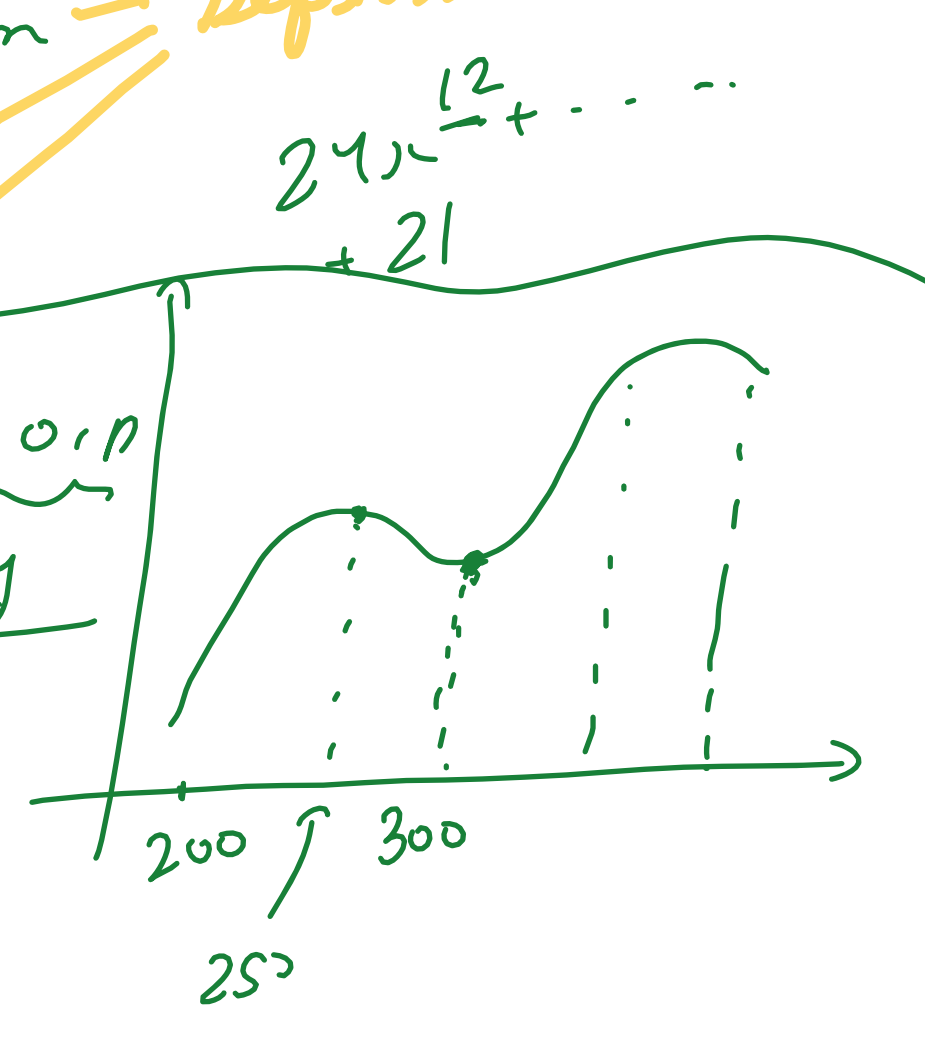
Topics

→ Supervised learning
- ↳ linear regression & multiple regression
  - ↳ line equation, cost function, gradient descent, losses, accuracy, precision & recall, vectorization (matrices)
- ↳ logistic regression
  - ↳ line eq^n, cost function, gradient descent, loss, accuracy, precision & recall, decision boundary,
- ↳ Z-score normalization — Definition
- ↳ learning rate (α)
- ↳ feature engineering

i/p				o/p
x_1	x_2	x_3	x_4	y



$24x^{12} + \dots + 21$

$21x_1^2 + 12x_2^3 + 25x_3 + 14x_4^7$ → $x_1 x_2^2$, $x_1^2 x_2 x_3$

$25x^{12} + 21x^y$

2

imp: → Overfitting (high variance) & underfitting (high bias)

imp: → Regularization (to reduce overfitting)
- ↳ ridge (L2)
- ↳ LASSO (L1)

Unsupervised learning :-

→ K-means clustering
→ Decision tree, random forest (XG-boost)
→ Reinforcement learning (reward & punishment learning)
  ↳ Q-points
→ Anomaly detection
→ Recommender systems (eg.. Netflix), (Tensorflow)
→ Skewed datasets

