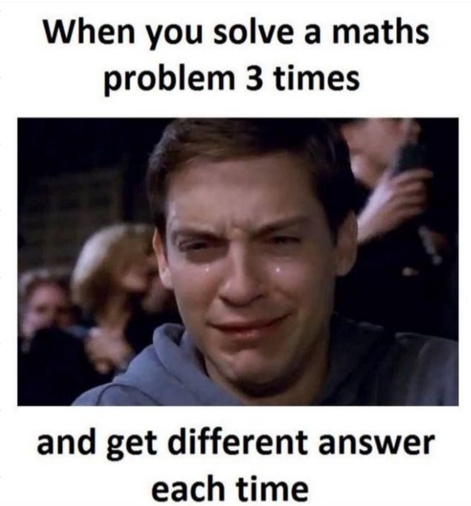
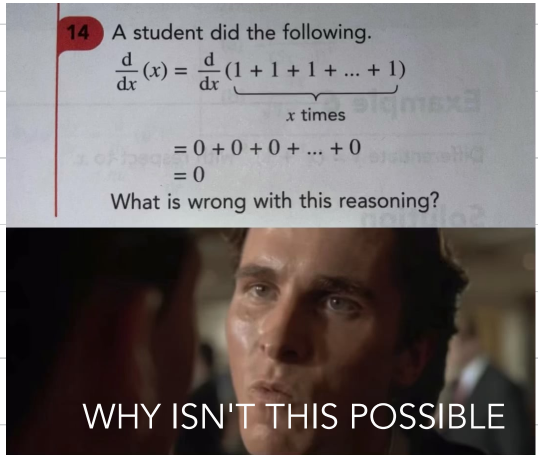
Session 7 Optimization-3 G.D. In Action Feb 13, 2024





Agenda

1. Understand where derivative Fits in ML
2. What are Gradients?
3. Linear Regression (Sort of :D)

① $f(x)\ 3x^4 - 5x^3 + 2x \rightarrow$ $f'(x),\ f''(x)$

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

$$f''(x) = 3\times 4\times 3x^2 - 5\times 3\times 2x$$

② $e^{2x}\cdot \cos(x)$ $f'(x),\ f''(x)$

$$f'(x)\quad e^{2x}\cdot \frac{d(\cos x)}{dx} + \frac{d(e^{2x})}{dx}\cdot \cos x$$

$$\Rightarrow e^{2x}\cdot(-\sin x) + \cos x\cdot e^{2x}\cdot 2$$

$$\Rightarrow 2\cdot e^{2x}\cdot \cos x - e^{2x}\cdot \sin x$$

$e^{2x} = f(g(x))$ $\rightarrow 2x$

$\rightarrow e^{x}$

$f'(g(x))\ g'(x)$

$$f'(x) = 2 \cdot e^{2x} \cdot \cos x - e^{2x} \cdot \sin x$$

$$f''(x) = 4 \cdot e^{2x} \cdot \cos x - 2e^{2x} \cdot \sin x$$

$$+$$

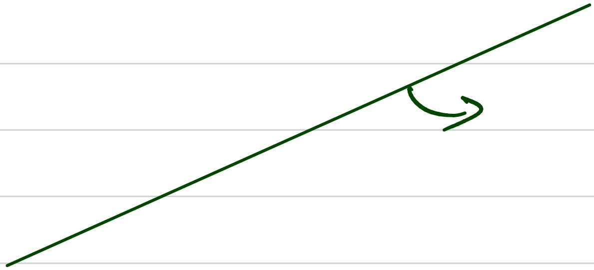
$$e^{2x} \cdot \sin x \Rightarrow -e^{2x} \cdot \cos x - \sin x \cdot 2 \cdot e^{2x}$$

$$\Rightarrow 3 e^{2x} \cdot \cos x - 3 \sin x \cdot e^{2x}$$

Why do we take derivative?

