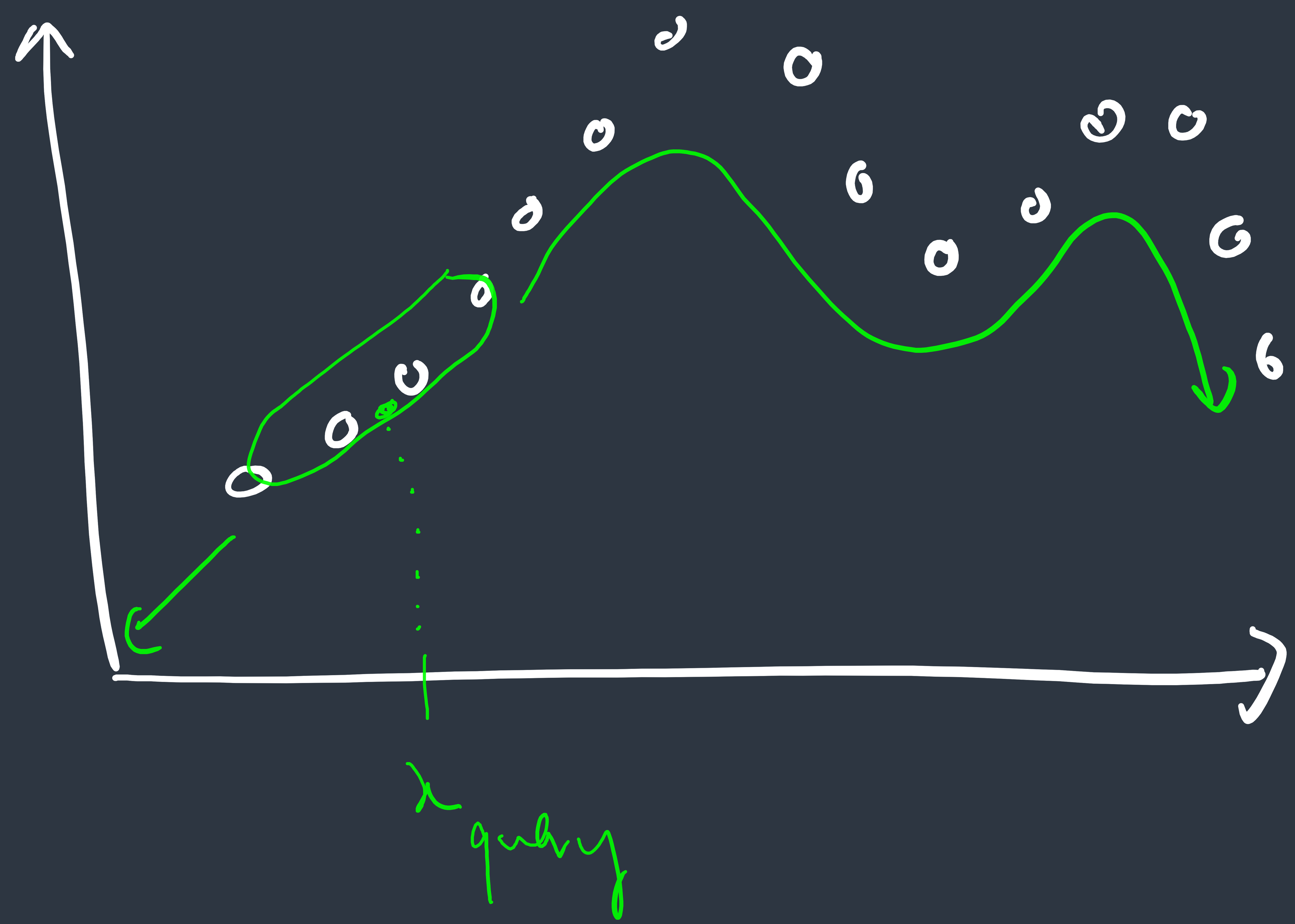
By -
[Prateek Narang]
CODING BLOCKS

Locally weighted Regression



→ Neighbours of the query point will have more.