Python → numpy, pandas, matplotlib & Seaborn,
- numpy → numerical & maths
- pandas → excel, CSV, manipulation
- matplotlib & Seaborn → plots & graphs

Scipy, Sklearn ☆, tensorflow,
- Scipy → scientific python
- Sklearn → (linear, logistic, decision trees)
- tensorflow → unsupervised learning

pytorch, pillow & CV,
- pytorch → (ML algorithms)
- pillow & CV → Image processing (computer vision)

Algebra → Machine learning → Deep learning → computer vision
↳ python → Machine learning
Machine learning → computer vision

ML

size of house (in sq.ft)	price (in K)
1000	250
1200	271
1400	301
1600	341

input

1315 sq → model → price

probability (80 – 100%)

ML → Supervised, Unsupervised

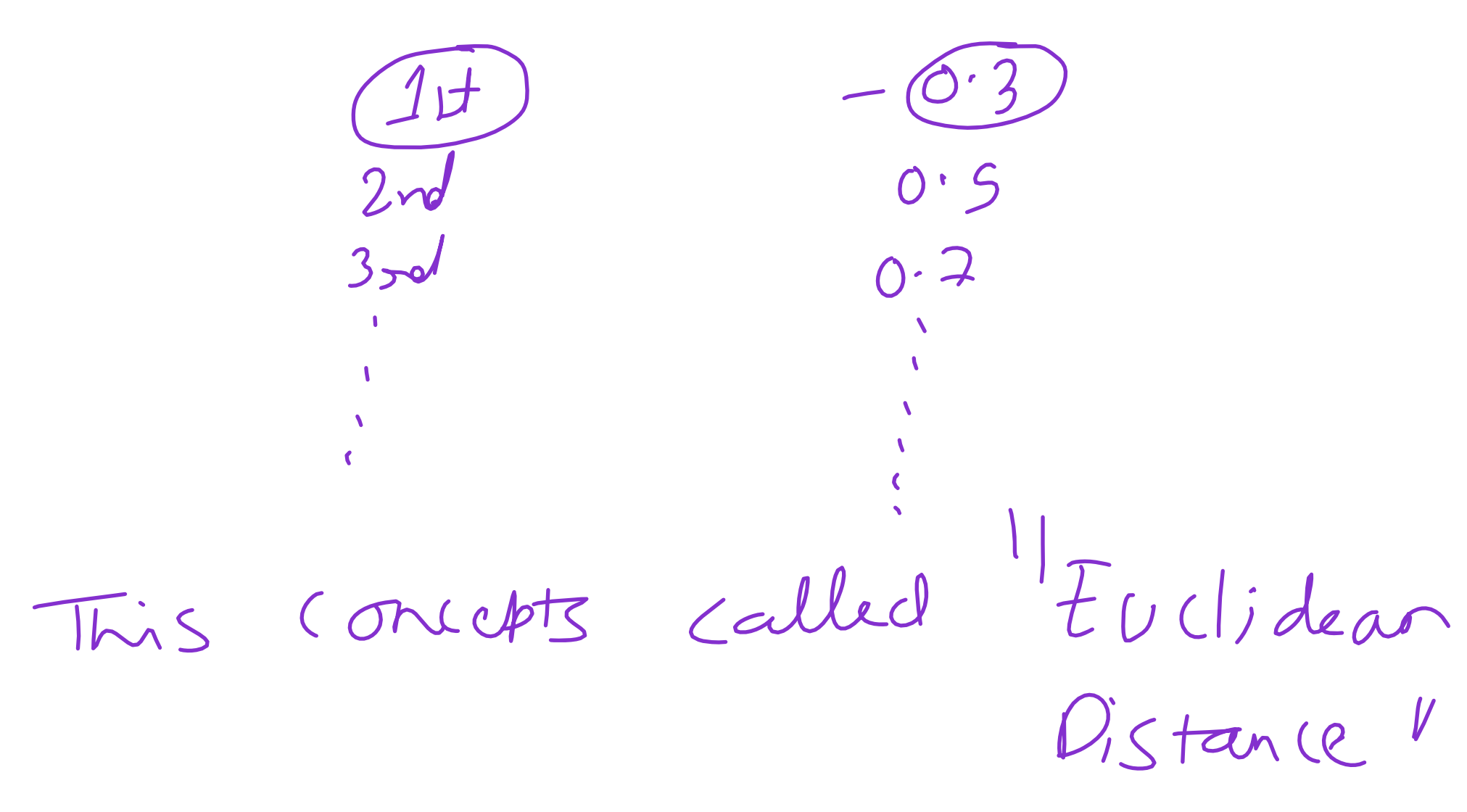
Supervised:

input	output

Unsupervised: input → news feed; netflix recommendation.

