

KNN - Examples

School Connect: Intro to DS & AI

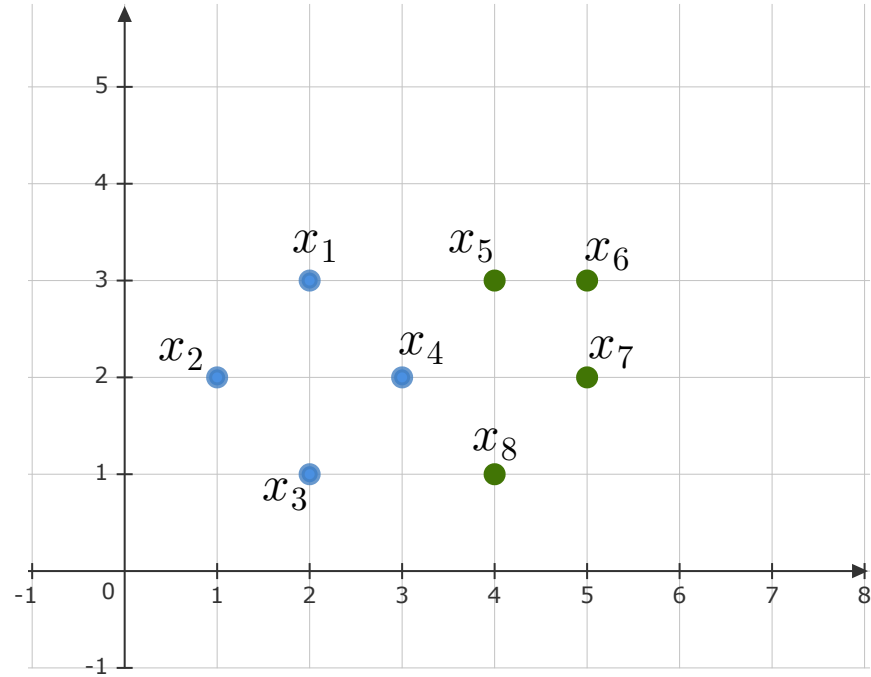
A Aniruddha

Indian Institute of Technology, Madras

Observe the dataset

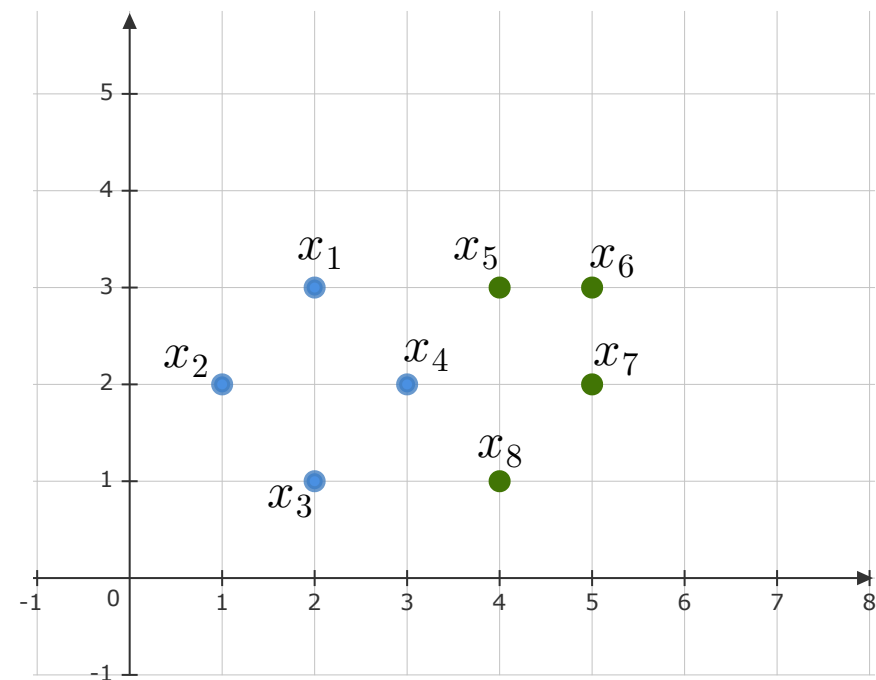
Observe the dataset

Consider the following dataset where blue points have class +1 and green points have class -1



Observe the dataset

Consider the following dataset where blue points have class +1 and green points have class -1



Point		Class
x_1	(2, 3)	+1
x_2	(1, 2)	+1
x_3	(2, 1)	+1
x_4	(3, 2)	+1
x_5	(4, 3)	-1
x_6	(5, 3)	-1
x_7	(5, 2)	-1
x_8	(4, 1)	-1

