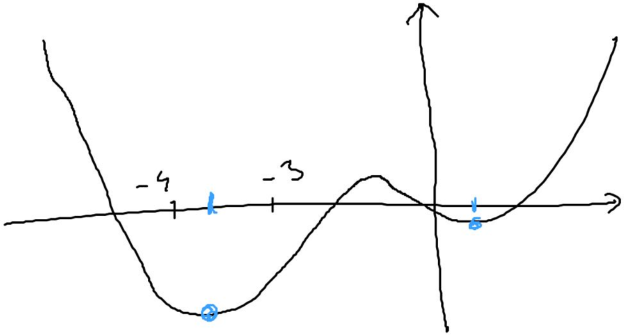
3)

$$f(x) = x^4 + 5x^3 - 7x$$

$$f'(x) = 4x^3 + 5 \cdot 3x^2 - 7$$

$$= 4x^3 + 15x^2 - 7$$



$$x_{n+1} = x_n - \alpha f'(x_n)$$

$$\alpha = 0{,}01$$

$$x_0 = -3{,}5$$

$$x_1 = -3{,}5 - 0{,}01 \cdot \left(4 \cdot (-3{,}5)^3 + 15(-3{,}5)^2 - 7\right)$$

$$\approx -3{,}5525$$

$$x_2 = -3{,}5525 - 0{,}01 \cdot \left(4 \cdot (-3{,}5525)^3 + 15 \cdot (-3{,}5525)^2 - 7\right)$$

$$= -3{,}5822$$

4) $f(x,y) = 3x^2 + 3xy + 2y^2 + 3x - 5y$

$$\frac{\partial}{\partial x} f(x,y) = 3 \cdot 2x + 3y + 3$$
$$= 6x + 3y + 3$$

$$\frac{\partial}{\partial y} f(x,y) = 3x + 2 \cdot 2y - 5$$
$$= 3x + 4y - 5$$

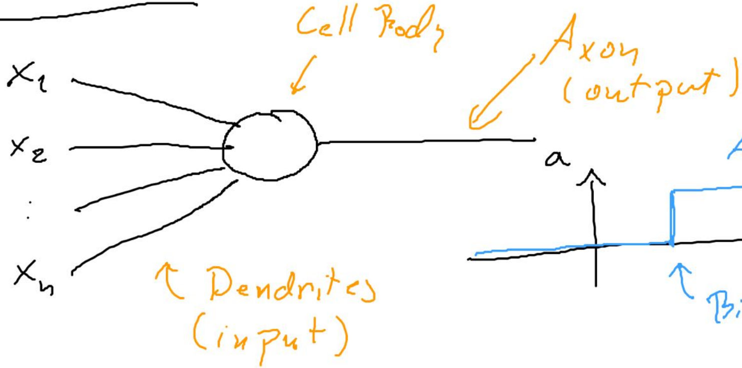
$$\nabla f = \begin{bmatrix} 6x + 3y + 3 \\ 3x + 4y - 5 \end{bmatrix} \qquad X_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix}$$

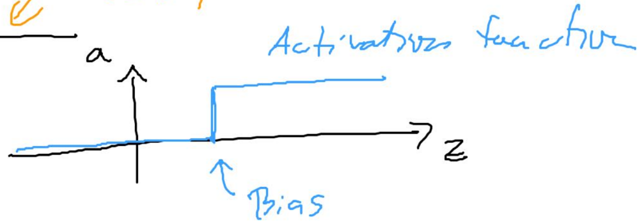
$$\boxed{X_{i+1} = X_i - \alpha \nabla f(X_i)}$$

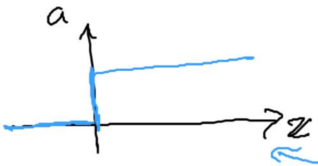
$$
\begin{cases}
x_{i+1} = x_i - \alpha \dfrac{\partial}{\partial x} f(x, y) \Big|_{\substack{x = x_i \\ y = y_i}} \\[2em]
y_{i+1} = y_i - \alpha \dfrac{\partial}{\partial y} f(x, y) \Big|_{\substack{x = x_i \\ y = y_i}}
\end{cases}
$$