crime,

news
- UP crime, crime patrol, Saavdhaan India, Aaj tak crime → crime
- Dating news → dating
- Business news → business



Euclidean distance $= \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$

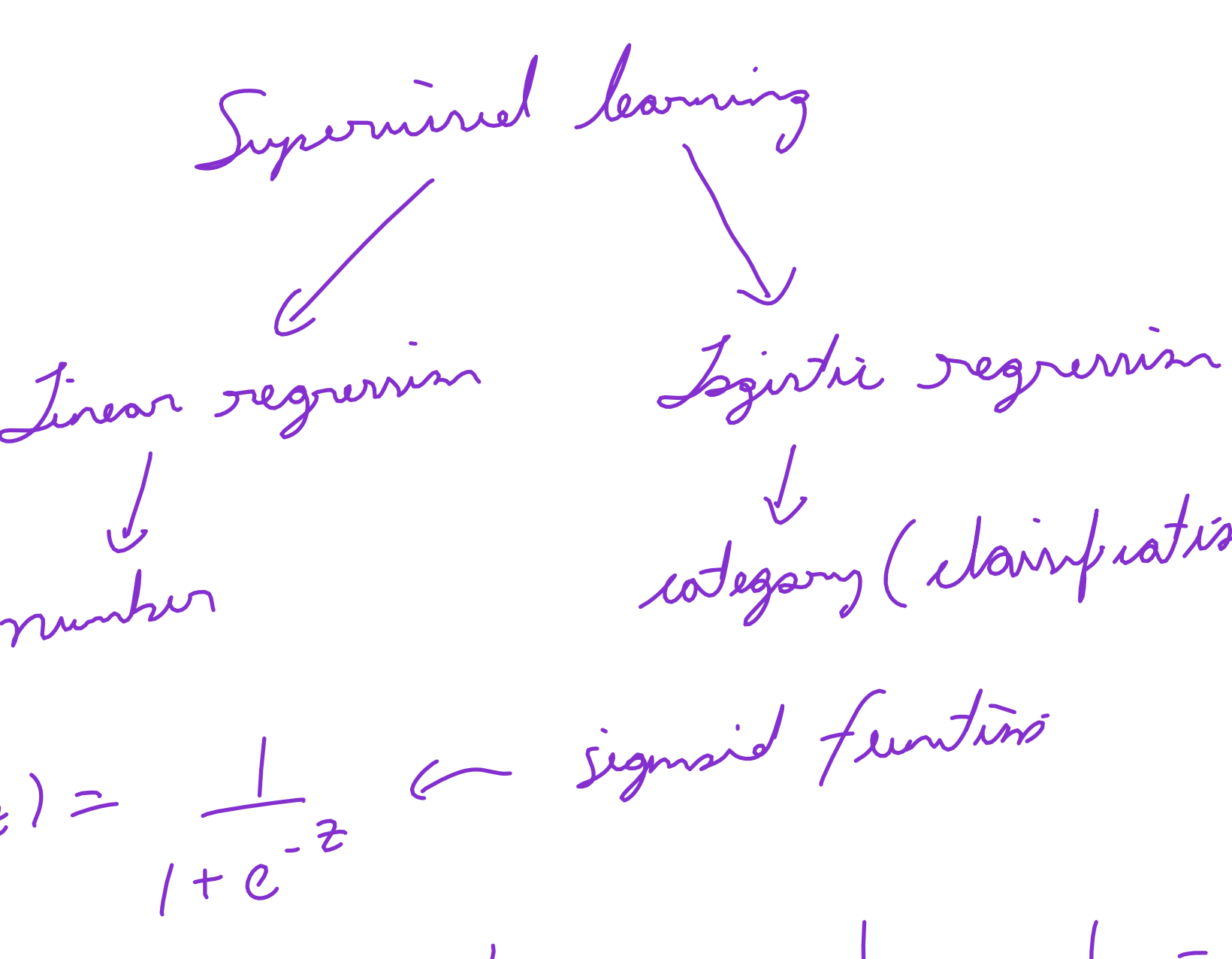
x_2, y_2 ×
× x_1, y_1

1st	– 0.3
2nd	0.5
3rd	0.7
⋮	⋮

This concepts called "Euclidean Distance"

Size of house	price (in K)
500	200
560	215
600	225
625	300

630 sq.



test on unseen data in supervised learning.