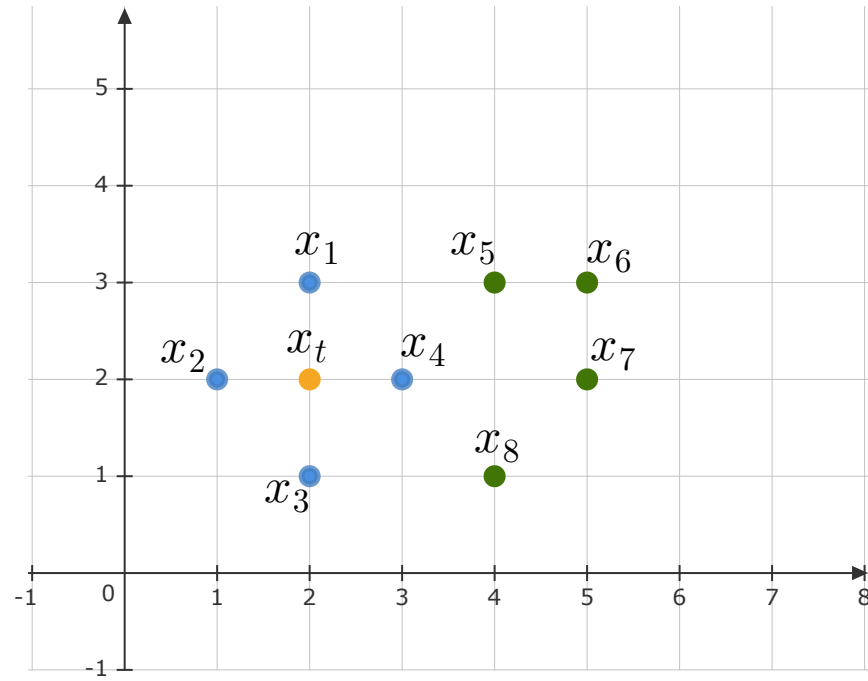
KNN - Procedure

KNN - Procedure

Consider a test point $x_t = (2,2)$ and assign a label for different values of K

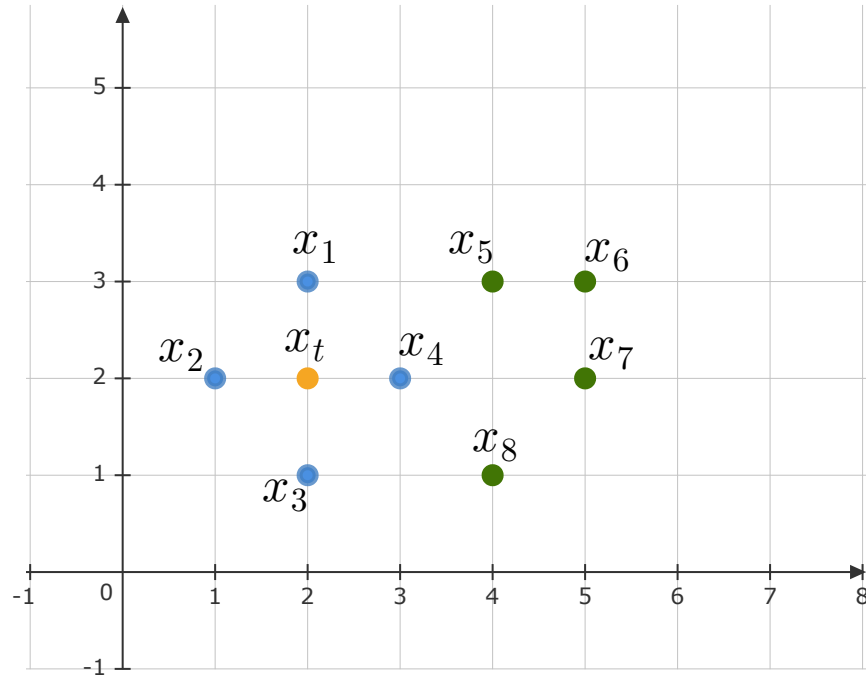
KNN - Procedure

Consider a test point $x_t = (2,2)$ and assign a label for different values of K



KNN - Procedure

Consider a test point $x_t = (2,2)$ and assign a label for different values of K

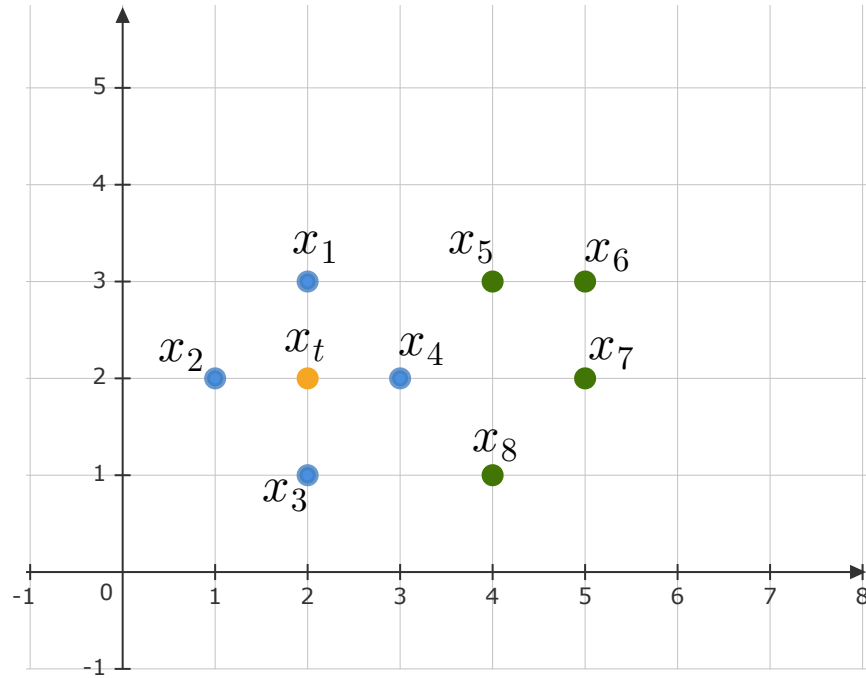


To assign a label to a test point,

1. Compute its distance from every other point in the dataset
2. Depending on the value of K, we choose the K closest points and assign the label corresponding to the majority of the points

KNN - Procedure

Consider a test point $x_t = (2,2)$ and assign a label for different values of K



To assign a label to a test point,

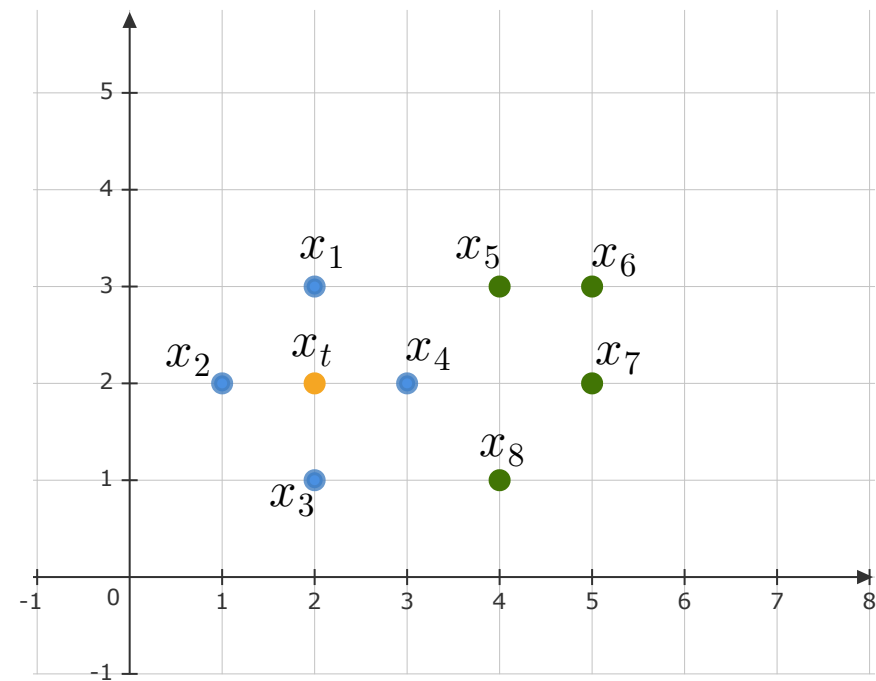
1. Compute its distance from every other point in the dataset
2. Depending on the value of K, we choose the K closest points and assign the label corresponding to the majority of the points

The distance of the point x_t from x_1 is given by,

$$\begin{aligned}\text{Distance} &= \sqrt{(2-2)^2 + (2-3)^2} \\ &= 1\end{aligned}$$

