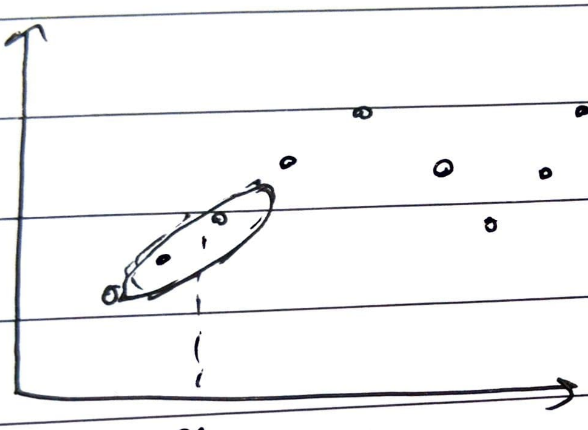
Locally Weighted Regression (LOWESS)

Idea:

It is not going to learn parameter/function for the curve like linear reg.



According to query:

(i) Neighbours of the query point will have more weight.

(ii) Far away points will have less weight.

(iii) weight will dec as we move away from point.

In LR:

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

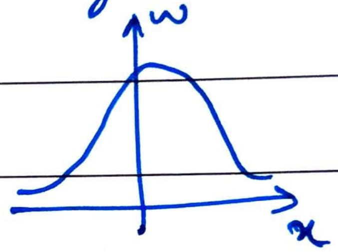
$$\text{loss} = \sum_i \left(y^i - h_\theta(x^i)\right)^2$$

$x \rightarrow$ If data pts have x from 1 to 100, algo will fit linear model for all points i.e. total 100 models.

* In lowess, we are going to have weighted loss.

$$\text{loss} = \sum_i \underbrace{w^{(i)}}_{\text{Only new thing added}} \left(y^{(i)} - h_\theta(x^{(i)})\right)^2$$

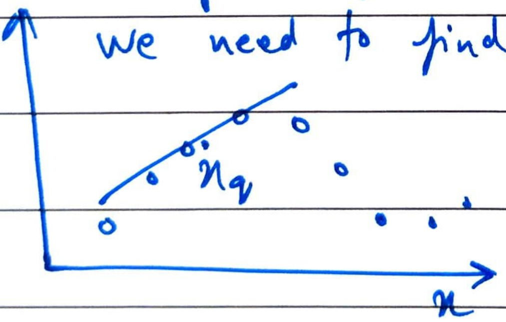
$$\boxed{w^{(i)} = e^{-\left(\frac{x^{(i)} - x}{2\tau^2}\right)^2}}$$



$x^{(i)} \rightarrow$ Any given pt. in dataset

$x \rightarrow$ Query points (usually but not always the data pts. in your sample)

we need to find slope & bias at this pt.



* Pts. which are closer to x_q have more weight in determining the slope at that point.

$\tau \rightarrow$ Bandwidth parameter (Controls how quickly weight falls)

Notice, $(x^{(i)} - x)$ is distance,

the $w^{(i)}$ will be :

* $\boxed{e^{-\infty} = 0}$ if they are far

* $x^{(i)} - x = 0$ if $x^{(i)} \simeq x$

$e^0 = 1$

$\boxed{w^{(i)} \text{ will lie btw } (0, 1)}$

So, we are adding more value to loss if points $x^{(i)} - x$ are close and less value if they are far away by multiplying $w^{(i)}$ in loss.

* It is a non-parametric algorithm.

Not going to learn any parameters.

OR: for every query point, learn some parameters

You can now use various methods to minimize loss function.

Closed form solution for loss:

$$\text{loss} = \sum_{i=1}^{m \text{ (examples)}} w^{(i)} \left[h_\theta(x^{(i)}) - y^{(i)} \right]^2$$

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

$$W = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ 0 & 0 & \ddots & w_m \end{bmatrix} \quad \left[\text{matrix of all weights}\right]$$

$$\boxed{\text{Find, } \nabla_\theta J(\theta) = 0}$$

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

(Closed form sol^n for LOWESS)