Supervised learning
- Linear regression → number
- Logistic regression → category (classification)

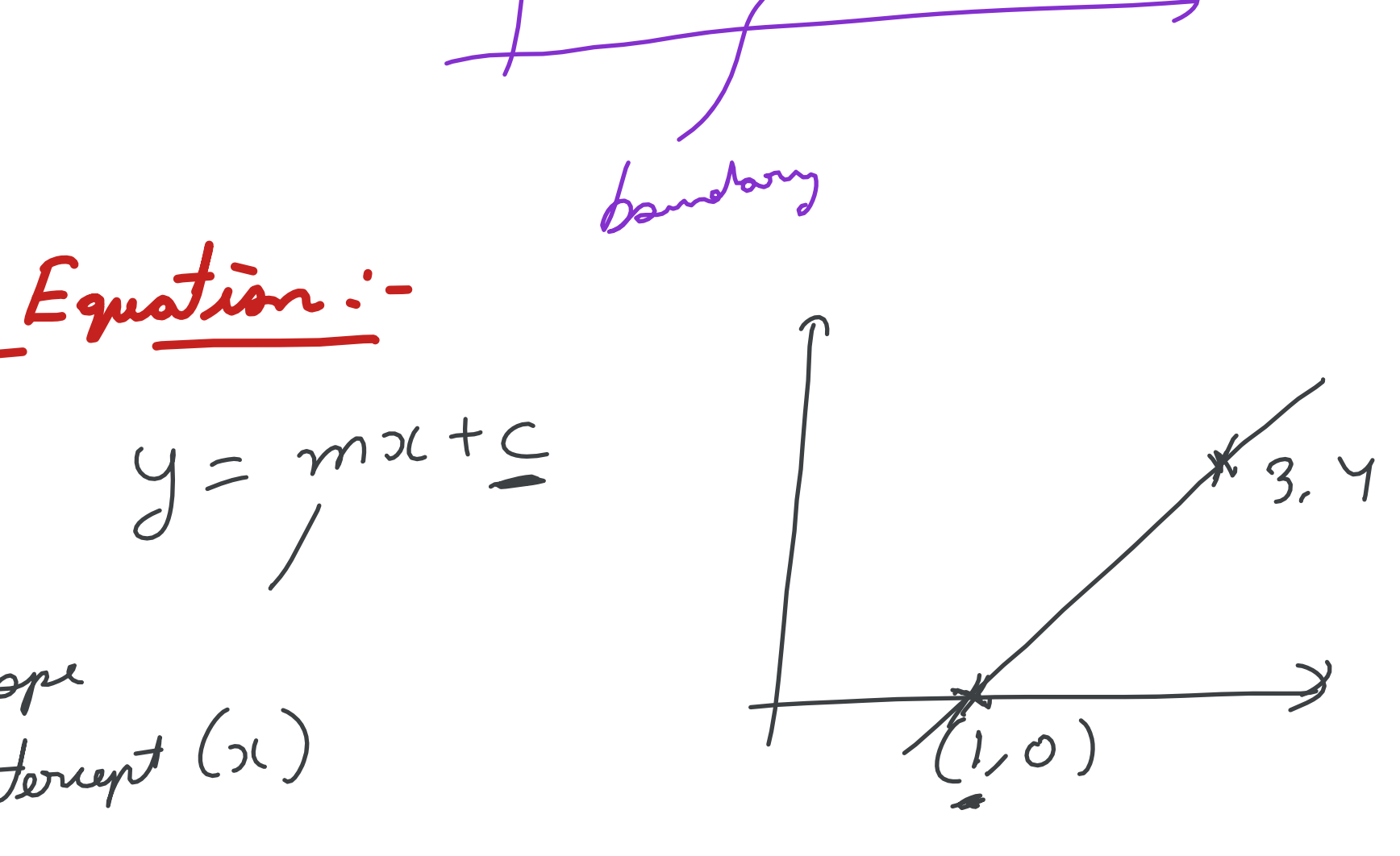
$f(z) = \dfrac{1}{1+e^{-z}}$ ← sigmoid function

