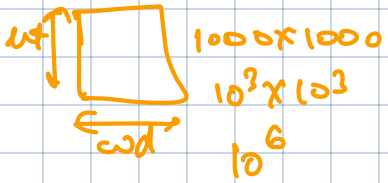
Supervised Learning Algo → $\{x^{(i)}, y^{(i)}\}_{i=1}^{m}$

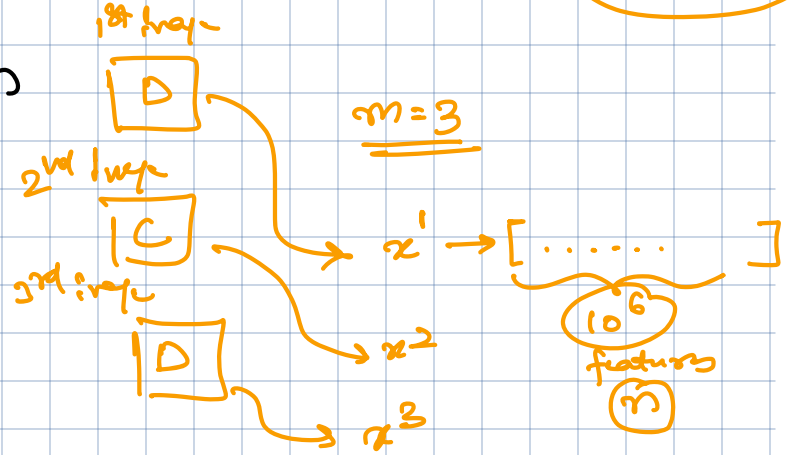
m = no. of examples

Classification

Regression

Dogs, Cats





$x^{(i)} = (x_1^{(i)}, x_2^{(i)}, x_3^{(i)}, \ldots, x_n^{(i)})$

$= (20, 30, 5, 0, \ldots 100)$

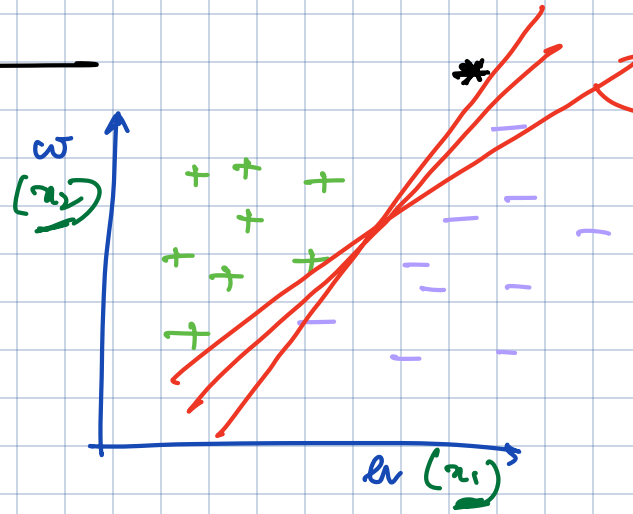
10^6 different values

$x_j^{(i)}$: j^{th} feature of i^{th} example

Classification:

(+) Cat → ht → w

(-) Dog → ht → w



Linear Classifier

Easier → $y = \theta_1 x + \theta_0$

$y = mx + c$

$\theta_2 x_2 + \theta_1 x_1 + \theta_0$

x_1, x_2

left side < 0

right side > 0

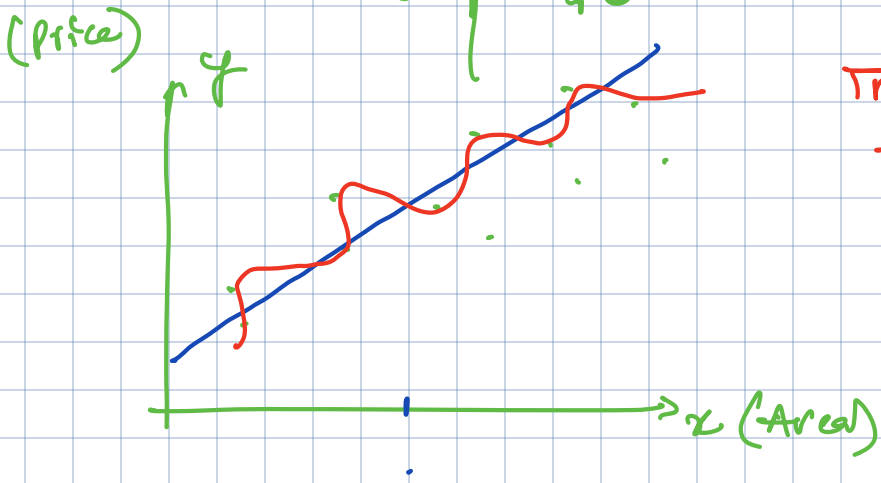
$\{x^{(i)}, y^{(i)}\}_{i=1}^{m}$

$y^{(i)}$ → Cat, → Dog } Discrete values

Regression

(x) Area	(y) Price
40	50
80	90
100	200
20	80

Q: 50 ?