Neural Networks

Simplified Neuron



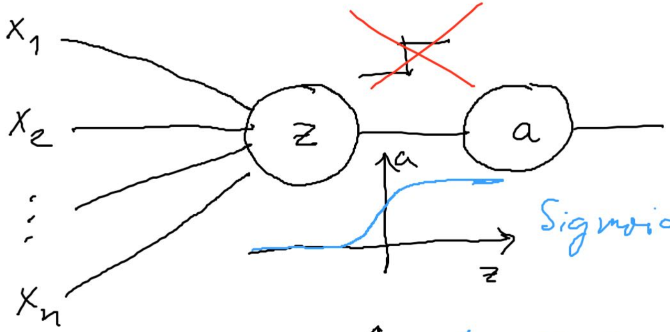




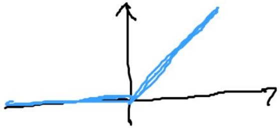
$$z = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n$$

$$z = w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_n x$$

Bias

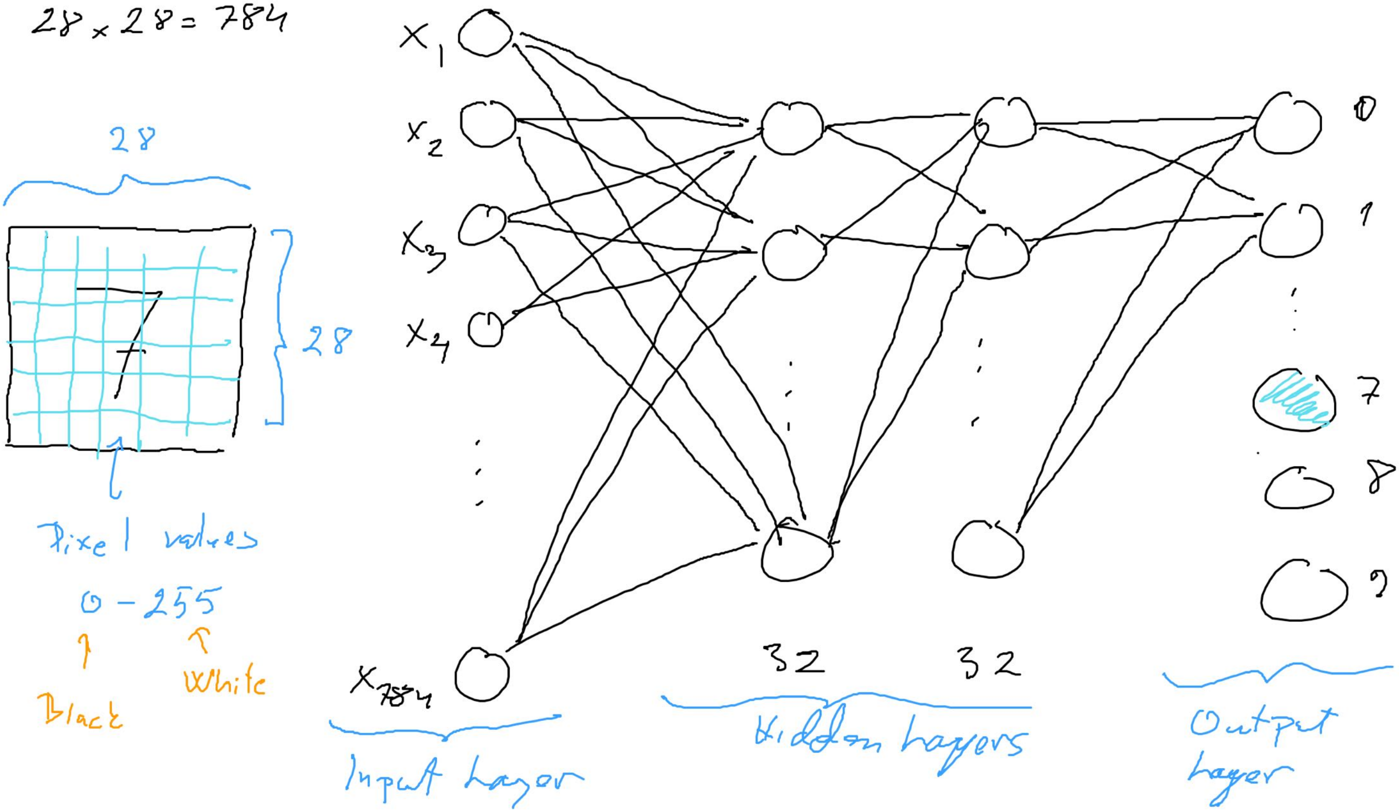


Sigmoid $\quad g(x) = \dfrac{1}{1 - e^{-z}}$



ReLU $\quad g(x) = \begin{cases} 0 & , x < 0 \\ x & , x \geq 0 \end{cases}$

$$\boxed{\begin{aligned} z &= \Theta_0 + \Theta_1 x_1 + \Theta_2 x_2 + \cdots + \Theta_n x_n \\ a &= g(z) = \frac{1}{1 - e^{-x}} \end{aligned}}$$

