

KNN

Algorithm

KNN - k Nearest Neighbors

↳ Distance based ✓

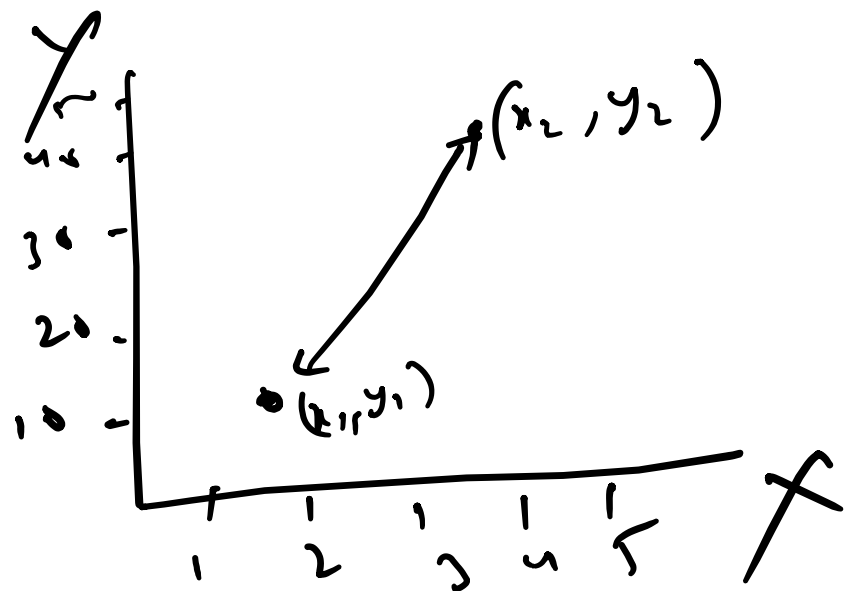
↳ Choose 'k'