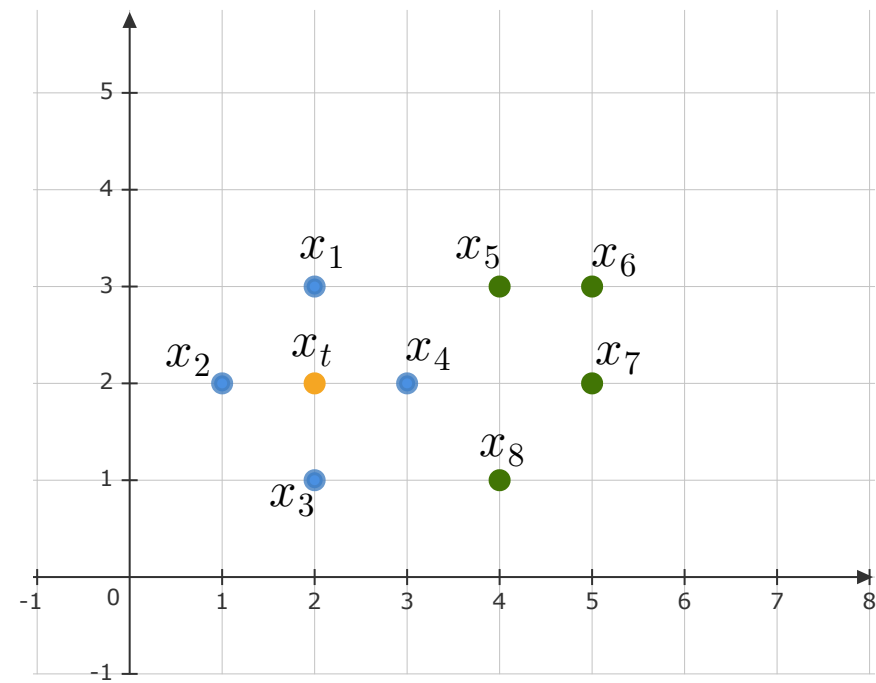
Distance of test point from training data

Distance of test point from training data



Distance of test point from training data

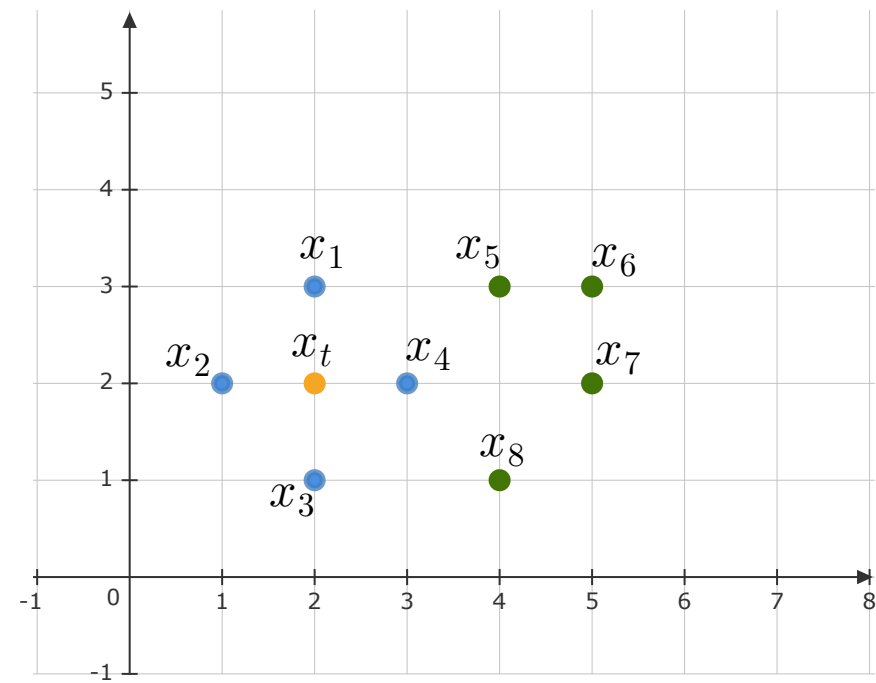
The distance of the test data point from each training data point is,



Point		Class	Distance
x_1	(2, 3)	+1	1
x_2	(1, 2)	+1	1
x_3	(2, 1)	+1	1
x_4	(3, 2)	+1	1
x_5	(4, 3)	-1	$\sqrt{5}$
x_6	(5, 3)	-1	$\sqrt{10}$
x_7	(5, 2)	-1	3
x_8	(4, 1)	-1	$\sqrt{5}$

Distance of test point from training data

The distance of the test data point from each training data point is,

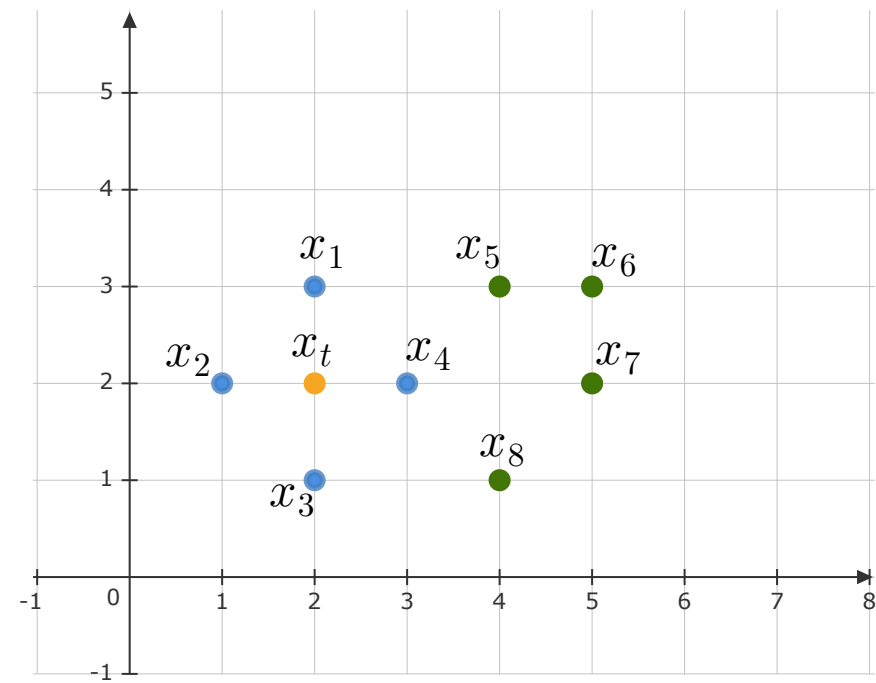


We now consider different values of K and assign the label accordingly

Point		Class	Distance
x_1	(2, 3)	+1	1
x_2	(1, 2)	+1	1
x_3	(2, 1)	+1	1
x_4	(3, 2)	+1	1
x_5	(4, 3)	-1	$\sqrt{5}$
x_6	(5, 3)	-1	$\sqrt{10}$
x_7	(5, 2)	-1	3
x_8	(4, 1)	-1	$\sqrt{5}$