Training Data

Test Data → Overfitting

Linear Regression

Eg:

Time Spent	Marks
1	4
3	7
10	8
20	10

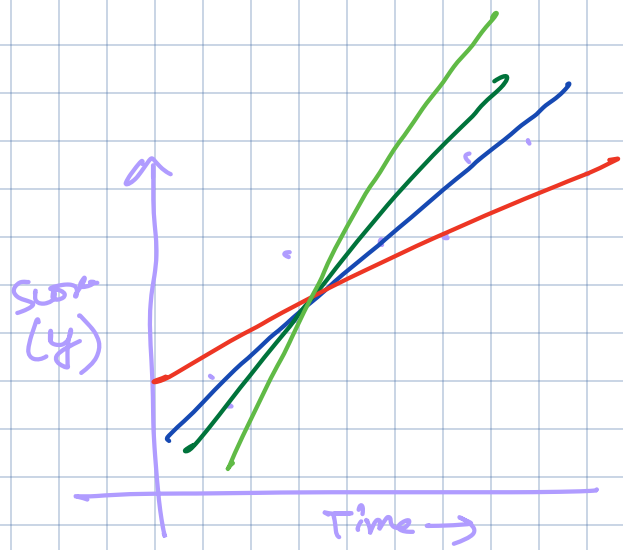
Training Data

Q: 8 hrs ? Score ?

$y = mx + c$

$y = \theta_1 x + \theta_0$

Aim: θ_0, θ_1 ?



hypothesis

$h_\theta(x) = \theta_1 x + \theta_0$ $\quad$ $f(x) = mx + c$

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}$$

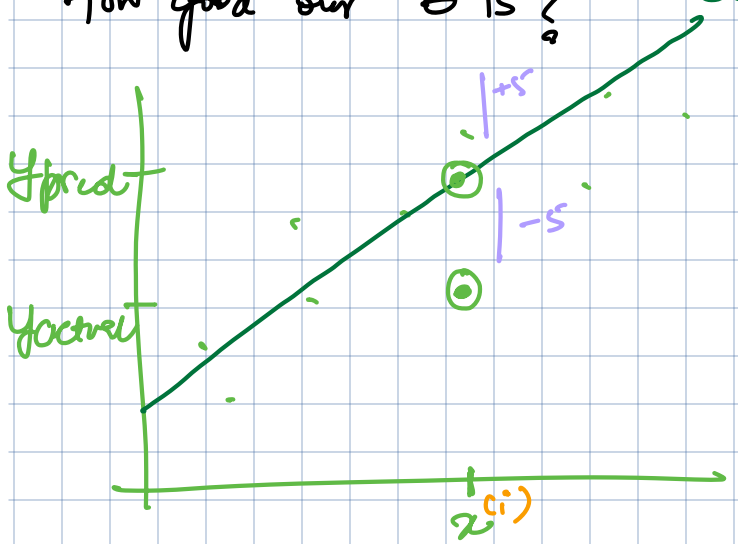
Aim: to learn best curve which fits through data points

(a) Random value of θ start

(b) how good our θ is ?

(c) θ change/update good performance.

How good our θ is ?



$$\varepsilon^{(i)} = \left| y^{(i)}_{pred} - y^{(i)}_{actual} \right|$$

$\varepsilon^{(i)}$ → error for ith example

Total error for all points:

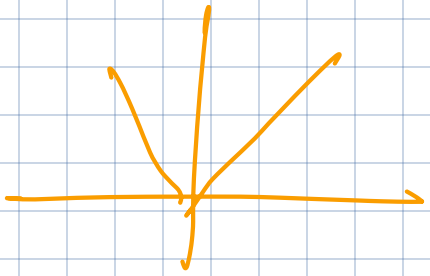
$$\sum_{i=1}^{m} \left| y^{(i)}_{pred} - y^{(i)}_{actual} \right|$$

(m = no. of examples/data points)

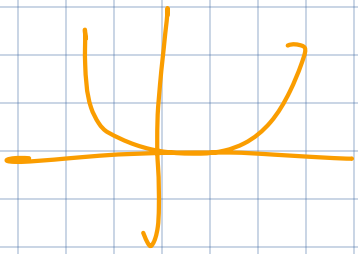
$$\sum_{i=1}^{m} \left| \hat{y}^{(i)} - y^{(i)} \right|$$

Average Absolute Error $= \frac{1}{m} \sum_{i=1}^{m} \left| \hat{y}^{(i)} - y^{(i)} \right| \rightarrow \varepsilon^{(i)}$

[-4]



x^2



J

Mean squared Error

$MSE = \frac{1}{m} \sum_{i=1}^{m} \left[\hat{y}^{(i)} - y^{(i)} \right]^2$

MSE → Loss or error function

$\hat{y}^{(i)}$ → predicted

$y^{(i)}$ → Actual

Attendance

428	501	EC/S
458	502	38
459	510	90
489	525	141
495	528	227
497	532	