AGENDA.

1) Chain Rule.
2) Backpropagation derivation.
3) TensorFlow.

Chain rule

Example.

$$f(x) = \frac{1}{\underbrace{1+e^{-x}}} = \sigma(x) \qquad \frac{df}{dx} = ?$$

Let $\underbrace{p = -x}_{}$ ($p(x)$), $1+e^{-x} = \underbrace{1+e^{p}}_{q(p)} = q$

$$\frac{1}{1+e^{-x}} = \frac{1}{1+e^{p}} = \frac{1}{q} = f \quad f(q)$$

$$f(q) = \frac{1}{q}$$

$$\frac{df}{dq} = \frac{dq^{-1}}{dq} = -q^{-2} = -\frac{1}{q^2} \quad \text{— (1)}$$

$$\underbrace{q(p) = 1+e^{p}}$$

$$\frac{dq}{dp} = 0 + e^{p} = e^{p} \quad \text{— (2)}$$

$$p(x) = -x$$

$$\frac{dp}{dx} = -1 = \longrightarrow (3)$$

$$\frac{df}{dx} = \frac{df}{dq} \cdot \frac{dq}{dp} \cdot \frac{dp}{dx} = \left(-\frac{1}{q^2}\right) \cdot \left(e^{p}\right) \cdot (-1)$$

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

replaced the value of q & p

$$\frac{df}{dx} = \frac{e^{-x}}{(1+e^{-x})^2}$$

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

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

$$\frac{\cancel{A}}{A^2} - \frac{1}{A^2}$$

$$\frac{df}{dx} = \frac{1}{(1+e^{-x})} - \frac{1}{(1+e^{-x})^2}$$

$$\frac{1}{A} - \frac{1}{A^2}$$

$$B - B^2$$

$\sigma(x)$

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

$$\underbrace{\frac{1}{A}\left[1 - \frac{1}{A}\right]}$$

$$\boxed{\begin{aligned} \frac{df}{dx} &= \underbrace{\sigma(x)}\left(1 - \underbrace{\sigma(x)}\right) \\ \sigma'(x) &= \sigma(x)\left(1 - \sigma(x)\right) \end{aligned}}$$

$$\underline{f(x)(1 - f(x))}$$

$$\boxed{f(x) \longrightarrow \underbrace{f(q(p(x)))} \Rightarrow \frac{df}{dx} = \underline{\frac{df}{dq} \cdot \frac{dq}{dp} \cdot \frac{dp}{dx}}}$$

Backpropagation

$x \xrightarrow{\;w,b\;} \boxed{NN} \xrightarrow{\;FP\;} \underbrace{\hat{y} - \overset{t}{y}}_{err}$

$\xleftarrow{\;bp\;}$

Assumption,

i) Each layer has only one neuron

ii) bias = 0

$x_1 \longrightarrow \bigcirc \xrightarrow{\;w_1\;} (1) \xrightarrow{\;w_2\;} (2) \longrightarrow$

$a_0 = x_1$ buffer layer

$z_1 = w_1 a_0$

$a_1 = \sigma(z_1)$

$z_2 = w_2 a_1$

$a_2 = \sigma(z_2)$

$\hat{y} = a_2$

y : actual value

$e = (y - \hat{y})^2$

Weight update rule

$$w = w - \eta \nabla c = w - \eta \frac{\partial e}{\partial w}$$

Observation,

$$e = (y - \hat{y})^2 = (y - a_2)^2 \;, \quad a_2 = \sigma(z_2) \qquad z_2 = w_2 a_1$$

$f(a_2)$ $\quad f(z_2)$ $\quad f(w_2, a_1)$

$$e \to f\left(a_2\left(z_2\left(w_2, a_1\right)\right)\right)$$

$$\frac{\partial e}{\partial w_2} = \frac{\partial e}{\partial a_2} \cdot \frac{\partial a_2}{\partial z_2} \cdot \frac{\partial z_2}{\partial w_2} = -2(y - a_2) \cdot \sigma(z_2)\left[1 - \sigma(z_2)\right] a_1$$

