

Binary Classification

0 ✓
[0.01
[0.05
[0.1
[0.4
[0.6
Probability
of 0

1 ✓
0.99 } $< 0.5 = 0$
0.95 } $> 0.5 = 1$
0.9
0.6
0.4
Probability
1

$\frac{N}{0} \rightarrow 0$ $\frac{Y}{1} \rightarrow 1$

K-NN

k - Nearest Neighbour

How many
Neighbours to
consider?
[elbow method]

Distance

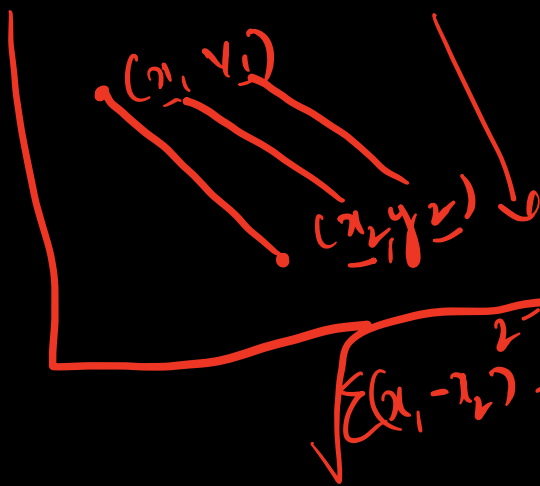
0 1
0 0
1 0

Hamming
distance

Euclidean

Manhattan

Minkowski



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

$$\sum (x_1 - x_2) + (y_1 - y_2)$$

Linear Reg \Rightarrow $y = mx + b$ ✓
 Logistic Reg \Rightarrow $y = mx + b$ ✓
 Predict
 Unseen Data

K-NN \Rightarrow No model
lazy model ✓
Huge Dataset
 Huge performance overhead
 distance
 select Neig
 mode, means
 to predict
 Unseen data

Large Samples (N): Large storage ✓
 row
 space complexity
 time complexity

Large column
 curse of dimensionality: Lower Accuracy



samples

50 | data record 50

1 |

1 |

k_2

100

10大

A hand-drawn graph on a grid showing a path from (1,2) to (9,5). The path is marked with green dots and arrows, passing through (2,3), (3,2), (4,6), (5,6), (8,9), and (9,5). The path is labeled with circled numbers 1, 2, and 3. The graph also shows other points like (1,2), (5,6), (8,9), and (9,5) labeled in red. The axes are labeled with numbers 1 through 9.

1	5	8	3	9
		↓		
1	3	5	8	9
		↑		