$z = -\infty = \dfrac{1}{1+e^{-(-\infty)}} = \dfrac{1}{1+e^{\infty}} = \dfrac{1}{\infty} = 0$

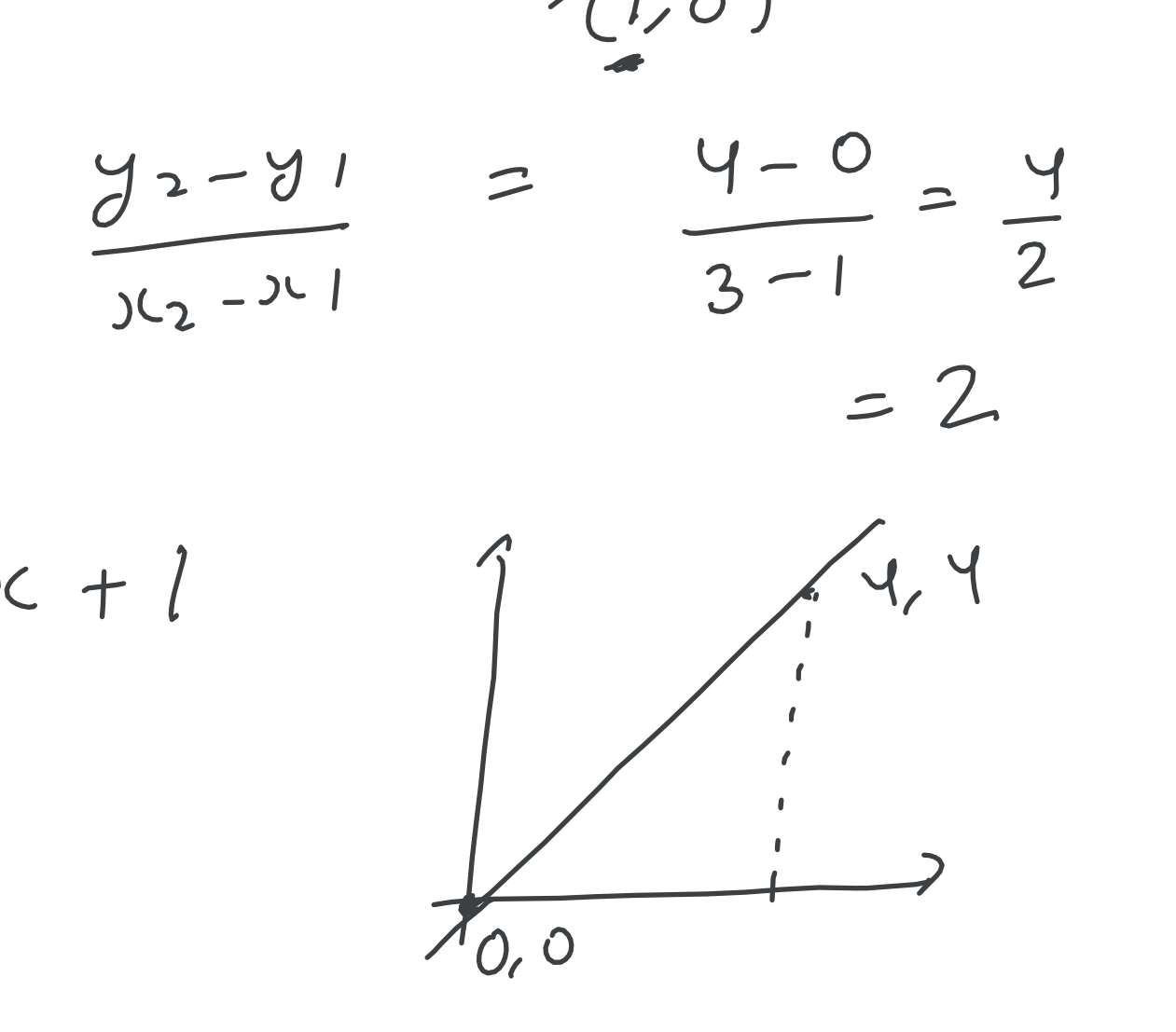
$z = +\infty = \dfrac{1}{1+e^{-(\infty)}} = \dfrac{1}{1+e^{-\infty}} = \dfrac{1}{1+0} = 1$