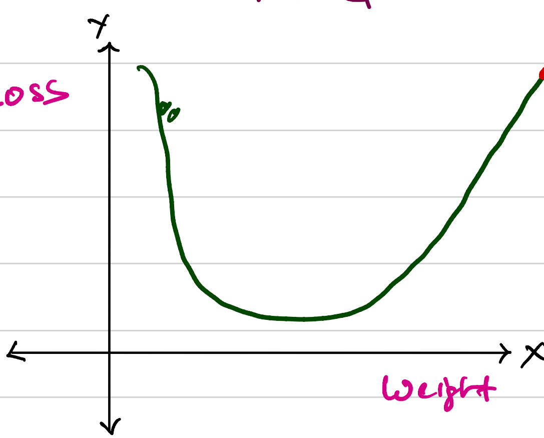
→ In my ML model → Reduce loss

$\vec{w}$

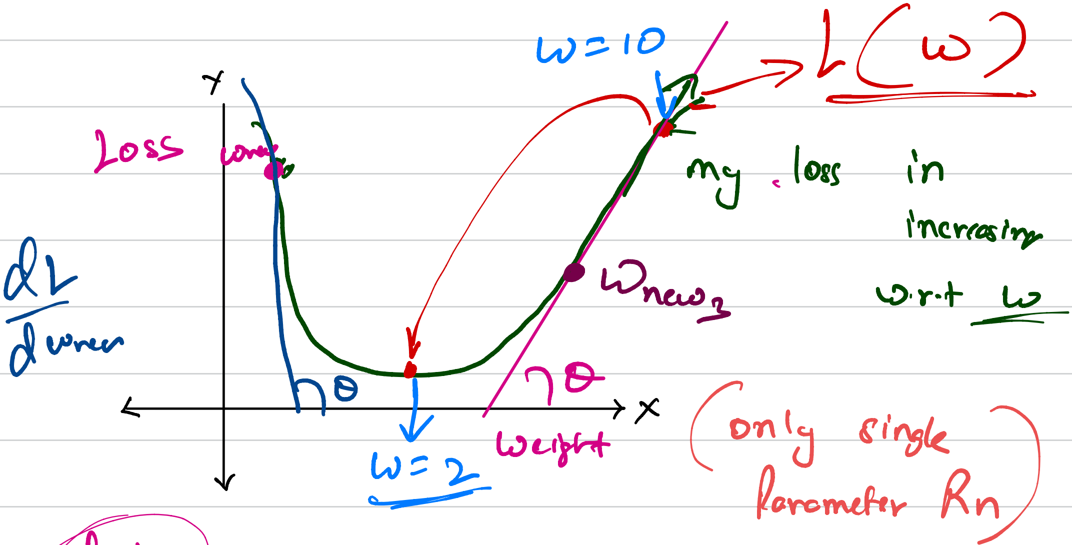


$$\max_{\vec{w},\, w_0} \sum_{i=1}^{n} y_i \times \frac{w^T x_i + w_0}{||w||}$$ gain function

- My ML model has only 1 parameter



(only single parameter Rn)



$\frac{dL}{dw}$

$$W_{new_1} = W_{old} - \frac{dL}{dw}$$

Idea can in Gradient descent

$$W_{new_2} = W_{new_1} - \frac{dL}{dw_{new_1}}$$

$\eta \rightarrow$ Learning Rate

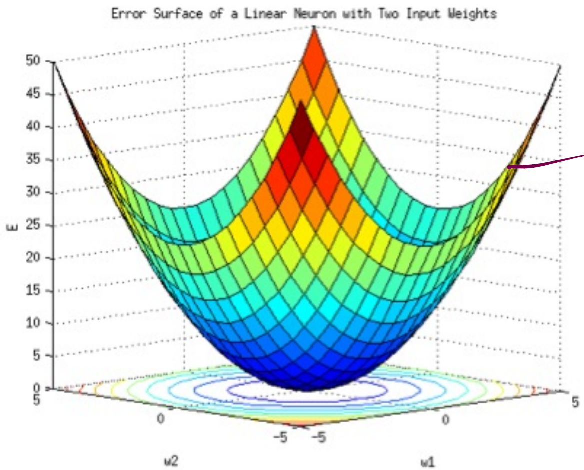
Eeta

10^{-3}

$$\boxed{W_{new} = W_{old} - \eta \times \frac{dL}{dw}}$$

GPT-3.5 → 135 Billion Para.

→ Multiple Parameters   Loss function ≈ Cost function



→ $L(w_1, w_2)$

$f(x, y)$ → multi variable func   1 output / >1 output

$= x^2 + y^2 + 3$

Domain: (4, 3), (1, 1)

Range: (20), (50)

(4, 3) → (20)

(1, 1) → (50)

Functions with multiple variable accepts vectors as input & returns a single value as output

PARTIAL DERIVATIVE

If my $L(w) \rightarrow \frac{dL}{dw}$) For updating my weights

If my $L(w_1, w_2) \rightarrow \frac{dL}{w_1, w_2}$??