Wt update $\Rightarrow$ $w_2 = w_2 - \eta \frac{\partial e}{\partial w_2}$

$f = (y - a_2)^2$

$f = y^2 + a_2^2 - 2 y a_2$

$\frac{\partial f}{\partial a_2} = 0 + 2a_2 - 2y$

$= -2(y - a_2)$

Observation,

$$\Rightarrow e = (y-\hat{y})^2 = (y - a_2)^2, \quad a_2 = \sigma(z_2), \quad z_2 = w_2 a_1$$

$$f(a_2) \qquad f(z_2) \qquad f(w_2, a_1)$$

$$a_1 = \sigma(z_1) \qquad z_1 = w_1 a_0$$

$$f(z_1) \qquad f(w_1, a_0)$$

$$\frac{\partial e}{\partial w_1} = \frac{\partial e}{\partial a_2} \cdot \frac{\partial a_2}{\partial z_2} \cdot \frac{\partial z_2}{\partial a_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_1}$$

$$-2(y-a_2) \cdot \sigma(z_2)\left[1-\sigma(z_2)\right] \cdot w_2 \cdot \sigma(z_1)\left[1-\sigma(z_1)\right] \cdot a_0$$

$$w_1 = w_1 - \eta \frac{\partial e}{\partial w_1}$$ } slow optimizer

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} - \eta \underbrace{\begin{bmatrix} \partial e/\partial w_1 \\ \partial e/\partial w_2 \end{bmatrix}}$$

$$\boxed{W = W - \eta \nabla e}$$

$$f(x,y) = \sqrt{x^2+y^2} = (x^2+y^2)^{1/2}$$

$$\nabla f(x,y) = \frac{\partial f}{\partial x}\hat{i} + \frac{\partial f}{\partial y}\hat{j}$$

$$\frac{\partial f}{\partial x} = (x^2+y^2)^{1/4}$$

$$x^n = n x^{n-1}$$

$$= \frac{1}{\cancel{2}}(x^2+y^2)^{1/2-1} \cdot (\cancel{2}x+0)$$

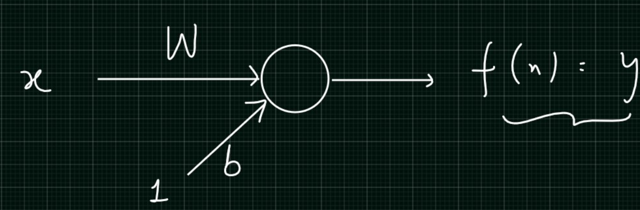
$$\frac{\partial f}{\partial x} = \frac{x}{\sqrt{x^2+y^2}}$$

$$\frac{\partial f}{\partial y} = \frac{y}{\sqrt{x^2+y^2}}$$

$$\nabla f(x,y) = \frac{x}{\sqrt{x^2+y^2}}\hat{i} + \frac{y}{\sqrt{x^2+y^2}}\hat{j}$$

$$\nabla f(3,2) = \frac{3}{\sqrt{9+4}}\hat{i} + \frac{2}{\sqrt{9+4}}\hat{j} =$$

$$f(x) = \underbrace{Wx+b}$$



$$\left\{\begin{matrix} W = 3 \\ b = 2 \end{matrix}\right\} \Rightarrow$$

Regression Example

∴ No actⁿ function used

True Value

$$y = \underbrace{\underset{W}{3}x + \underset{b}{2}} \leftarrow \text{True}$$

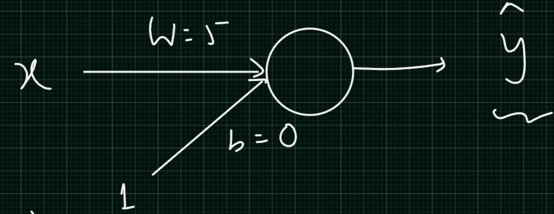
Data Generation

$$y = 3x + 2 + \text{noise}$$

Model

$$W = 5 \qquad b = 0$$

Train step :-



$$e = \text{error} = \frac{1}{m}\sum (y - \hat{y})^2 \qquad m = \text{no of samples}$$

error

$$\frac{\partial e}{\partial w} \qquad \frac{\partial e}{\partial b}$$

$$W = \underbrace{W - \eta \frac{\partial e}{\partial w}} \quad , \quad b = b - \eta \frac{\partial e}{\partial b}$$

W. ← assign_sub($\eta \times$)

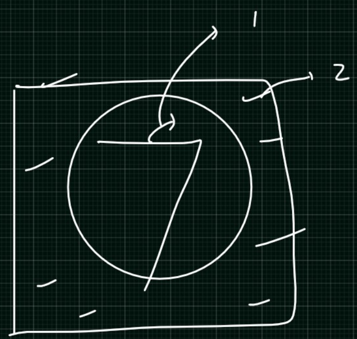
$$x = [\quad]$$

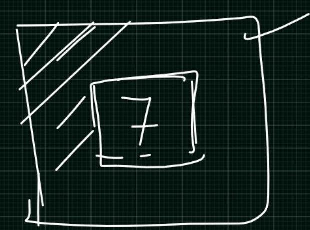
$$\hat{y} = W \cdot x + b$$

$$\hat{y} = 3 \times \begin{bmatrix} \\ \\ \end{bmatrix} + 2 \quad \{\text{Broadcast}\}$$