→ Far aways pts will have less weight

Linear Regression

$$h_\theta(x) = \theta^T x$$

$$\text{Loss} = \sum_i \left(y^i - h_\theta(x^i)\right)^2$$

more loss

weighted loss

$$\sum_i w^{(i)} \left(y^{(i)} - h_\theta(x^{(i)})\right)^2$$

Any point

Query point

$$e^0 \approx 1$$

$$w^{(i)} = e^{-\frac{(x^{(i)} - x)^2}{2\tau^2}}$$

τ Bandwidth parameter [Hyperparam

$(0,1)$ if query point is close to given point

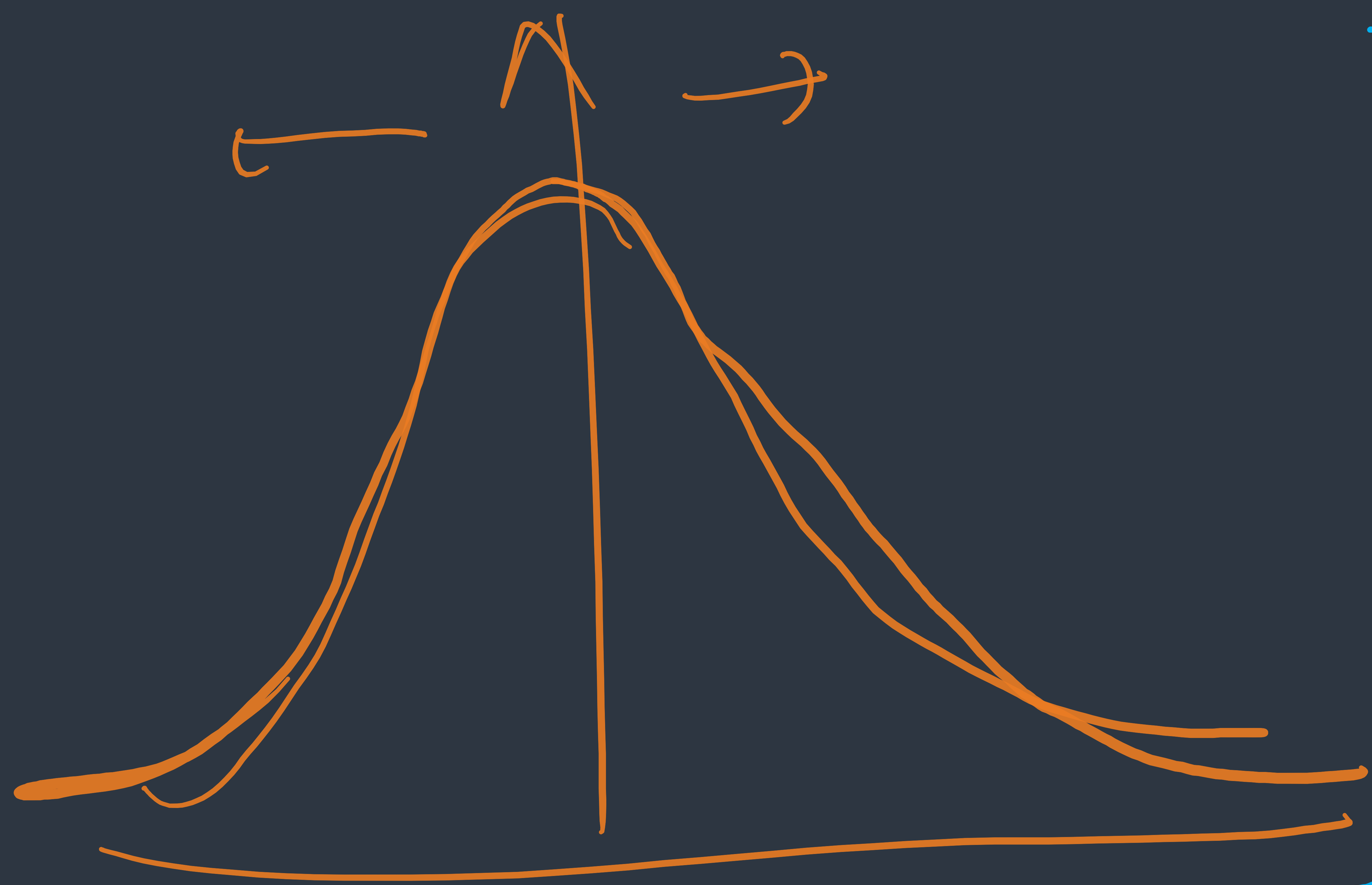
$$\frac{x' - x}{\tau} \to 0$$

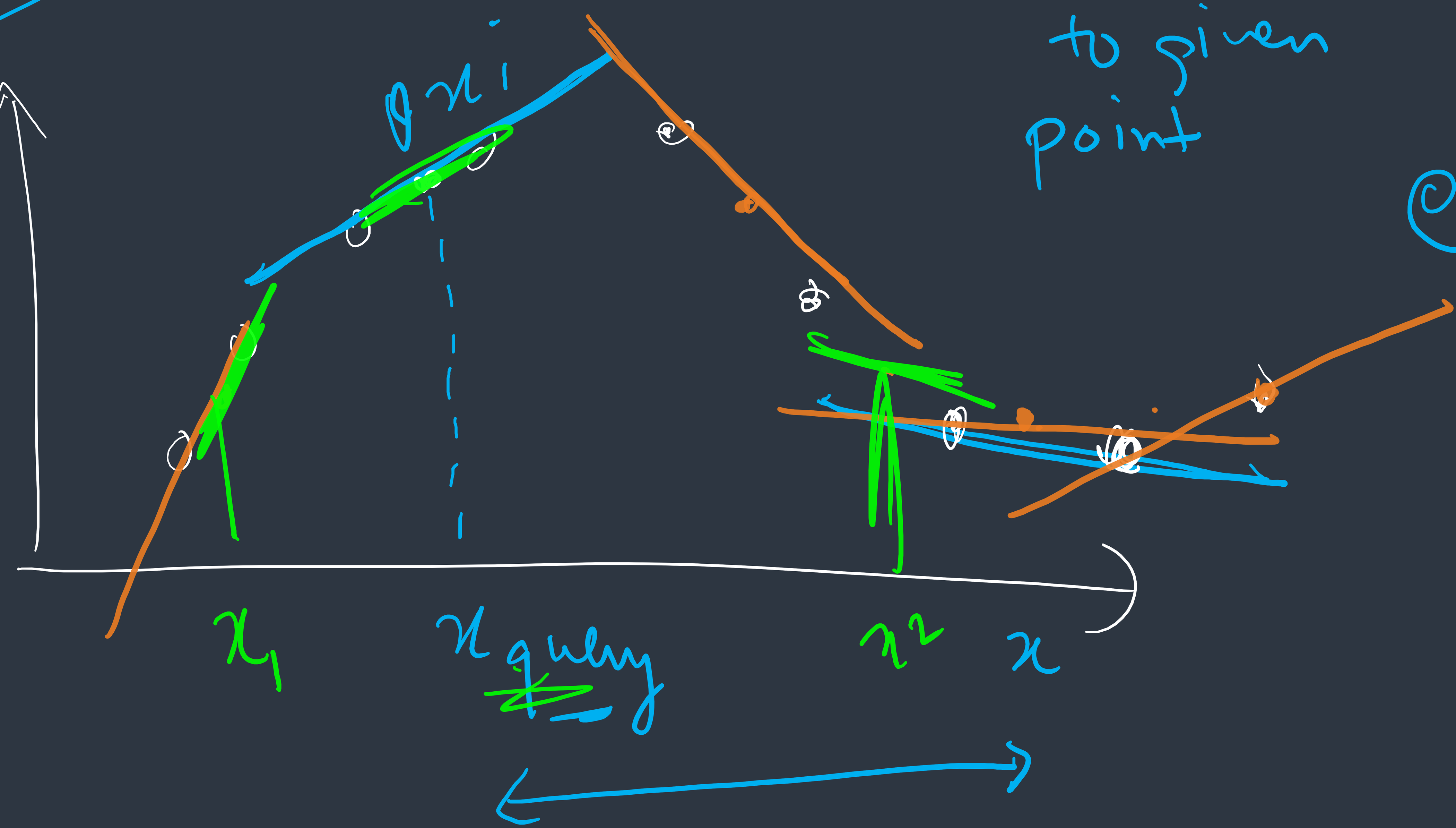
$$\boxed{e^0 \approx 1}$$

$$e^{-\infty} = \boxed{0}$$

$$x^{(i)} \approx x$$

$$x^{(i)} - x = 0$$





$e^0 = 1$

⇒ No general

Non-Parametric

⇒ For every test point, learn θ.