↳ Lazy algorithm

Classification

Regression

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



Euclidean distance

Manhattan

Minkowski

$$\left((x_2 - x_1)^2 + (y_2 - y_1)^2 \right)^{1/2}$$

(x_1, y_1) 1 pt $\rightarrow (2, 10)$
 (x_2, y_2) 2 pt $\rightarrow (4, 40)$

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

$$= \sqrt{(5 - 2)^2 + (3 - 2)^2}$$

$$= \sqrt{3^2 + 1} = \sqrt{10} = 3.16$$

Scaling →

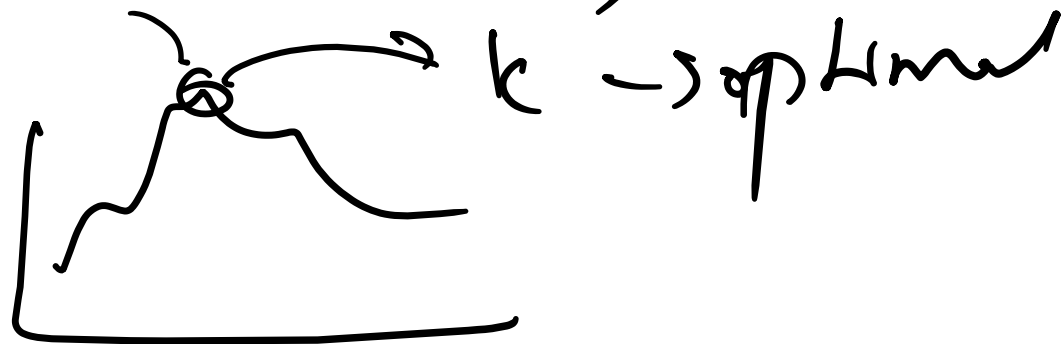
2-score
Normalization

Distance
band



$k=5$
 $k=3$

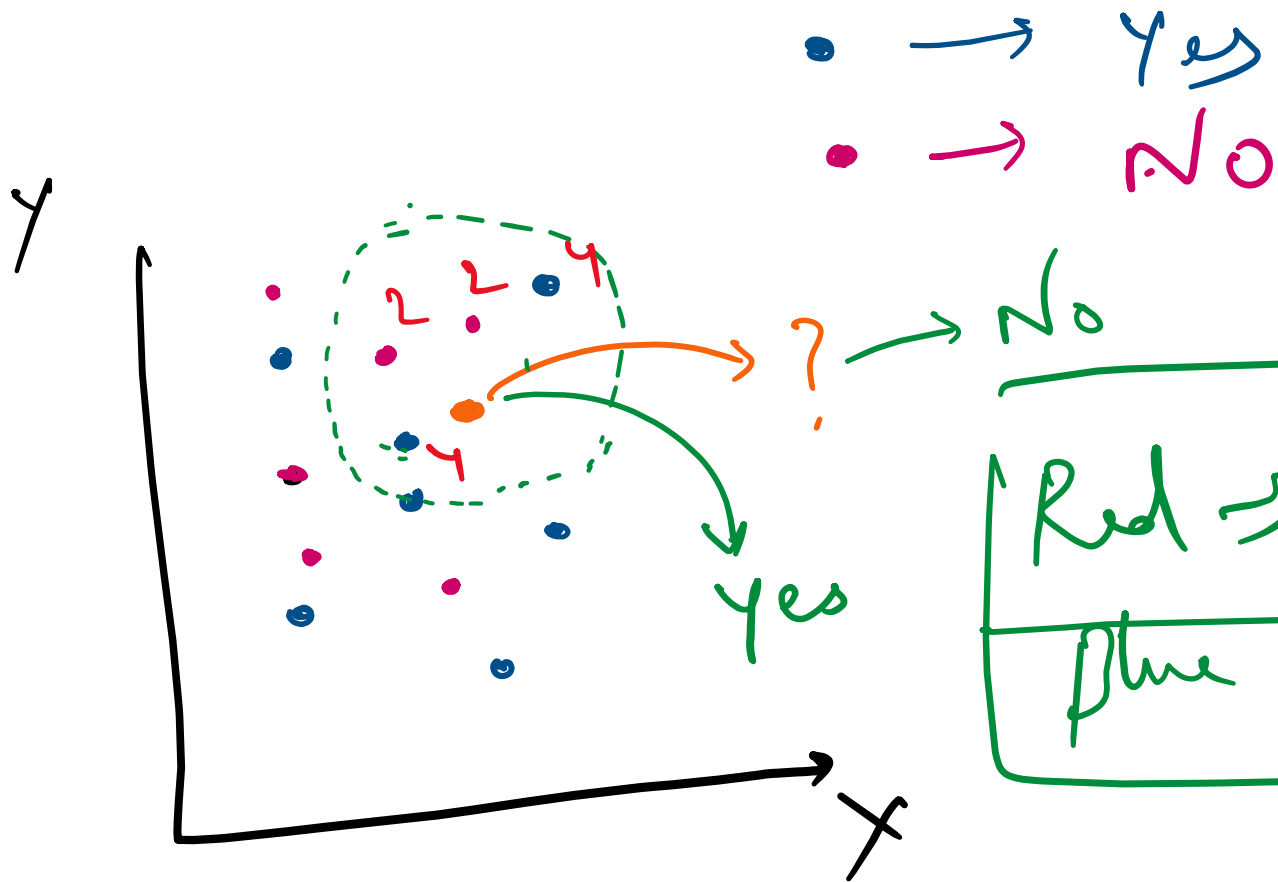
(nearest value)