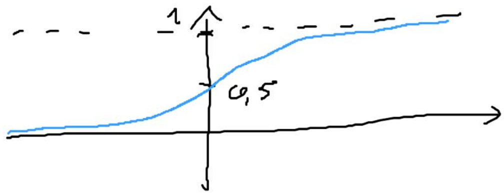

28 x 28 = 784
28
28
7
Pixel values
0 - 255
Black
White
x_1
x_2
x_3
x_4
x_{784}
Input layer
32
32
Hidden layers
0
1
7
8
9
Output layer

Logistic Regression with Gradient Descent

$$\hat{y} = g(\Theta_0 + \Theta_1 x)$$

where

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



$$\hat{y} = \frac{1}{1 + e^{-(\Theta_0 + \Theta_1 x)}}$$

Cost functions

$\log(x) = \ln(x)$

$$\cancel{S = \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)}\right)^2}$$

$$J(\theta_0, \theta_1) = \sum_{i=1}^{m} -\left[y^{(i)}\log(\hat{y}^{(i)}) + (1-y^{(i)})\log(1-\hat{y}^{(i)})\right]$$

$$\boxed{J(\theta_0, \theta_1) = -\left[y\log(\hat{y}) + (1-y)\log(1-\hat{y})\right]}$$

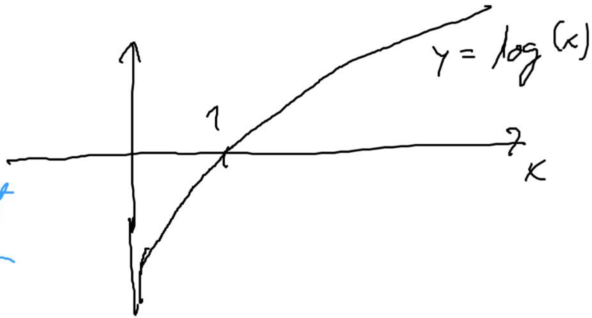
$y = 1$ and $\hat{y} \approx 1$

$$J = -\left[1 \cdot 0 + 0 \cdot (-\infty)\right] = 0$$

$y = 0$ and $\hat{y} \approx 0$

$$J = -\left[0(-\infty) + 1 \cdot 0\right] \approx 0$$

Correct answer



$$J = -\left[y \log \hat{y} + (1-y)\log(1-\hat{y})\right]$$

$y = 1$ and $\hat{y} = 0$

$$J = -\left[1 \cdot (-\infty) + 0 \cdot 0\right] \approx \infty$$

$y = 0$ and $\hat{y} \approx 1$

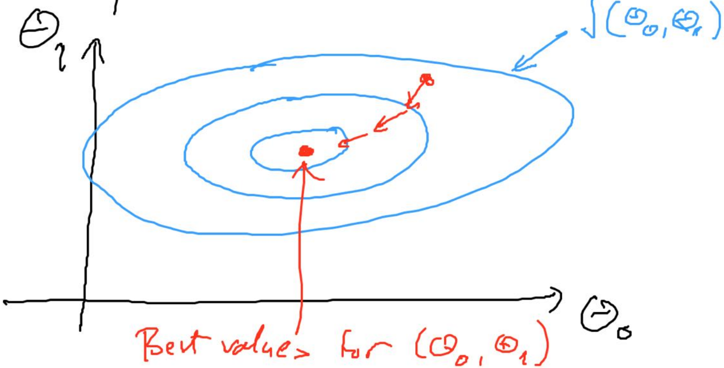
$$J = -\left[0 \cdot 0 + 1 \cdot (-\infty)\right] \approx \infty$$