You treat other variable as Cons

∂ $z = f(x_1, x_2)$

Rho

$\frac{\partial z}{\partial x_1} \Rightarrow f'(x_1, x_2)$ Constant

Ex. $f(x,y) = 2x^2y + 3y^3x^2 + 3y$

$$\frac{\partial f(x,y)}{\partial x} = 2y \cdot 2x + 3y^3 \cdot 2x + \cancel{\frac{d\,3y}{dx}} \rightarrow 0$$

$$\frac{\partial f(x,y)}{\partial y} = 2x^2 \cdot 1 + 3x^2 \cdot 3y^2 + 3\cancel{\frac{dy}{dy}} \rightarrow 1$$

Goal is

$$\min_{w_0, w_1, \ldots w_n} L = f(w_0, w_1, w_2 \ldots w_n)$$

minimum

$$\frac{\partial L}{\partial w_0}, \frac{\partial L}{w_1}, \frac{\partial L}{w_2} \ldots \frac{\partial L}{w_n}$$

Change in Loss w.r.t. $w_0, w_1, \ldots w_n$

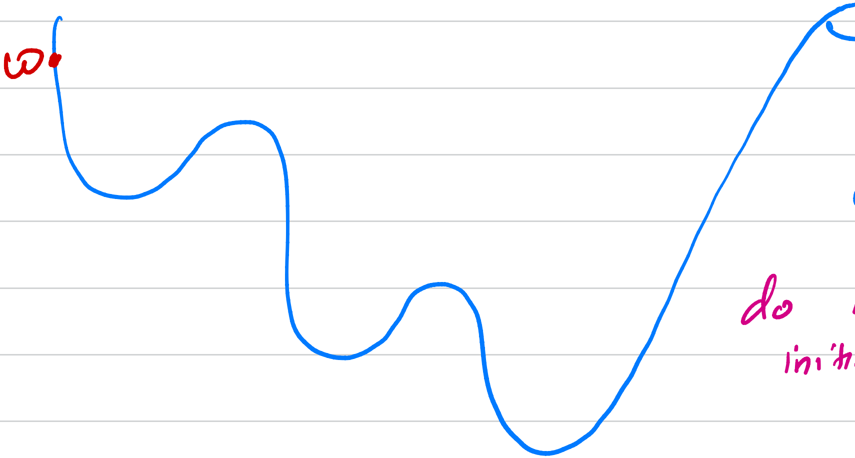
$\frac{\partial L}{\partial w_0}$ → Gradient

$$w = w - \eta \frac{\partial L}{\partial w}$$

→ Gradient

$$\nabla L = \begin{bmatrix} \frac{\partial L}{\partial w_0} \\ \frac{\partial L}{\partial w_1} \\ \vdots \\ \frac{\partial L}{\partial w_n} \end{bmatrix} \qquad w = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_n \end{bmatrix}$$

$$\boxed{w = w - \eta \times \nabla L}$$



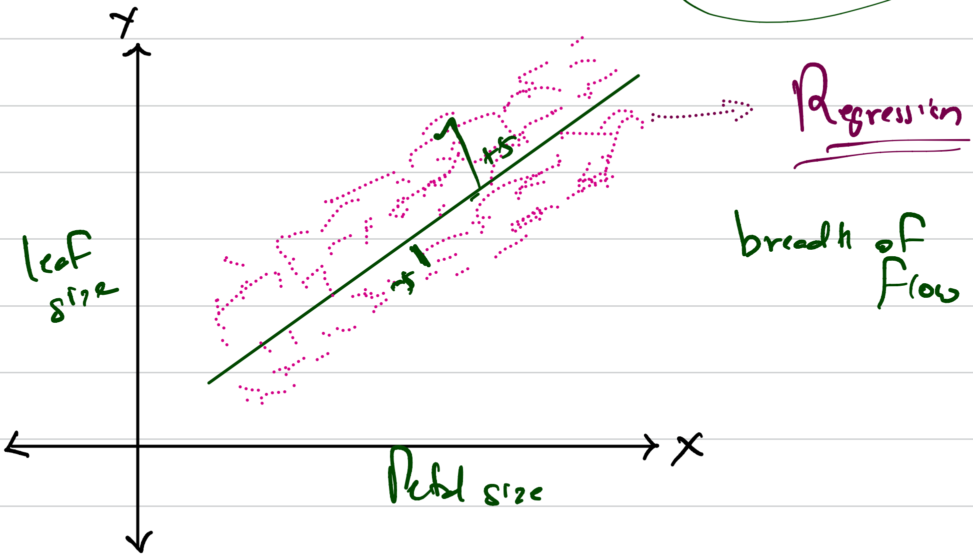
Loss func

w.r.t. one Parameter

do Random initialization multiple times

Regression Problem

$+5 \; -5 = 0$



$$\hat{y}_i = m x_i + c$$

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = J(m, c)$$

another notation for loss function

I want to optimize m & c ??