Distance of test point from training data

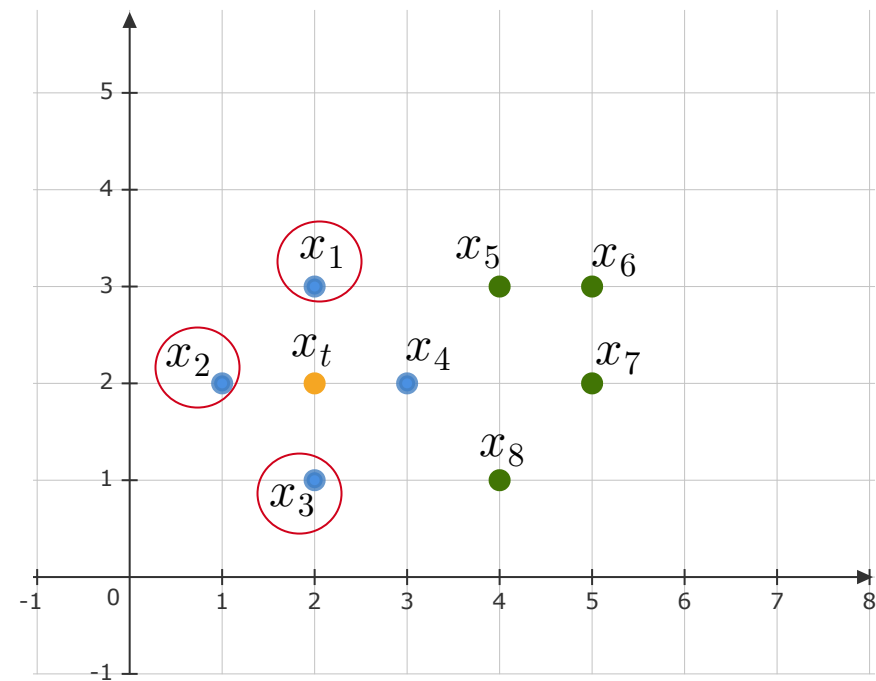
For $K = 3$, we take the three closest points and assign the class corresponding to the majority



Point		Class	Distance
x_1	(2, 3)	+1	1
x_2	(1, 2)	+1	1
x_3	(2, 1)	+1	1
x_4	(3, 2)	+1	1
x_5	(4, 3)	-1	$\sqrt{5}$
x_6	(5, 3)	-1	$\sqrt{10}$
x_7	(5, 2)	-1	3
x_8	(4, 1)	-1	$\sqrt{5}$

Label Assignment

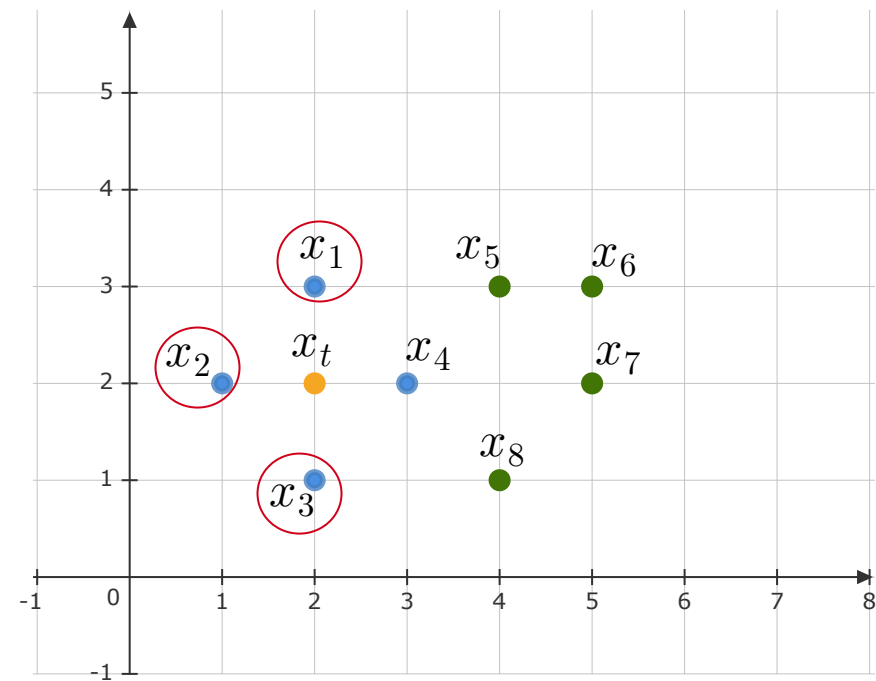
For $K = 3$, we take the three closest points and assign the class corresponding to the majority



Point		Class	Distance
x_1	(2, 3)	+1	1
x_2	(1, 2)	+1	1
x_3	(2, 1)	+1	1
x_4	(3, 2)	+1	1
x_5	(4, 3)	-1	$\sqrt{5}$
x_6	(5, 3)	-1	$\sqrt{10}$
x_7	(5, 2)	-1	3
x_8	(4, 1)	-1	$\sqrt{5}$

Label Assignment

For $K = 3$, we take the three closest points and assign the class corresponding to the majority

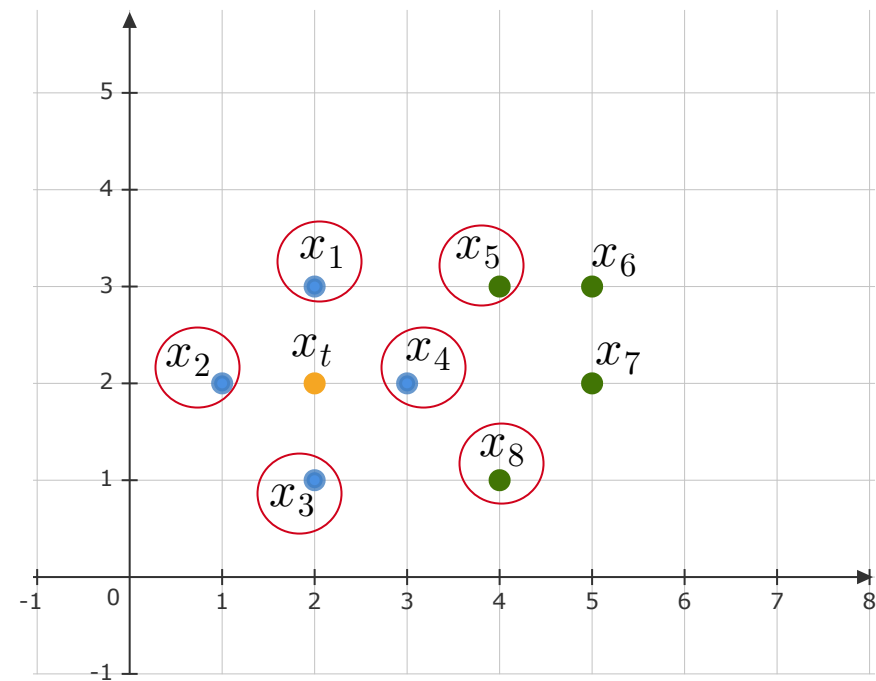


Here, we see that the three closest points have a label of +1 and so we will assign the test point with the same label

Point		Class	Distance
x_1	(2, 3)	+1	1
x_2	(1, 2)	+1	1
x_3	(2, 1)	+1	1
x_4	(3, 2)	+1	1
x_5	(4, 3)	-1	$\sqrt{5}$
x_6	(5, 3)	-1	$\sqrt{10}$
x_7	(5, 2)	-1	3
x_8	(4, 1)	-1	$\sqrt{5}$

