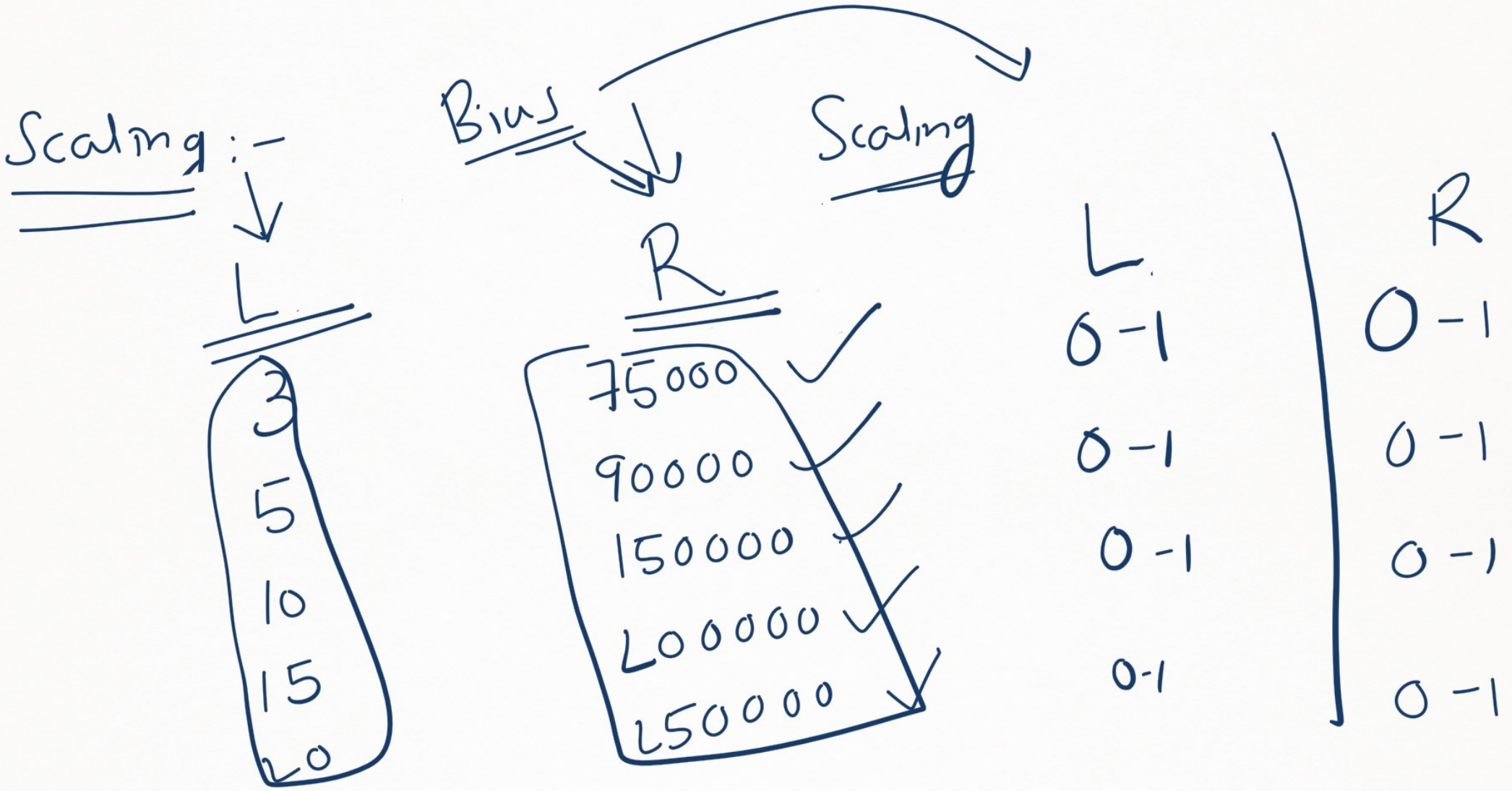
1. KNN Definition
2. K in KNN
3. Distance Metrics (ED, MD, MD')
4. Solving the problem
5. Scipy distance metrics
6. Implementation of KNN

Agenda:-

1. Template for ML Classification Algo.
2. Evaluation of Classification Algo
    a. Confusion Matrix
    b. Precision
    c. Recall
    d. f1 score
    e. Accuracy
3. KNN Regression Algorithm


Scaling :-
L
3
5
10
15
20
Bias
Scaling
R
75000
90000
150000
200000
250000
L
0-1
0-1
0-1
0-1
R
0-1
0-1
0-1
0-1

Confusion Matrix :- Evaluate a Classification model performance.

Eg :- Assume you have some symp of fever. You visit a doctor.

L possibilites

(1) Have fever
(2) No fever.

ML Algorithm

(1) Have fever
(2) No fever.

Actual :- Have Fever
No Fever

ML Prediction :- Have Fever
No Fever.

ML Prediction →	Actual (Doctor) (Reality) No	Actual (Doctor) (Reality) Yes
No	Doc :- No fever ML Algo :- No fever (True Negative)	Doc :- Has fever ML Algo :- No fever (False Negative)
Yes	Doc :- No fever ML Algo :- Has fever (False positive)	Doc :- Has fever ✓ ML Algo :- Has fever ✓ (True positive)

n = 100

Doctor

ML Algo		No	Yes
	No	56 (TN)	11
	yes	9	24 (TP)

24 = TP

= Doc → Have Fever

ML → Have Fever

56 = TN

⇒ Doc → No Fever

ML Algo → No Fever

9 = FP

$\Rightarrow$ Doc = No Fever

ML = Has Fever

Algo

11 = FN

$\Rightarrow$ Doc = Has Fever

ML = No Fever.

Algo

---

① TP & TN should have big nos.

② FP & FN should have small nos.

$$\begin{bmatrix} 64 & 12 \\ 4 & 20 \end{bmatrix}$$

Purchased
(Social Network Ads)

$TP = 20 \rightarrow R = Yes$ ML Algo. $\Rightarrow$ Yes

$TN = 64 \Rightarrow R = No$ ML Algo $\Rightarrow$ No

$FP = 4 \Rightarrow R = No$ ML Algo $\Rightarrow$ Yes

$FN = 12 \Rightarrow R = Yes$ ML Algo $\Rightarrow$ No

Precision :- What proportion of positive identifications were actually correct?

$$\text{Precision} :- \frac{TP}{TP+FP} = \frac{20}{20+4} = 83\%.$$

*: A model with '0' FP has a precision of 1

$$P = \frac{TP}{TP+0} = \frac{\cancel{TP}}{\cancel{TP}} = 1$$

Recall :- What proportion of actual positives were identified correctly

$$R = \frac{TP}{TP + FN} = \frac{20}{20+12} = \frac{20}{32} = 62.5\%$$

* A model with '0' FN has a recall of 1

$$R = \frac{TP}{TP+FN} = \frac{TP}{TP+0} = \frac{TP}{TP} = 1$$

Accuray :- $\frac{TP + TN}{TP + TN + FP + FN}$

$= \frac{20 + 64}{20 + 64 + 4 + 12} = 84\%$.

Error_rate $= \frac{FP + FN}{TP + TN + FP + FN} = \frac{4 + 12}{20 + 64 + 4 + 12}$

$= 16\%$.

$$f1\_score = \frac{2 * Precision * recall}{precision + recall.}$$

$$= \frac{2 * 0.83 * 0.625}{0.83 + 0.625}$$

$$= 71.4\%$$