Find gradient of m & c

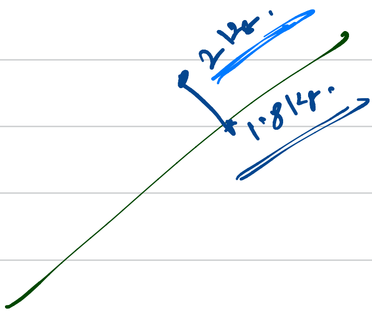
$$\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \Longleftrightarrow \sum_{i=1}^{n}\left(y_i-(mx_i+c)\right)^2$$

$$\frac{\partial J}{\partial m} \quad \& \quad \frac{\partial J}{\partial c}$$

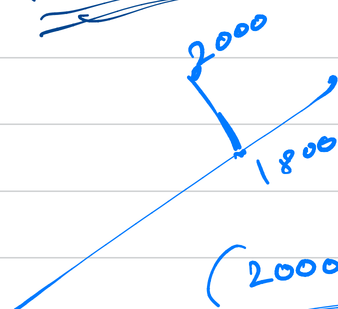
① $$\frac{\partial J}{\partial m} = \sum_{i=1}^{n} 2\cdot\left(y_i-(mx_i+c)\right)(-x_i)$$

$$\Rightarrow -\frac{2}{N}\sum_{i=1}^{n}\left(y_i-(mx_i+c)\right)\times x_i$$

② $$\frac{\partial J}{\partial c} = \frac{2}{N}\cdot\sum_{i=1}^{n}\left(y_i-(mx_i+c)\right)\cdot(-1)$$



$$(2-1.8)^2$$



$$(2000-1800)^2$$

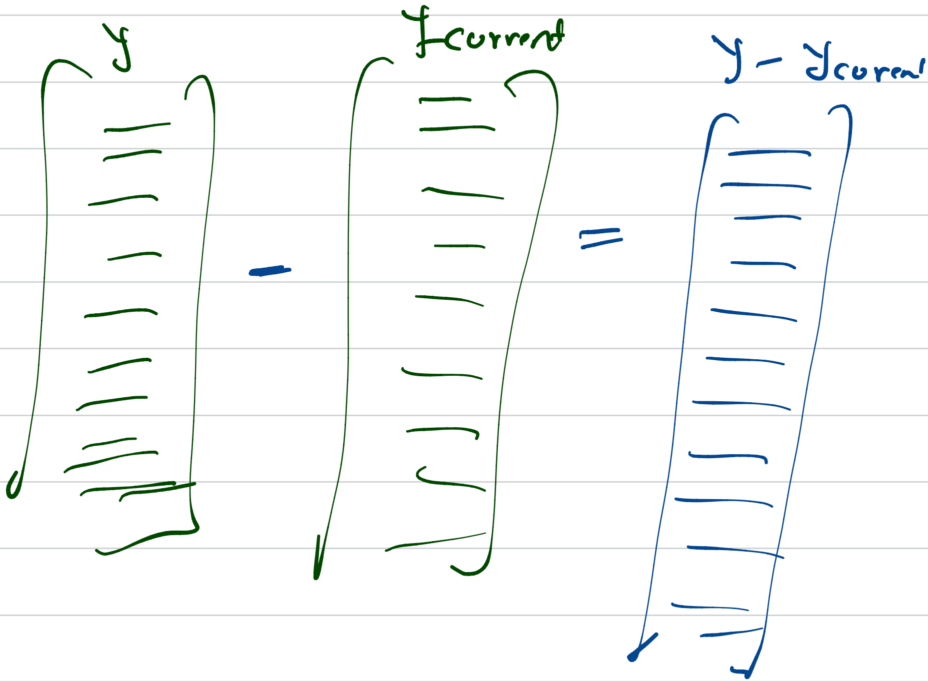
→ m_current

→ c_current

$y\_current = m\_current \; x + c\_current$

target Price

Feature area



d dimen

$$\vec{w}^{(t+1)} = \bar{w}^{(t)} - \eta \nabla_{\vec{w}} \times \alpha(X, \vec{w}, w_0)$$

Assignment

Q-5

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

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

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

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

$\frac{d}{dx} x^n = n \cdot x^{n-1}$

$$f'(x) = \frac{1 \times d(1+e^{-x})}{(1+e^{-x})^2}$$

$\frac{d f(g(x))}{dx} = f'(g(x)) \cdot g'(x)$

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

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

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

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

$$\sigma(x) \to \sigma(x)(1-\sigma(x))$$

$$\frac{d\, f(g(x))}{dx} = \underbrace{\underbrace{f'(g(x))}\cdot\underbrace{g'(x)}}$$