Label Assignment

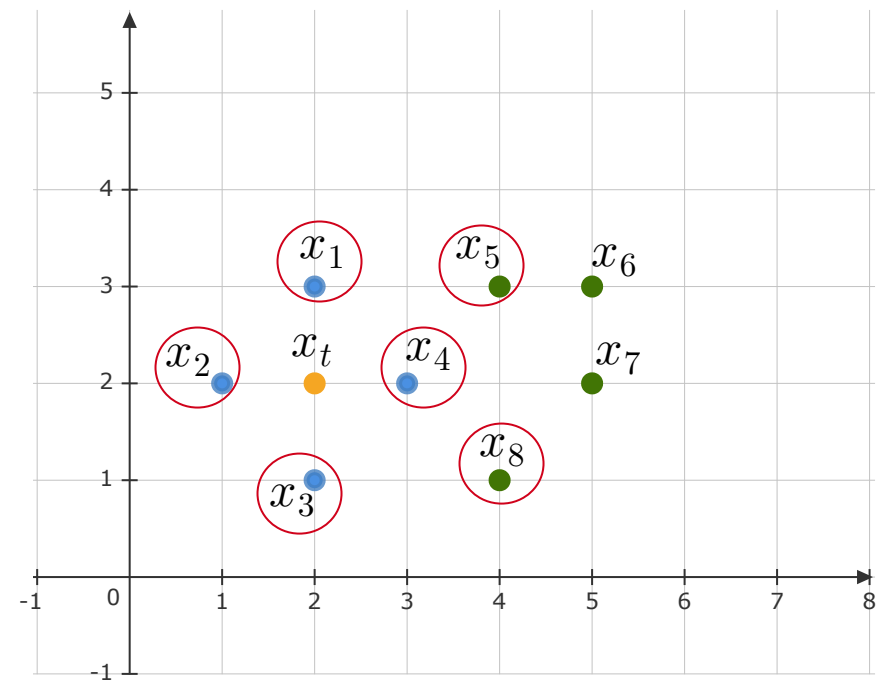
For $K = 6$, we take the six closest points and assign the class corresponding to the majority



Point		Class	Distance
x_1	(2, 3)	+1	1
x_2	(1, 2)	+1	1
x_3	(2, 1)	+1	1
x_4	(3, 2)	+1	1
x_5	(4, 3)	-1	$\sqrt{5}$
x_6	(5, 3)	-1	$\sqrt{10}$
x_7	(5, 2)	-1	3
x_8	(4, 1)	-1	$\sqrt{5}$

Label Assignment

For $K = 6$, we take the six closest points and assign the class corresponding to the majority

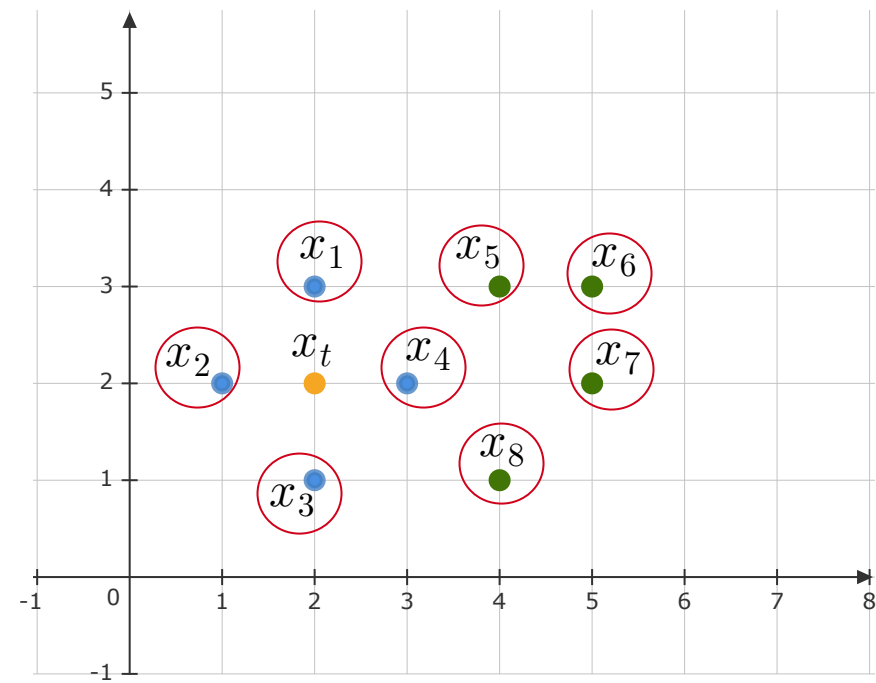


Here, we see that four points have a label of +1 and two points have a label of -1. Since the majority has label +1, we assign the same to x_t

Point		Class	Distance
x_1	(2, 3)	+1	1
x_2	(1, 2)	+1	1
x_3	(2, 1)	+1	1
x_4	(3, 2)	+1	1
x_5	(4, 3)	-1	$\sqrt{5}$
x_6	(5, 3)	-1	$\sqrt{10}$
x_7	(5, 2)	-1	3
x_8	(4, 1)	-1	$\sqrt{5}$

Label Assignment

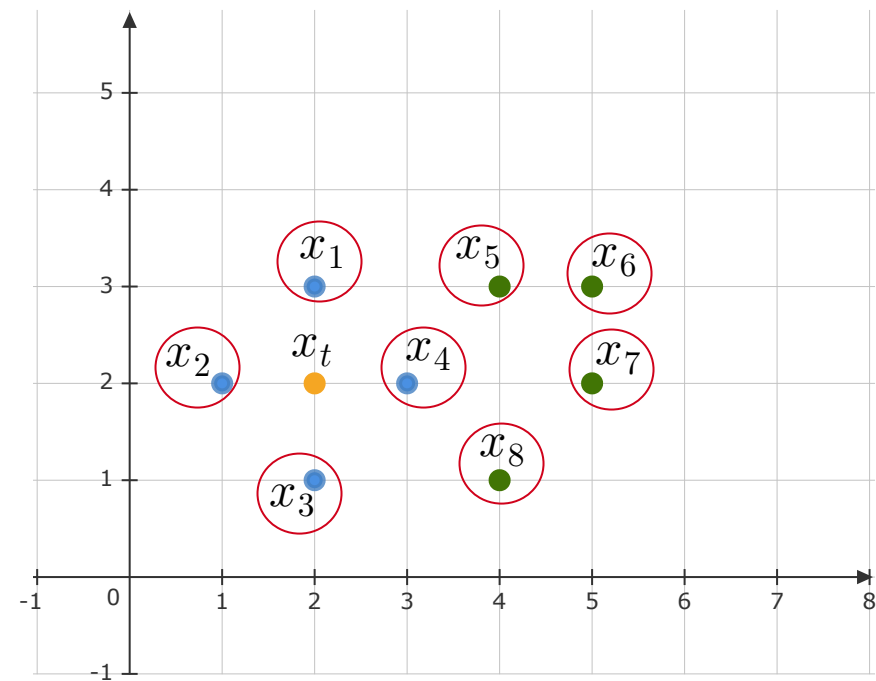
For $K = 8$, we take all the points and assign the class corresponding to the majority