↳ Closed Form Soln θ = formula

↳ Gradient Descent ✓ → 1st

iterative

↳ Newton's Descent ✓ → 2nd order

Closed Form solution for Locally weighted Regression

$$\boxed{\min_{\theta}} \; Loss = \sum_{i=1}^{m} w^{(i)} \left[h_\theta(x^{(i)}) - y^{(i)} \right]^2 = (X\theta - Y)^2$$

scalar is different for every i

$i \in m$

(Scaling Loss)

$$= (X\theta - Y)^T (X\theta - Y)$$

$1 \times m$ $\quad$ $m \times 1$

Diagonal Matrix

$$\underbrace{\begin{bmatrix} K_1 & 0 & 0 \\ 0 & K_2 & 0 \\ 0 & 0 & K_3 \end{bmatrix}}_{\substack{3\times 3 \\ n\times n}} \underbrace{\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}}_{\substack{3\times 1 \\ n\times 1}} = \underbrace{\begin{bmatrix} K_1 \cdot 1 \\ K_2 \cdot 2 \\ K_3 \cdot 3 \end{bmatrix}}_{\substack{3\times 1 \\ n\times 1}}$$

scaled

Linear Algebra

$$\underbrace{\underbrace{\begin{bmatrix} (X\theta - y)^T \end{bmatrix}}_{1\times m} \begin{bmatrix} w_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & w_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & w_3 & 0 & 0 & 0 \\ 0 & 0 & & \ddots & & \\ & & & & & w_m \end{bmatrix}_{m\times m} \overbrace{\begin{bmatrix} h_\theta x^1 - y^1 \\ h_\theta x^2 - y^2 \\ h_\theta x^3 - y^3 \\ \vdots \end{bmatrix}}^{X\theta - y}_{m\times 1}}_{\text{Total Scalar}}$$

$$= 1\times 1 = \text{Scalar} \checkmark$$

$$\underbrace{X}_{m\times n}\ \underbrace{\theta}_{n\times 1} - \underbrace{y}_{m\times 1}$$

$$= [m\times 1]$$

$$\left(h_\theta(x^{(i)}) - y^{(i)}\right)$$

$$= w_1\left(h_\theta x^{(1)} - y^{(1)}\right)$$

$$\vdots$$

$$\underbrace{w^{(i)}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2}$$

$$J(\theta) = \sum_{i=1}^{m} w^{(i)} \left[h_\theta(x^{(i)}) - y^{(i)} \right]^2 = (X\theta - Y)^T \boxed{W} (X\theta - Y)$$

dependent on query point

$$W = \begin{bmatrix} w^1 & 0 & 0 & 0 \\ 0 & w^2 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & w^M \end{bmatrix}$$

Find $\nabla_\theta J(\theta) = 0$

$$\frac{\partial}{\partial \theta} J(\theta) = 0 = (\theta^T X^T - Y^T)(WX\theta - WY)$$

$$= \frac{\partial}{\partial \theta} \left(\theta^T X^T W X \theta - \theta^T X^T W Y - Y^T W X \theta + \cancel{Y^T W Y} \right)$$

$$= \frac{\partial}{\partial \theta}\left(\underline{\theta^T X^T W X \theta} - \underline{2\theta X^T W Y}\right)$$

$$\Rightarrow \quad 2 X^T W X \theta - 2 X^T W Y = 0$$

$$\Rightarrow \quad \theta = \frac{X^T W Y}{X^T W X}$$

$$\Rightarrow \boxed{\theta = (X^T W X)^{-1} (X^T W Y)}$$

closed form. solution for locally weighted Regres

if $\tau \to$ large
$W \to I$

$$\theta = (X^T X)^{-1} (X^T Y) \Rightarrow$$ Linear Regression