Linear Regression – Recap.

$(x^{(i)}, y^{(i)}) \longrightarrow i^{th}$ training example.

$x^{(i)} \rightarrow \mathbb{R}^{d+1}$ $\quad x_0, x_1 \ldots x_d$

$x_0 = 1$

$y^{(i)} \rightarrow \mathbb{R}$

$d \rightarrow$ No. of features

$n \rightarrow$ No. of training samples

$$h_\theta(x) = \sum_{j=0}^{d} \theta_j x_j = \theta^T x$$

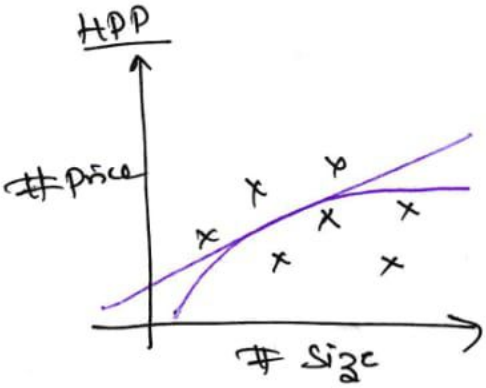
$$\theta^T = [\theta_0 \ \theta_1 \ \theta_2 \cdots \theta_d]_{1 \times d}$$

Note: Dimension should be 1x(d+1)

$$X = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_d \end{bmatrix}_{d \times 1}$$

Note: Dimension should be (d+1)x1

$$J(\theta) = \frac{1}{2n} \sum_{i=1}^{n} (h_\theta(x^{(i)}) - y^{(i)})^2$$



$\theta_0 + \theta_1 x_1 \qquad y_1$

$\qquad \downarrow \qquad\qquad \Downarrow$

$\quad$ # Size $\quad$ # Price

$\theta_0 + \theta_1 x_1 + \theta_2 x^2$

quadratic

$\theta_0 + \theta_1 x_1 + \theta_2 \sqrt{x} + \ldots$

$\qquad\quad \Downarrow \qquad\quad \Downarrow$

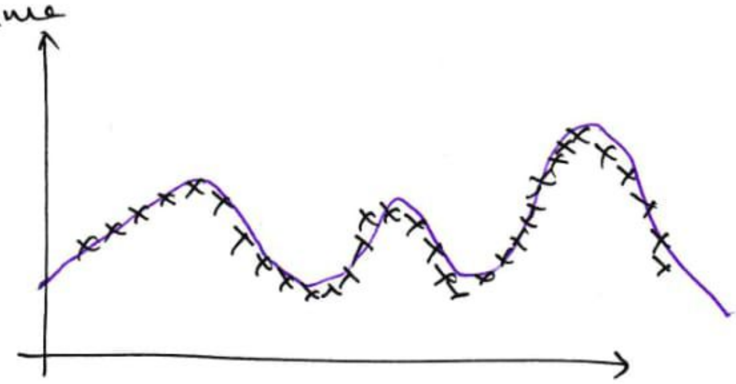
New $x_1 \qquad x_2$

So which kind of features do we use?!

x_1^2 or $\sqrt{x}$??

Locally Weighted Regression (LWR)

Assume



So, how to fit a curve onto this data??

x^2 or $\sqrt{x}$ or $\log x$??

Terminology.

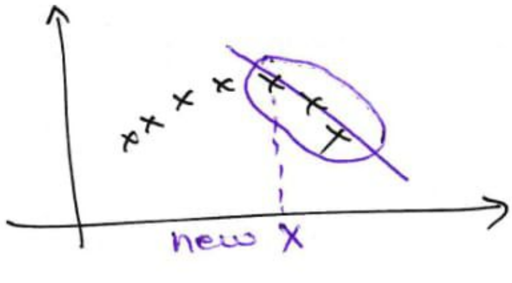
"Parametric" learning algorithm.

fixed set of parameters (θ) $\xrightarrow{\text{fitted}}$ data

"Non Parametric" learning algorithm

parameters (θ) $\underset{\text{linearly}}{\xrightarrow{\text{grows}}}$ size of data set.

But does help with the above data without manually selecting features.



LR

Fit θ to minimize

$$J(\theta) = \frac{1}{2n} \sum_{i=1}^{n} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

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

new x.

LWR

Fit θ to minimize

$$J(\theta) = \frac{1}{2n} \sum_{i=1}^{n} W^{(i)} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

$W^{(i)} \rightarrow$ weight parameter

$$w^{(i)} = \exp\left(-\frac{(x^{(i)}-x)^2}{2\tau^2}\right) \quad 0 < w^{(i)} < 1$$

Intuition :-

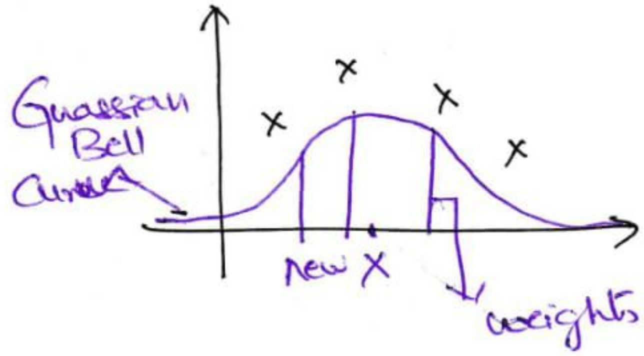
If $|x^{(i)} - x|$ is small $\Rightarrow \exp(\approx 0) = 1$

$\Rightarrow w^{(i)} \approx 1$

If $|x^{(i)} - x|$ is large $\Rightarrow \exp(\approx \text{large})$

$\Rightarrow w^{(i)} \approx 0$

$\sum_{i=1}^{n} w^{(i)} (h_\theta(x^{(i)}) - y^{(i)})^2 \Rightarrow$ Ends up Summing only Squared Error for Samples that are close to new x.



$\Rightarrow$ near points to new x get bigger $w^{(i)}$ values

Width of the Guassian bell curve

τ = "band width" (hyper) parameter

$\tau \rightarrow$ Tow

depending on the value bigger or narrower curve.

decides

How many examples to consider near to new 'x'

$\tau \Rightarrow$ too broad $\Rightarrow$ overfitting

too narrow $\Rightarrow$ underfitting