Point		Class	Distance
x_1	$(2, 3)$	$+1$	1
x_2	$(1, 2)$	$+1$	1
x_3	$(2, 1)$	$+1$	1
x_4	$(3, 2)$	$+1$	1
x_5	$(4, 3)$	-1	$\sqrt{5}$
x_6	$(5, 3)$	-1	$\sqrt{10}$
x_7	$(5, 2)$	-1	3
x_8	$(4, 1)$	-1	$\sqrt{5}$

Label Assignment

For $K = 8$, we take all the points and assign the class corresponding to the majority

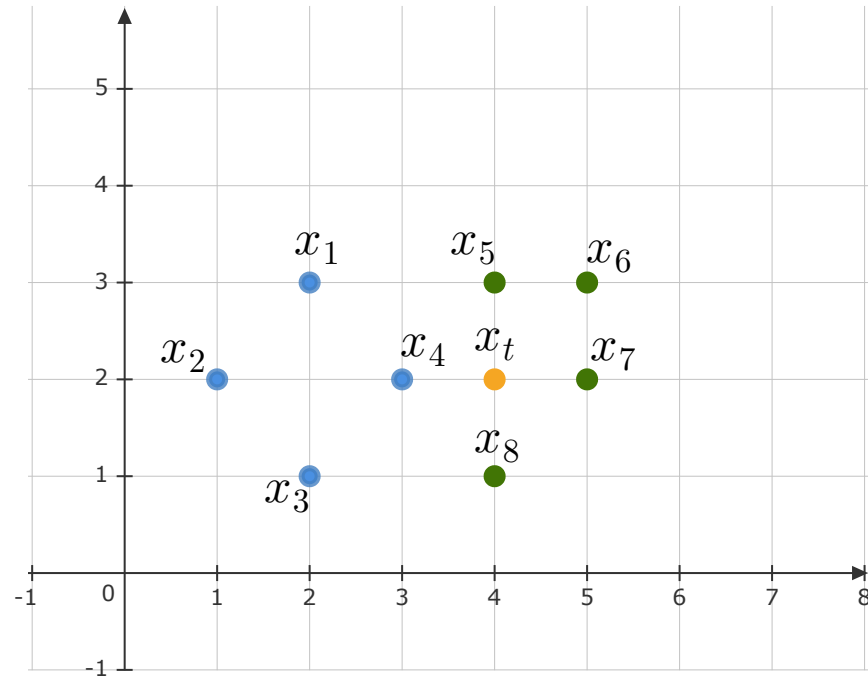


Here, we see that four points have a label of +1 and four points have a label of -1. Since there is no clear majority, we can assign any label

Point		Class	Distance
x_1	(2, 3)	+1	1
x_2	(1, 2)	+1	1
x_3	(2, 1)	+1	1
x_4	(3, 2)	+1	1
x_5	(4, 3)	-1	$\sqrt{5}$
x_6	(5, 3)	-1	$\sqrt{10}$
x_7	(5, 2)	-1	3
x_8	(4, 1)	-1	$\sqrt{5}$

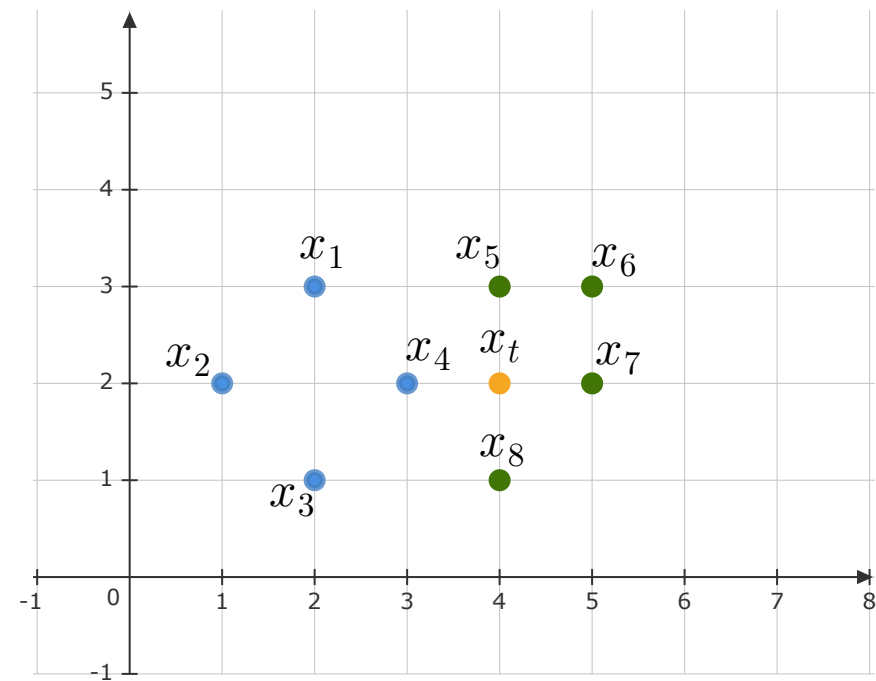
KNN - Procedure

Consider a test point $x_t = (4, 2)$ and assign a label for different values of K



Distance of test point from training data

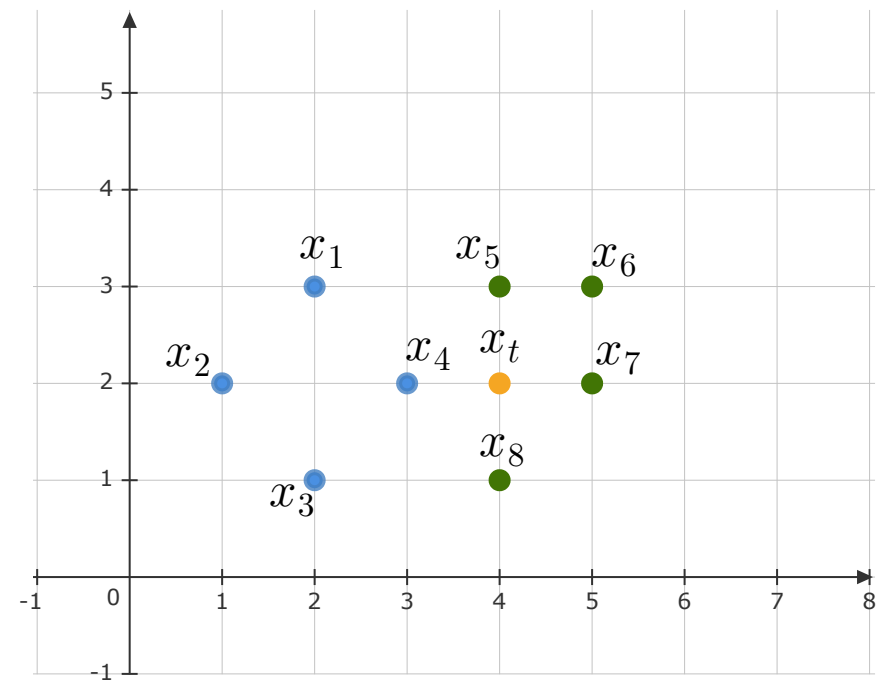
The distance of the test data point from each training data point is,