< 20

< 40

1, 2, 4, 6, 8, 10, ... 40
1 — 20
No magic 'k' value



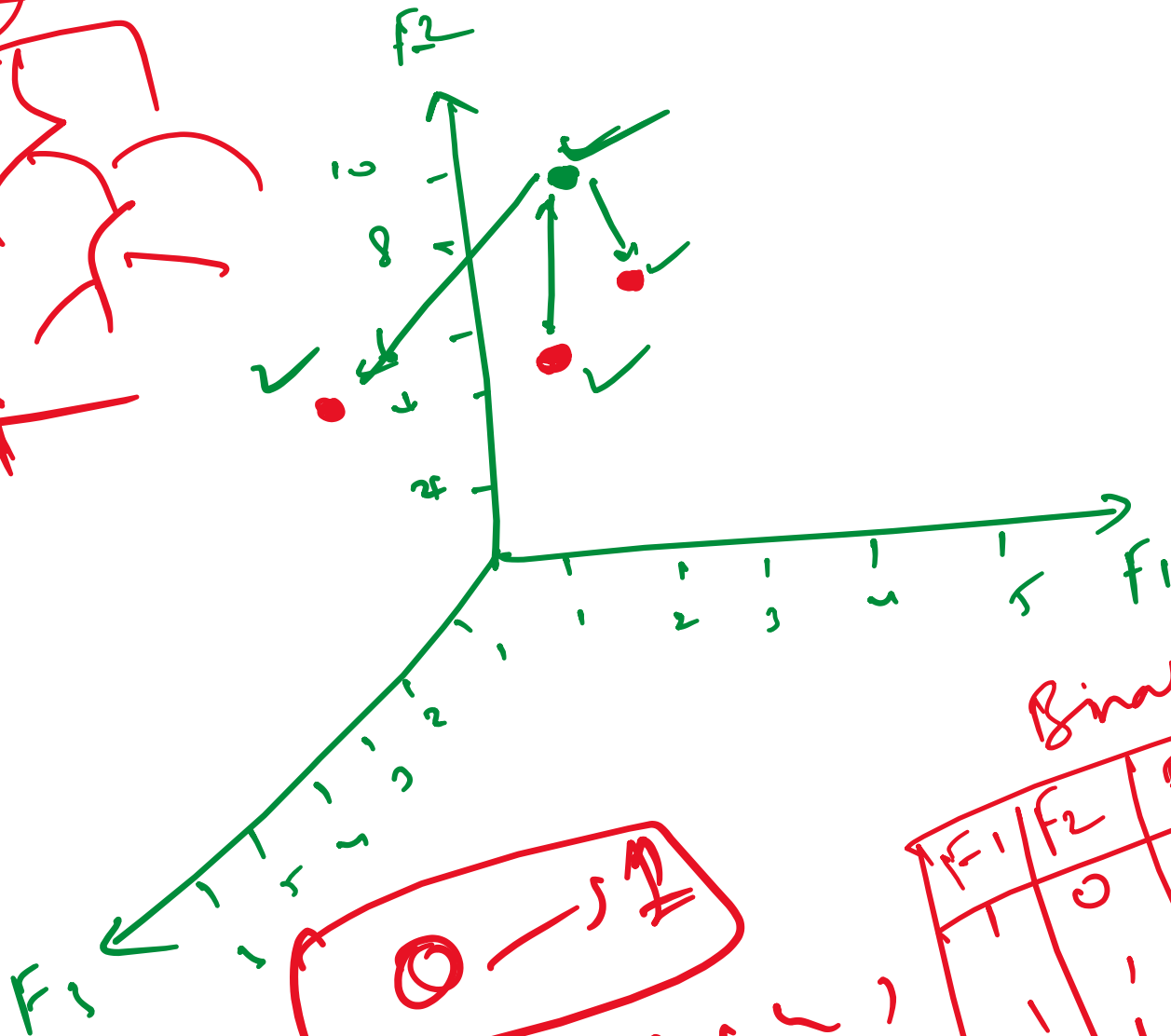
$k=3$
$k=5$

Red → 2
Blue → 3

kNN Regression
kNN classifier

$k_1 \rightarrow 2$