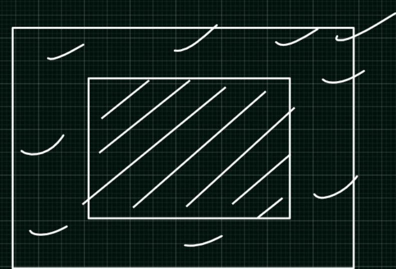
$$\hat{y} = \begin{bmatrix} 3 \\ 3 \\ \vdots \end{bmatrix} + 2$$

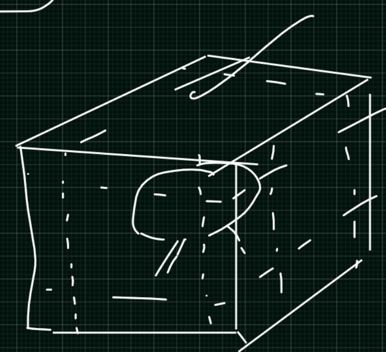
$$\hat{y} = \begin{pmatrix} 3\dots + 2 \\ 3\dots + 2 \\ \\ \end{pmatrix}_{1000\times 1} \qquad y = \begin{pmatrix} \\ \\ \end{pmatrix}_{1000\times 1}$$





Am





$$\underbrace{\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_{100} \end{bmatrix}}_{\text{new}} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_{100} \end{bmatrix} - \eta \underset{\sim}{\nabla C}$$

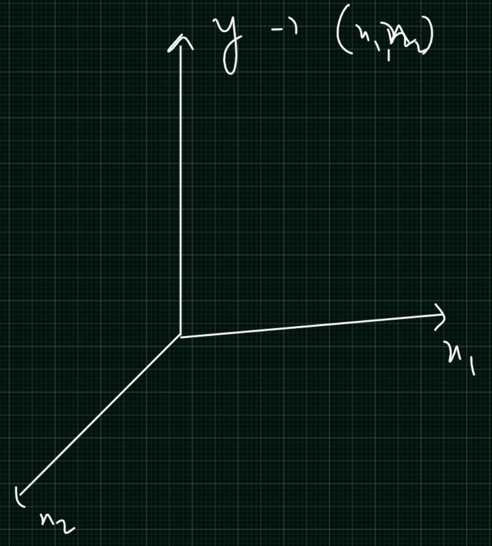
$$\Downarrow$$

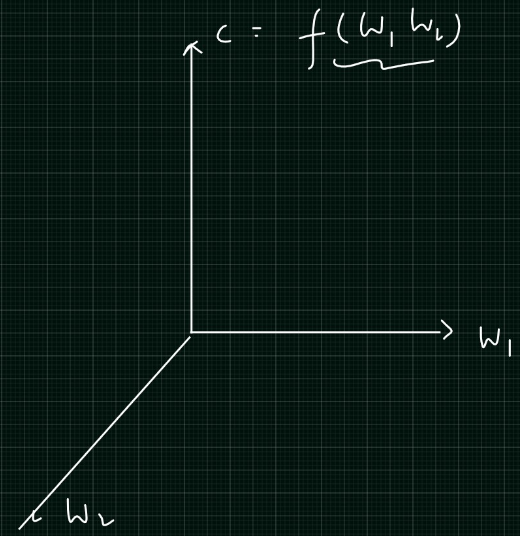
$$\begin{bmatrix} \frac{\partial C}{\partial w_1} \\ \frac{\partial C}{\partial w_2} \\ \vdots \\ \frac{\partial C}{\partial w_{100}} \end{bmatrix}$$

$$C(w_1, w_2 \dots\ b_1, \dots b_2)$$

$$\nabla C = \underbrace{\frac{\partial C}{\partial w_1}\hat{i} + \frac{\partial C}{\partial w_2}\hat{j}}_{} + \dots \frac{\partial C}{\partial b_1} + \frac{\partial C}{\partial b_2} + \dots$$

$$\underset{\sim}{\Delta W} = \eta \begin{pmatrix} \partial C/\partial w_1 \\ \partial C/\partial w_2 \\ \vdots \end{pmatrix}$$





$x_1 \quad x_2 \quad y$

$w_1 \quad w_2 \quad C$

$$C = (y - \hat{y})^2$$

$$C = \big(\underbrace{y - \phi(w_1 x_1 + w_2 x_2)}\big)^2$$

$$f = w_1 x_1 + w_2 x_2 + w_3 x_3 + \text{bias}$$