Point		Class	Distance
x_1	$(2, 3)$	+1	$\sqrt{5}$
x_2	$(1, 2)$	+1	3
x_3	$(2, 1)$	+1	$\sqrt{5}$
x_4	$(3, 2)$	+1	1
x_5	$(4, 3)$	-1	1
x_6	$(5, 3)$	-1	$\sqrt{2}$
x_7	$(5, 2)$	-1	1
x_8	$(4, 1)$	-1	1

Distance of test point from training data

The distance of the test data point from each training data point is,

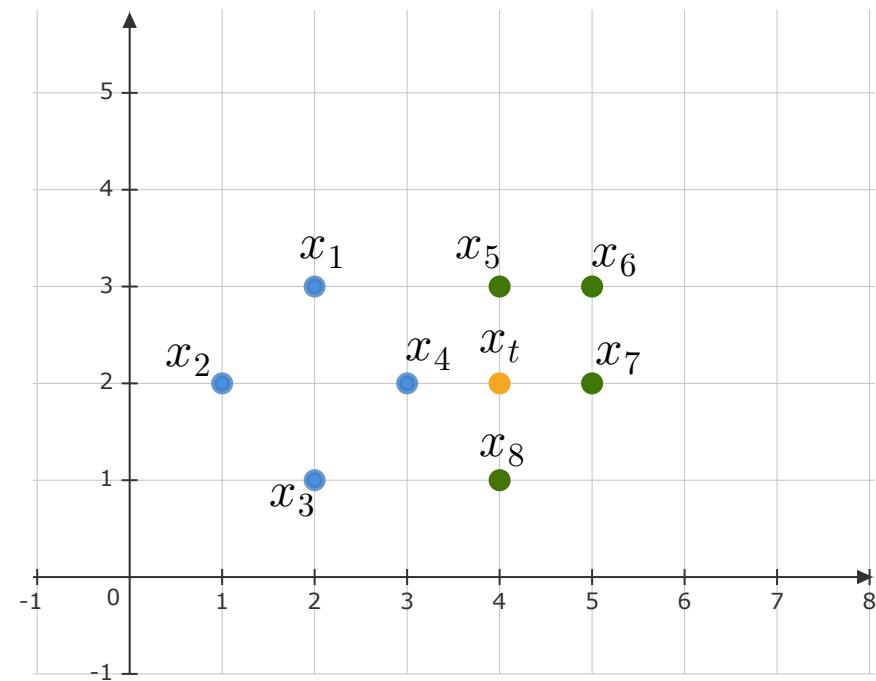


Point		Class	Distance
x_1	(2, 3)	+1	$\sqrt{5}$
x_2	(1, 2)	+1	3
x_3	(2, 1)	+1	$\sqrt{5}$
x_4	(3, 2)	+1	1
x_5	(4, 3)	-1	1
x_6	(5, 3)	-1	$\sqrt{2}$
x_7	(5, 2)	-1	1
x_8	(4, 1)	-1	1

We now consider different values of K and assign the label accordingly

Distance of test point from training data

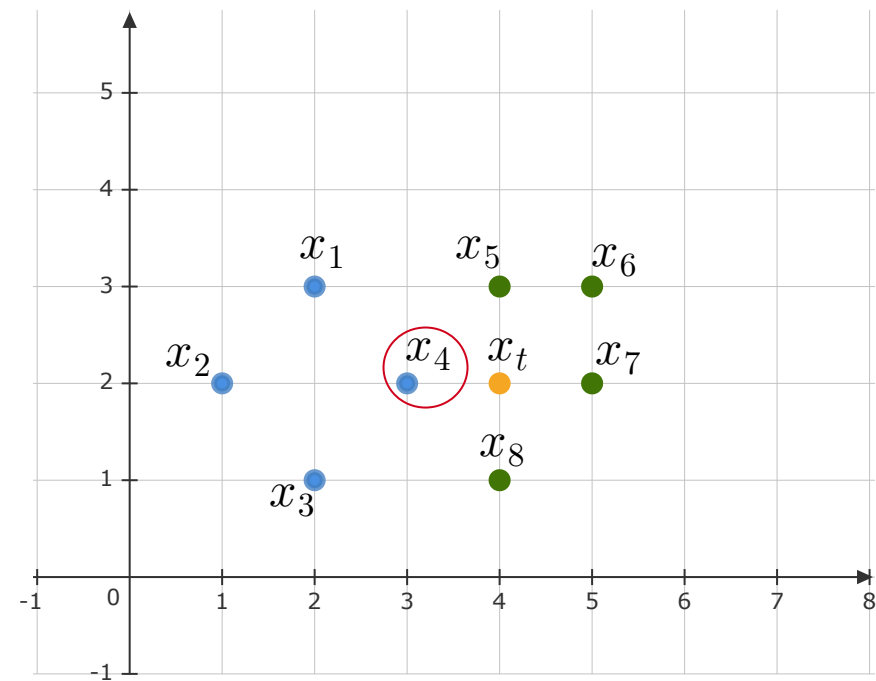
For $K = 1$, we take one of the closest points and assign its class to the test point



Point		Class	Distance
x_1	(2, 3)	+1	$\sqrt{5}$
x_2	(1, 2)	+1	3
x_3	(2, 1)	+1	$\sqrt{5}$
x_4	(3, 2)	+1	1
x_5	(4, 3)	-1	1
x_6	(5, 3)	-1	$\sqrt{2}$
x_7	(5, 2)	-1	1
x_8	(4, 1)	-1	1

Label Assignment

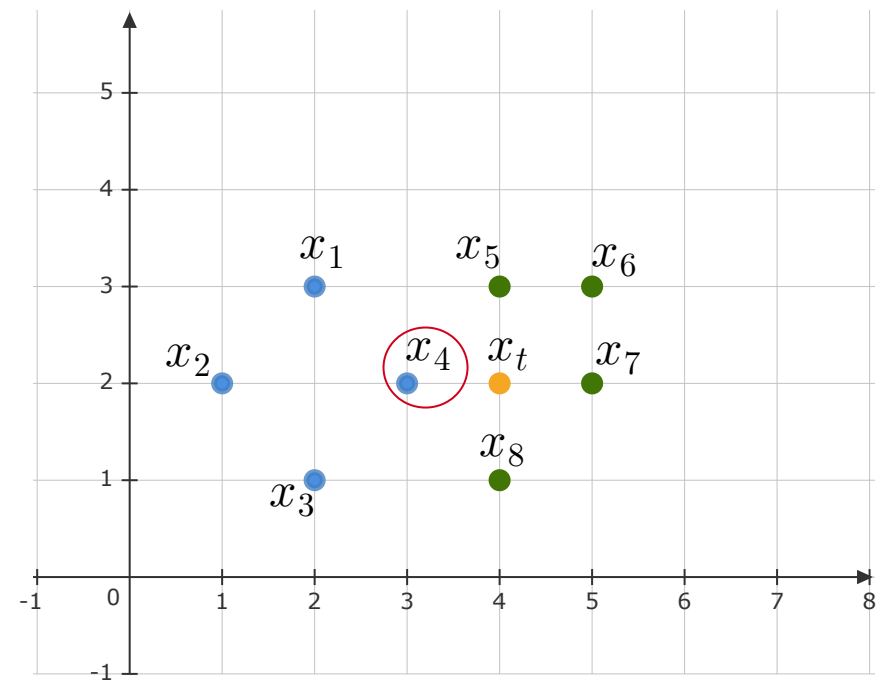
For $K = 1$, we take one of the closest points and assign its class to the test point