 $k_2 \rightarrow 4$
 $k_3 \rightarrow 4$

~~New~~
 New - $2 + 4 + 4 = 10/3$
 $= 3.3$



F1	F2	F3
1	2	4
4	2	4
3	2	1

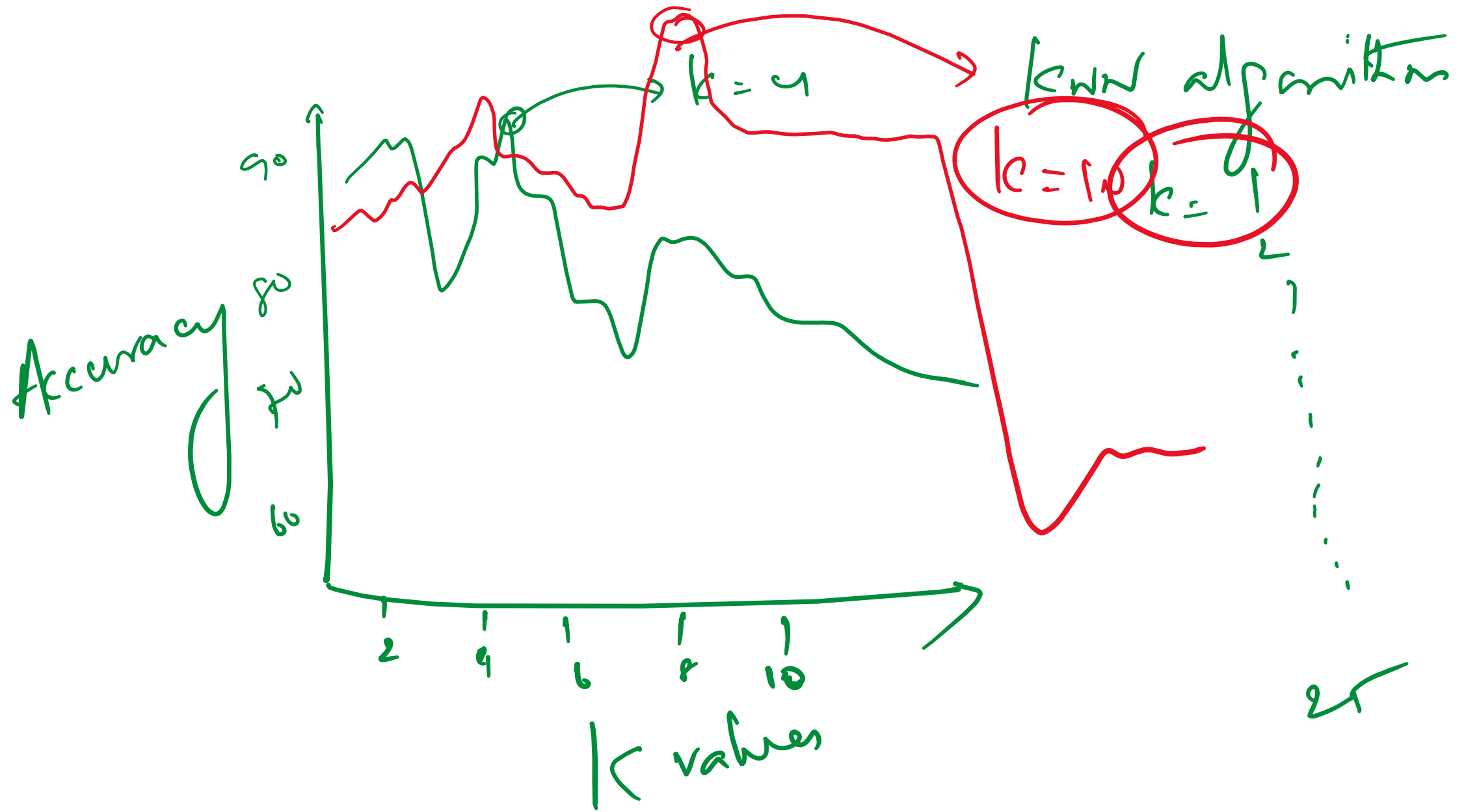


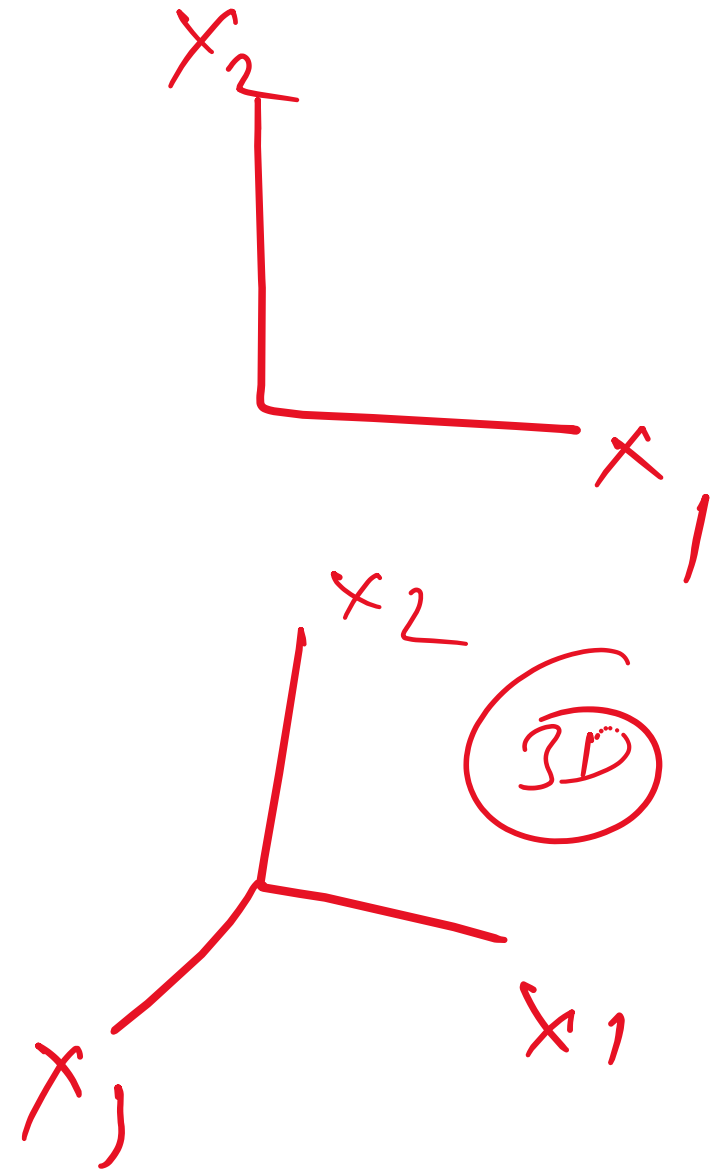
