

$$w_{45} = w_{54} = \|[1\ 0\ 0\ 0] - [0\ 0\ 0\ 1]\| = \sqrt{1^2 + 0^2 + 0^2 + (-1)^2} = \sqrt{2}$$

$$w_{46} = w_{64} = \|[1\ 0\ 0\ 0] - [1\ 1\ 1\ 0]\| = \sqrt{0^2 + (-1)^2 + (-1)^2 + 0^2} = \sqrt{2}$$

$$w_{56} = w_{65} = \|[0\ 0\ 0\ 1] - [1\ 1\ 1\ 0]\| = \sqrt{(-1)^2 + (-1)^2 + (-1)^2 + 1^2} = 2$$

| | [1 1 0 0] | [0 0 1 1] | [1 0 1 0] | [1 0 0 0] | [0 0 0 1] | [1 1 1 0] |
|---|---|---|---|---|---|---|
| [1 1 0 0] | 0 | 2 | $\sqrt{2}$ | 1 | $\sqrt{3}$ | 1 |
| [0 0 1 1] | 2 | 0 | $\sqrt{2}$ | $\sqrt{3}$ | 1 | $\sqrt{3}$ |
| [1 0 1 0] | $\sqrt{2}$ | $\sqrt{2}$ | 0 | 1 | $\sqrt{3}$ | 1 |
| [1 0 0 0] | 1 | $\sqrt{3}$ | 1 | 0 | $\sqrt{2}$ | $\sqrt{2}$ |
| [0 0 0 1] | $\sqrt{3}$ | 1 | $\sqrt{3}$ | $\sqrt{2}$ | 0 | 2 |
| [1 1 1 0] | 1 | $\sqrt{3}$ | 1 | $\sqrt{2}$ | 2 | 0 |

Ömer Yılmaz
2017721043

② a)

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

Letting main diagonal be consisting of all zeros:

$$H_1 = \begin{bmatrix} 0 & -1 & 1 & -1 & -1 \\ -1 & 0 & -1 & 1 & 1 \\ 1 & -1 & 0 & -1 & -1 \\ -1 & 1 & -1 & 0 & 1 \\ -1 & 1 & -1 & 1 & 0 \end{bmatrix}$$

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

$$\Rightarrow H_2 = \begin{bmatrix} 0 & -1 & -1 & 1 & -1 \\ -1 & 0 & 1 & -1 & 1 \\ -1 & 1 & 0 & -1 & 1 \\ 1 & -1 & -1 & 0 & -1 \\ -1 & 1 & 1 & -1 & 0 \end{bmatrix}$$

Then the weight matrix is just the addition of $H_1$ & $H_2$

$$H = H_1 + H_2 = \begin{bmatrix} 0 & -2 & 0 & 0 & -2 \\ -2 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & -2 & 0 \\ 0 & 0 & -2 & 0 & 0 \\ -2 & 2 & 0 & 0 & 0 \end{bmatrix}$$

Ömer Yılmaz
2017721048

2 b) Threshold is 0.

✓ Update node 1: $[0\ -2\ 0\ 0\ -2] \cdot [1\ 0\ 1\ 0\ 1]^T = -2 \Rightarrow 0$

Update node 3: $[0\ 0\ 0\ -2\ 0] \cdot [0\ 0\ 1\ 0\ 1]^T = 0 \Rightarrow 1$

Update node 4: $[0\ 0\ -2\ 0\ 0] \cdot [0\ 0\ 1\ 0\ 1]^T = -2 \Rightarrow 0$

✓ Update node 2: $[-2\ 0\ 0\ 0\ 2] \cdot [0\ 0\ 1\ 0\ 1]^T = 2 \Rightarrow 1$

Update node 5: $[-2\ 2\ 0\ 0\ 0] \cdot [0\ 1\ 1\ 0\ 1]^T = 2 \Rightarrow 1$

Update node 1: $[0\ -2\ 0\ 0\ -2] \cdot [0\ 1\ 1\ 0\ 1]^T = -4 \Rightarrow 0$

Update node 3: $[0\ 0\ 0\ -2\ 0] \cdot [0\ 1\ 1\ 0\ 1]^T = 0 \Rightarrow 1$

Update node 4: $[0\ 0\ -2\ 0\ 0] \cdot [0\ 1\ 1\ 0\ 1]^T = -2 \Rightarrow 0$

Update node 2: $[-2\ 0\ 0\ 0\ 2] \cdot [0\ 1\ 1\ 0\ 1]^T = 2 \Rightarrow 1$

Update node 5: $[-2\ 2\ 0\ 0\ 0] \cdot [0\ 1\ 1\ 0\ 1]^T = 2 \Rightarrow 1$

We have updated each node in the net without changing them. Therefore, we converged.

$[0\ 1\ 1\ 0\ 1]$, After 4 updates

Ömer Yılmaz
2017721048

③ a) $n$ is the number of inputs.

$\Rightarrow 2^n$ is the number of different input combinations.