$z = 0 = \dfrac{1}{1+e^{-0}} = \dfrac{1}{1+1} = \dfrac{1}{2} = 0.5$

Sigmoid f^n :-



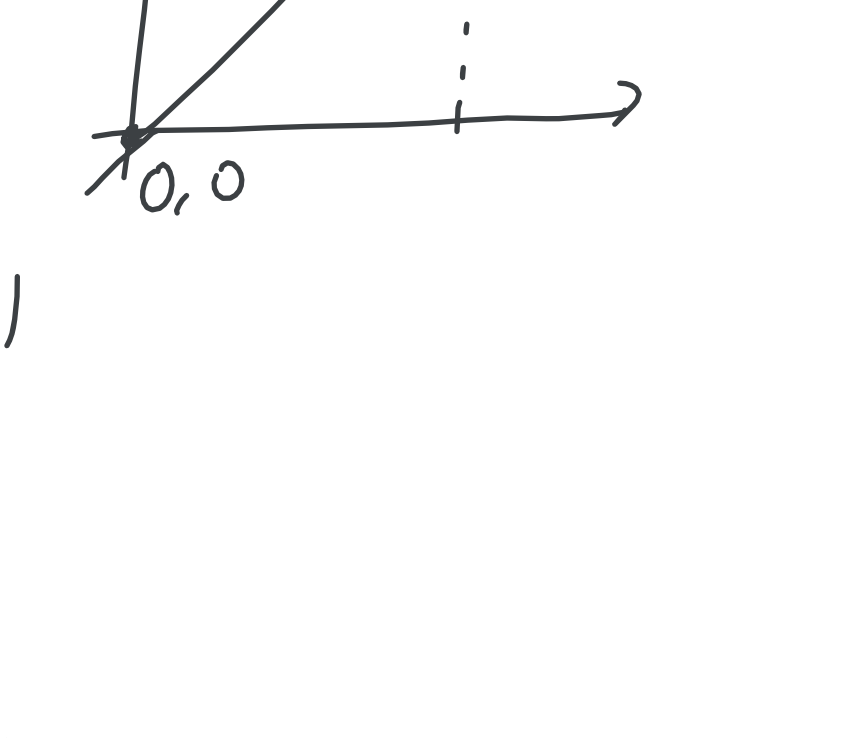
threshold (boundary) value = 0.5



Line Equation :-

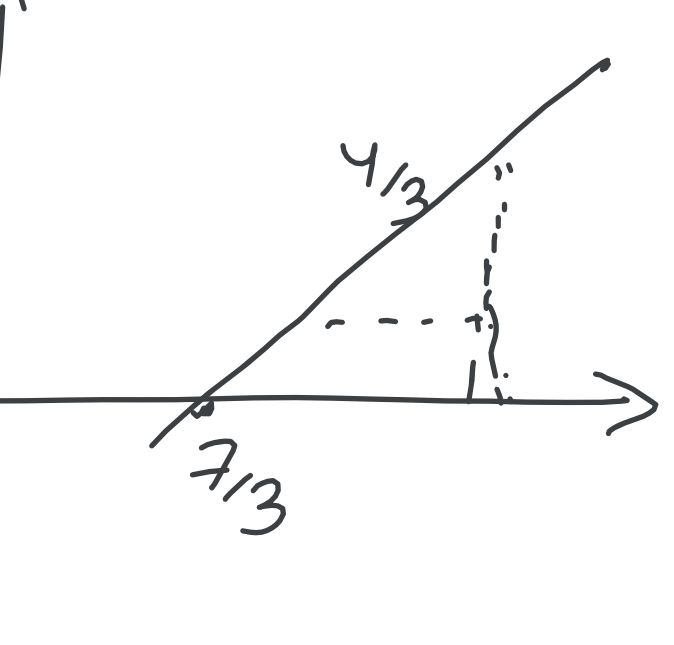
$y = mx + c$

m = slope
c = intercept (x)