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

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

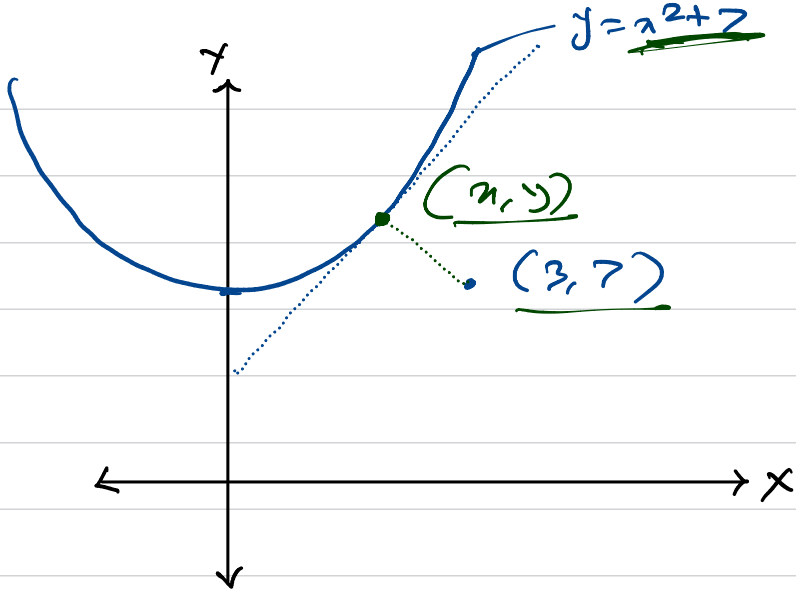
Dist b/w Pt & line $\frac{ax_1 + by_1 + c}{\|w_i\|}$

$y = x^2 + 7$ $\neq$ Linear line / Plane

<

> x

v



$$\sqrt{(x-3)^2 + (y-7)^2}$$

$$f(x) = (x-3)^2 + (y-7)^2$$

$$\rightarrow (x-3)^2 + (x^2+\cancel{7}-\cancel{7})^2$$

$$\Rightarrow x^2+9-6x \quad + x^4$$

$$f'(x) = 2x-6 \quad + 4x^3$$

$$= 4x^3+2x-6 = 0$$

$$x = 1$$

$y = \frac{\log x}{x}$

y'

$= \frac{-\log x}{x^2} + \frac{1}{x^2} = 0$

$\frac{\log x}{\cancel{x^2}} = \frac{1}{\cancel{x^2}} \Rightarrow \log x = 1$

$x = e^1$

$x = (4, 3, 5)$

$a = (1, 2, -1)$

$b = (3, 1, 5)$

$\frac{5}{\sqrt{6}} a + \frac{40}{\sqrt{35}} b$

$a = \frac{1}{\sqrt{6}} \quad \frac{2}{\sqrt{6}} \quad \frac{-1}{\sqrt{6}}$

$b = \frac{3}{\sqrt{35}} \quad \frac{1}{\sqrt{35}} \quad \frac{5}{\sqrt{35}}$

x $\quad$ y $\quad$ z

$\frac{5}{\sqrt{6}} \times \frac{1}{\sqrt{6}} \qquad \frac{5}{\sqrt{6}} \times \frac{2}{\sqrt{6}} \qquad \frac{5}{\sqrt{6}} \times \frac{-1}{\sqrt{6}}$

$\frac{40 \times 3}{35} \qquad \frac{40}{35} \qquad \frac{40 \times 5}{35}$

$\underbrace{\frac{5}{6} + \frac{120}{35}}_{x}$ $\underbrace{\frac{10}{6} + \frac{40}{35}}_{y}$ $\underbrace{\frac{-5}{6} + \frac{200}{35}}_{z}$

$\frac{5 \times 35 + 120 \times 6}{35 \times 6}$ $\frac{10 \times 35 + 40 \times 6}{35 \times 6}$ $\frac{-5 \times 35 + 200 \times 6}{35 \times 6}$

$\Rightarrow \frac{895}{210} x_i$, $\frac{590}{210} y_i$, $\frac{1025}{210} z$

local outlier cluster

1. Outlier / IQR / Impute / z-method
3-std

2. Standardization Vs Normalization
→ Know the range
↓
$\cancel{\#} > 100$
$\frac{x_i - x_{min}}{x_{max} - x_{min}}$

3. One-hot encoding

Don't know the range