For $n=3$, $2^3 = 8$ different possible input scenarios exist.

b) $y = \varphi(w_1 x_1 + w_2 x_2 + w_3 x_3 + b)$

For $P_1$: $y = \varphi(2.1 - 4.0 + 1.0 + 0) = \varphi(2) = 1$

For $P_2$: $y = \varphi(2.0 - 4.1 + 1.1 + 0) = \varphi(-3) = 0$

For $P_3$: $y = \varphi(2.1 - 4.0 + 1.1 + 0) = \varphi(3) = 1$

For $P_4$: $y = \varphi(2.1 - 4.1 + 1.1 + 0) = \varphi(-1) = 0$

c) AND function should give 1 output only if all three inputs are 1. That is also means only if all three inputs are 1, then $v$ (the argument of $\varphi$) should be greater than or equal to 2.

$$y = \varphi(1.x_1 + 1.x_2 + w_3.x_3 + 2)$$

$\Rightarrow 1.1 + 1.1 + w_3 + 2 \geq 2 \Rightarrow w_3 \geq -2$

$1.1 + 1.1 + 0 + 2 < 2$ Impossible with these parameters

AND function do not work properly.

d) I would change it in a way that only input $\{0, 0, 0\}$ would give 0 output. If $w_1 = 1$, $w_2 = 1$ and $b = 2$:

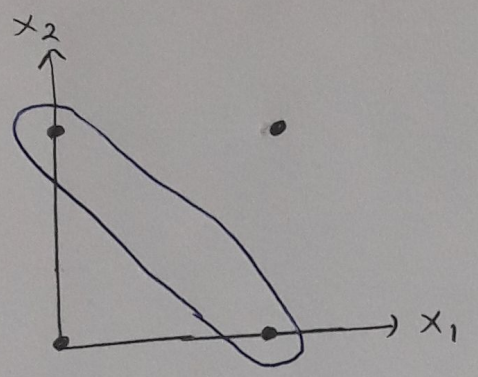
$1.0 + 1.0 + w_3.0 + 2 < 2$ This cannot be achieved without decreasing bias or increasing threshold ($v_{th} > 2$).

$1.0 + 1.0 + w_3.1 + 2 \geq v_{th} \Rightarrow w_3 + 2 \geq v_{th}$

If we choose $w_3$ to be 1, $v_{th} = 2.5$ would suffice

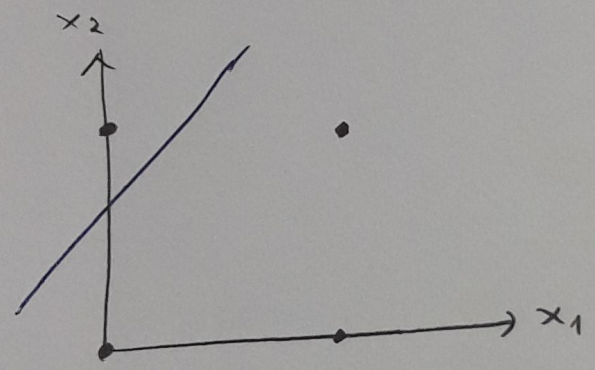
Ömer Yılmaz
2017721048

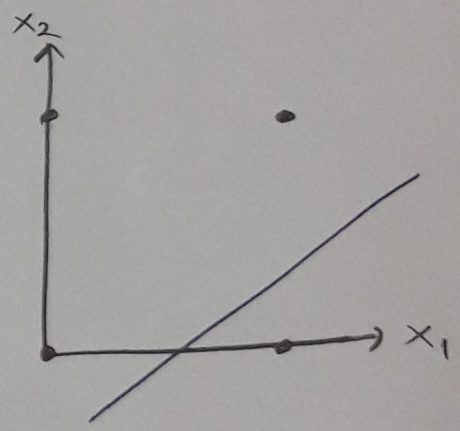
③ e)




⇒ This subset cannot be isolated with a single decision boundary.

⇒ It is impossible to implement XOR with a single neuron.




These two neurons can be ORed to give an XOR implementation. Similarly, one minterm of three or more dimensional inputs can be isolated always. The resulting layer of neurons can then be ORed.

Ömer Yılmaz
2017721048

4 a) For P1:

3 : $\varphi((-2).0+3.0) = \varphi(0) = 1$

4 : $\varphi(4.0+(-1).0) = \varphi(0) = 1$

5 : $\varphi(1.1+(-1).1) = \varphi(0) = 1$

6 : $\varphi((-1).1+1.1) = \varphi(0) = 1$

For P2:

3 : $\varphi((-2).1+3.0) = \varphi(-2) = 0$

4 : $\varphi(4.1+(-1).0) = \varphi(4) = 1$

5 : $\varphi(1.0+(-1).1) = \varphi(-1) = 0$

6 : $\varphi((-1).0+1.1) = \varphi(1) = 1$

For P3:

3 : $\varphi((-2).0+3.1) = \varphi(3) = 1$

4 : $\varphi(4.0+(-1).1) = \varphi(-1) = 0$

5 : $\varphi(1.1+(-1).0) = \varphi(1) = 1$

6 : $\varphi((-1).1+1.0) = \varphi(-1) = 0$

For P4:

3 : $\varphi((-2).1+3.1) = \varphi(1) = 1$

4 : $\varphi(4.1+(-1).1) = \varphi(3) = 1$

5 : $\varphi(1.1+(-1).1) = \varphi(0) = 1$

6 : $\varphi((-1).1+1.1) = \varphi(0) = 1$

Ömer Yılmaz
2017721048

④ b)

Training set is the input that is fed to a machine learning algorithm that tries to find patterns on it. Training set samples can be labeled or not. Labeled data can be used under supervision of the machine learning algorithm which constantly tries to decrease the errors of models prediction capability (this can be achieved by different methods like GD, CG, NM, QN, LM). Unlabeled training set can be divided into some clusters (unsupervised learning). Also, input values can be associated with other samples.

As an example, we can have a model which predicts outputs given input values. The real value of the outputs and the predictions can be compared. This is generally done with a cost function like average sum of squared errors. When we descript the cost function in terms of weigth parameters, we can then search for the weigth parameters that minimizes the cost. This may be achieved with a gradient descent algoritm. First, we analytically calculate gradient of the cost function then scale it with a learning rate and use it to modify the weigth parameters.

Epoch is a single pass through whole training set. This passing can be one-shot, one by one or passing through batches.