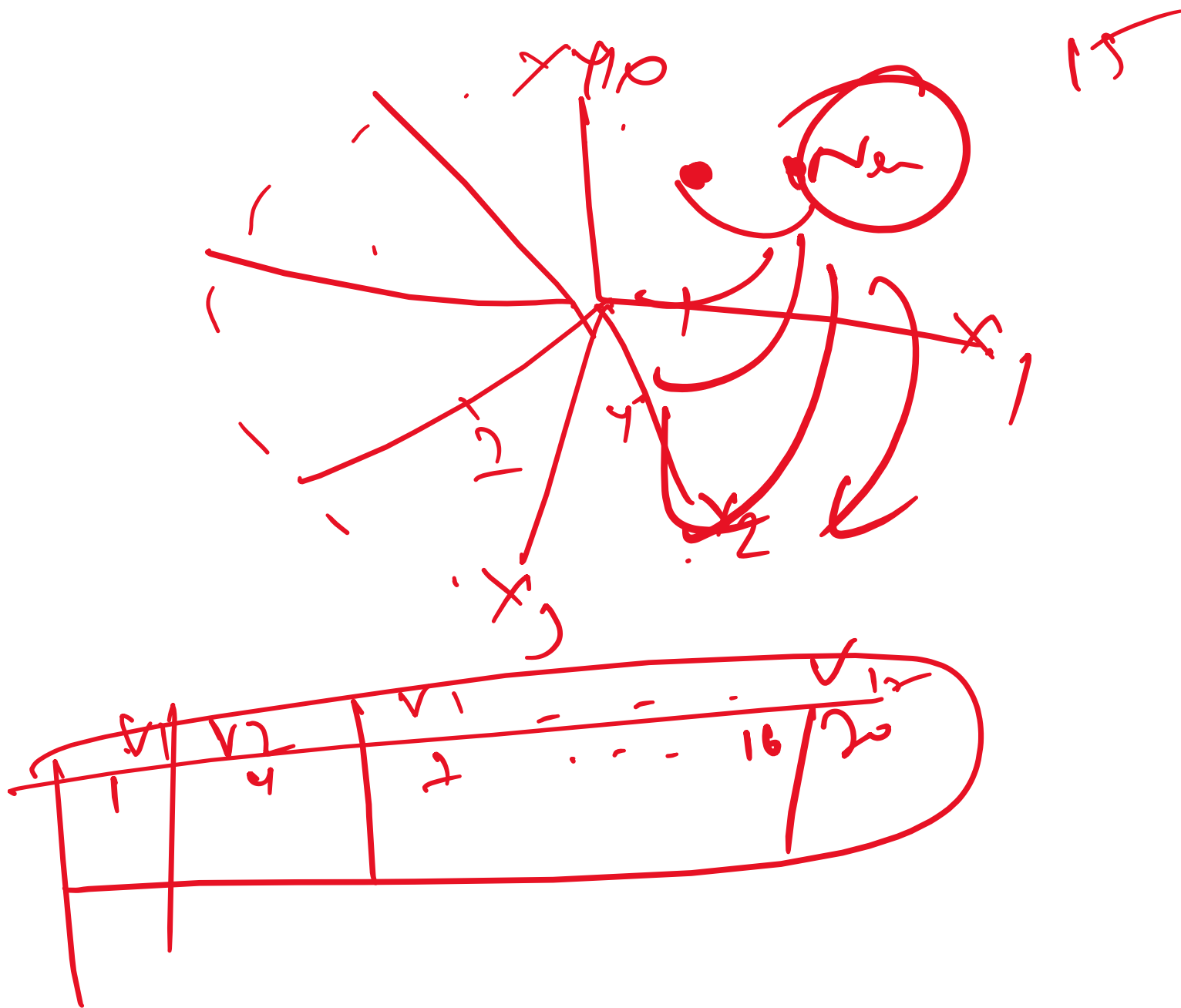
$f_2 = 2, 7, 1$
 $= 0, 1$

Binary

F1	F2	F3	F4	F5
1	0	1	0	1
1	1	0	1	1
0	1	0	1	1

KNN?





Count Vectorizer

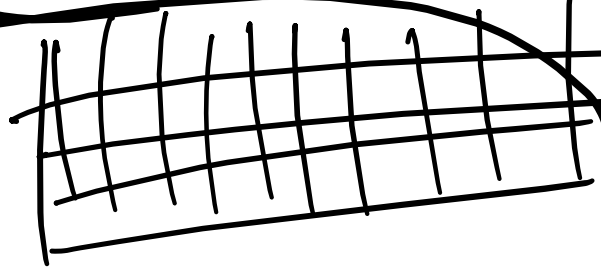
- 1 → Hello how are you
2 → Hi are you good
3 → What are you doing you

Matrix

	Hello	how	are	you	Hi	good	What
1	1	1	1	1	0	0	0
2	0	0	1	1	1	1	0
3	0	0	1	2	0	0	1

TF-IDF → 1000

Term Frequency



Inverse doc freq

NLP