$m = \dfrac{y_2 - y_1}{x_2 - x_1} = \dfrac{4-0}{3-1} = \dfrac{4}{2} = 2$

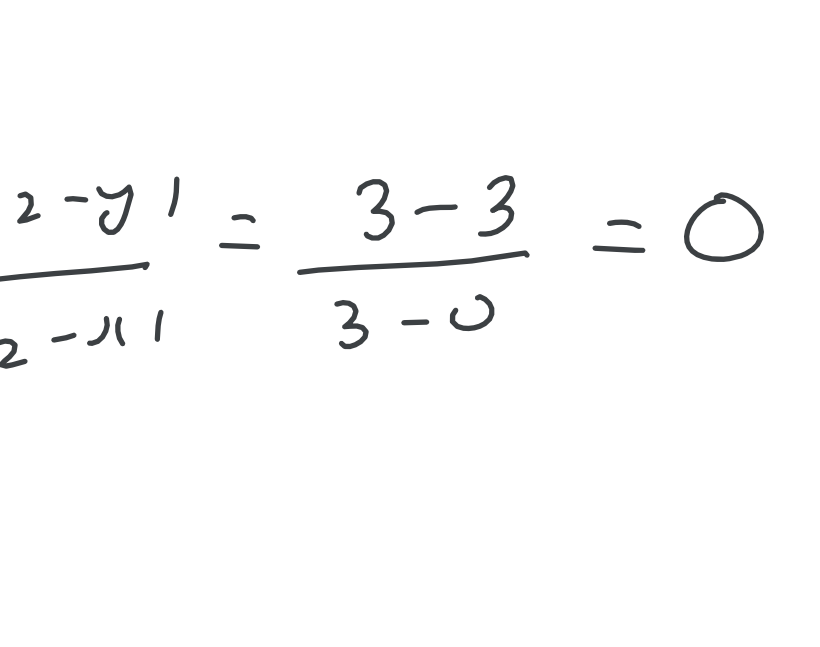
$y = 2x + 1$



$m = \dfrac{4-0}{4-0} = \dfrac{4}{4} = 1$

$y = mx + c$
$y = x + 0$; $y = x$.

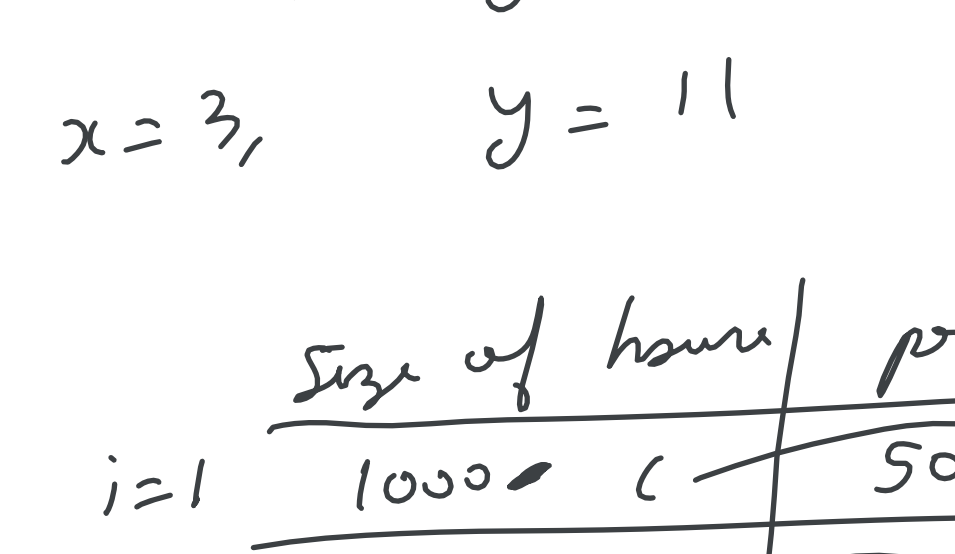
$3y = 4x + 7$
$y = \dfrac{4}{3}x + \dfrac{7}{3}$.



$f(x) = wx + b$ (w → m, b → c)

$f_{w,b}(x) = wx + b$.

w, b = parameters



$m = \dfrac{y_2 - y_1}{x_2 - x_1} = \dfrac{3-3}{3-0} = 0$

$y = mx + c$
$y = 0 \times x + 3$, $y = 3$

$y = 3x + 2$ $x = 1$

$y = 5$,

$x = 2$, $y = 8$
$x = 3$, $y = 11$

	Size of house	price
i=1	1000	50
i=2	1200	60
i=3	1400	70
i=4	1600	75
i=5	1800	85

$f_{\vec{w},\vec{b}}(\vec{x}^{(1)}) = wx^{(1)} + b$
$f_{w,b}(\vec{x}^{(2)}) = wx^{(2)} + b$
$f_{w,b}(\vec{x}^{(3)}) = wx^{(3)} + b$