Wrong answer

We aim to calculate the partial derivatives

$$\frac{\partial}{\partial \Theta_0} J(\Theta_0, \Theta_1)$$

$$\frac{\partial}{\partial \Theta_1} J(\Theta_0, \Theta_1)$$



First we calculate the derivative of

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

$$D\left(\frac{1}{1+e^{-x}}\right) = D\left(1+e^{-x}\right)^{-1}$$

$$= (-1)\left(1+e^{-x}\right)^{-2} \cdot D\left(1+e^{-x}\right)$$

$$= \cancel{(-1)}\left(1+e^{-x}\right)^{-2} \cdot \left(e^{-x} \cdot \cancel{(-1)}\right) = \frac{e^{-x}}{\left(1+e^{-x}\right)^{2}}$$

$$= \frac{1+e^{-x}-1}{\left(1+e^{-x}\right)^{2}} = \frac{1+e^{-x}}{\left(1+e^{-x}\right)^{2}} - \frac{1}{\left(1+e^{-x}\right)^{2}}$$

$$D\left(\frac{1}{1+e^{-x}}\right) = \frac{1}{1+e^{-x}} - \frac{1}{(1+e^{-x})^2}$$

$$= g(x) - [g(x)]^2$$

$$= g(x)\left(1 - g(x)\right)$$

$$\Rightarrow \boxed{D\,g(x) = g(x)\left(1 - g(x)\right)}$$

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

$$D \log(x) = \frac{1}{x}$$

The cost function is

$$J(\theta_0, \theta_1) = -\left[y \log(\hat{y}) + (1-y)\log(1-\hat{y}) \right]$$

$$= -\left[y \log\left(g(\theta_0 + \theta_1 x)\right) + (1-y)\log\left(1 - g(\theta_0 + \theta_1 x)\right) \right]$$

$$\frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) = -\left[\frac{y}{\cancel{g(\theta_0 + \theta_1 x)}} \cancel{g(\theta_0 + \theta_1 x)} \left(1 - g(\theta_0 + \theta_1 x)\right) \cdot 1 \right.$$

$$\left. + \frac{1-y}{\cancel{1 - g(\theta_0 + \theta_1 x)}} \left(- g(\theta_0 + \theta_1 x) \cancel{\left(1 - g(\theta_0 + \theta_1 x)\right)} \right) \cdot 1 \right]$$

$$= -\left[y\left(1 - g(\theta_0 + \theta_1 x)\right) - (1-y)\, g(\theta_0 + \theta_1 x) \right]$$

$$= -\left[y - \cancel{y g(\theta_0 + \theta_1 x)} - g(\theta_0 + \theta_1 x) + \cancel{y g(\theta_0 + \theta_1 x)} \right]$$

$$= -\left[y - g(\theta_0 + \theta_1 x) \right]$$

$$= g(\theta_0 + \theta_1 x) - y \quad = \hat{y} - y$$

$$\Rightarrow \boxed{\frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) = \hat{y} - y}$$

$$J(\theta_0, \theta_1) = -\left[y \log(g(\theta_0 + \theta_1 x)) + (1-y)\log(1 - g(\theta_0 + \theta_1 x))\right]$$

Same way as before, we get

$$\frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1) = (\hat{y} - y)x$$

We have

$$\frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) = \hat{y} - y$$

$$\frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1) = (\hat{y} - y)x$$

$$\Rightarrow \quad \nabla J(\theta_0, \theta_1) = \begin{bmatrix} \hat{y} - y \\ (\hat{y} - y)x \end{bmatrix}$$

$$M = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}$$

$$y = \begin{bmatrix} y^{(0)} \\ y^{(1)} \\ \vdots \\ y^{(m-1)} \end{bmatrix} \qquad X = \begin{bmatrix} 1 & x^{(0)} \\ 1 & x^{(1)} \\ \vdots & \vdots \\ 1 & x^{(m)} \end{bmatrix}$$

$$\hat{y} = \begin{bmatrix} \hat{y}^{(0)} \\ \hat{y}^{(1)} \\ \vdots \\ \hat{y}^{(m-1)} \end{bmatrix} = \begin{bmatrix} g(\theta_0 + \theta_1 x^{(0)}) \\ g(\theta_0 + \theta_1 x^{(1)}) \\ \vdots \\ g(\theta_0 + \theta_1 x^{(m-1)}) \end{bmatrix}$$

$$= g(XM)$$

We have

$$\nabla J(\theta_0, \theta_1)$$

$$= X^T(\hat{y} - y)$$

$$= X^T(g(Xu) - y)$$

Gradient Descent for the Logistic Regression

$$M \leftarrow M - \alpha \nabla J(\theta_0, \theta_1)$$

$$\boxed{M \leftarrow M - \alpha X^T(g(XM) - y)}$$