Point		Class	Distance
x_1	(2, 3)	+1	$\sqrt{5}$
x_2	(1, 2)	+1	3
x_3	(2, 1)	+1	$\sqrt{5}$
x_4	(3, 2)	+1	1
x_5	(4, 3)	-1	1
x_6	(5, 3)	-1	$\sqrt{2}$
x_7	(5, 2)	-1	1
x_8	(4, 1)	-1	1

Label Assignment

For $K = 1$, we take one of the closest points and assign its class to the test point

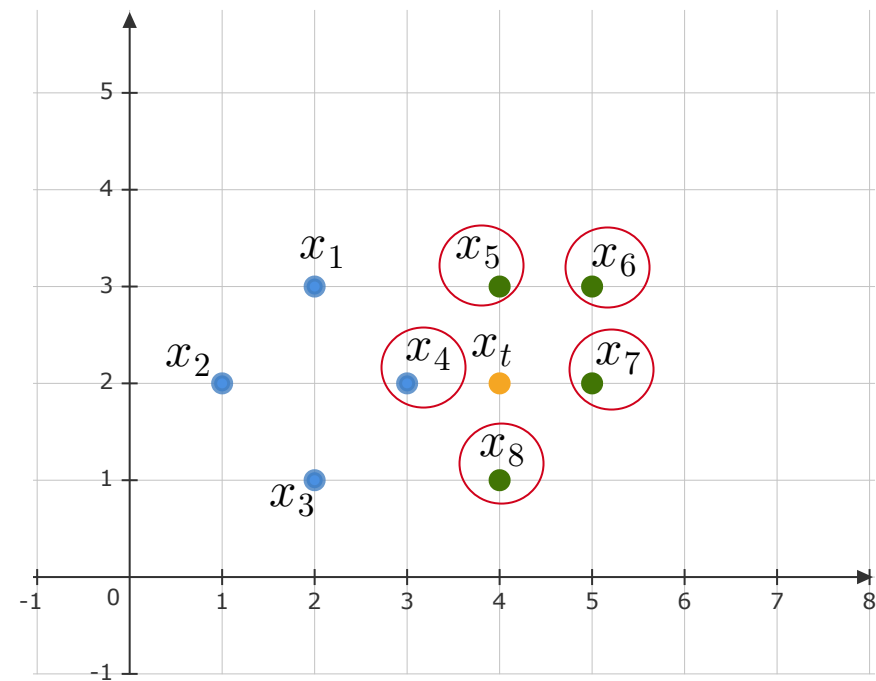


Here, we see that the point x_4 has a label of +1 and so we assign the same label to the test point

Point		Class	Distance
x_1	$(2, 3)$	+1	$\sqrt{5}$
x_2	$(1, 2)$	+1	3
x_3	$(2, 1)$	+1	$\sqrt{5}$
x_4	$(3, 2)$	+1	1
x_5	$(4, 3)$	-1	1
x_6	$(5, 3)$	-1	$\sqrt{2}$
x_7	$(5, 2)$	-1	1
x_8	$(4, 1)$	-1	1

Label Assignment

For $K = 5$, we take the five closest points and assign the class corresponding to the majority



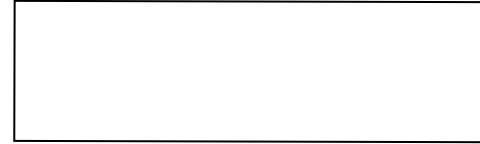
Point		Class	Distance
x_1	(2, 3)	+1	$\sqrt{5}$
x_2	(1, 2)	+1	3
x_3	(2, 1)	+1	$\sqrt{5}$
x_4	(3, 2)	+1	1
x_5	(4, 3)	-1	1
x_6	(5, 3)	-1	$\sqrt{2}$
x_7	(5, 2)	-1	1
x_8	(4, 1)	-1	1

Train set and Test set

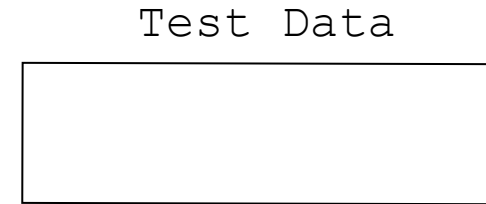
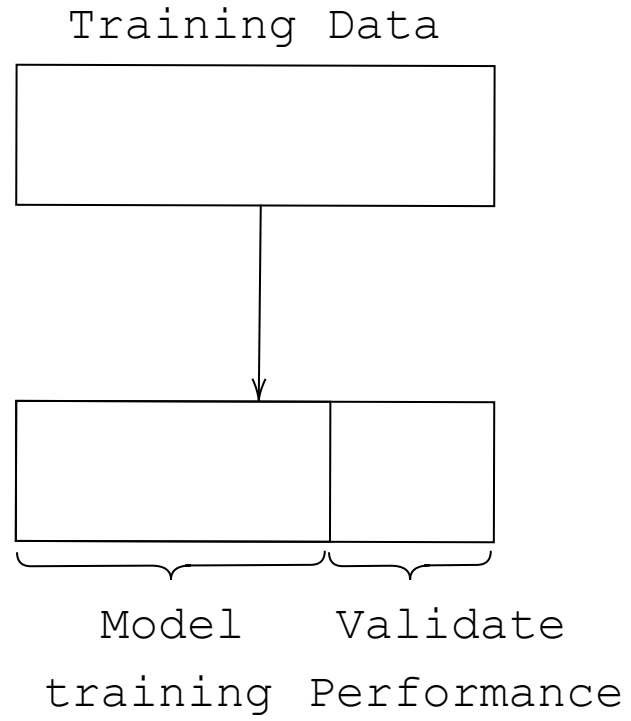
Training Data



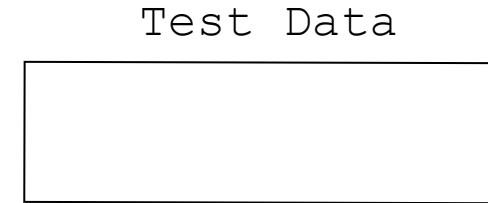
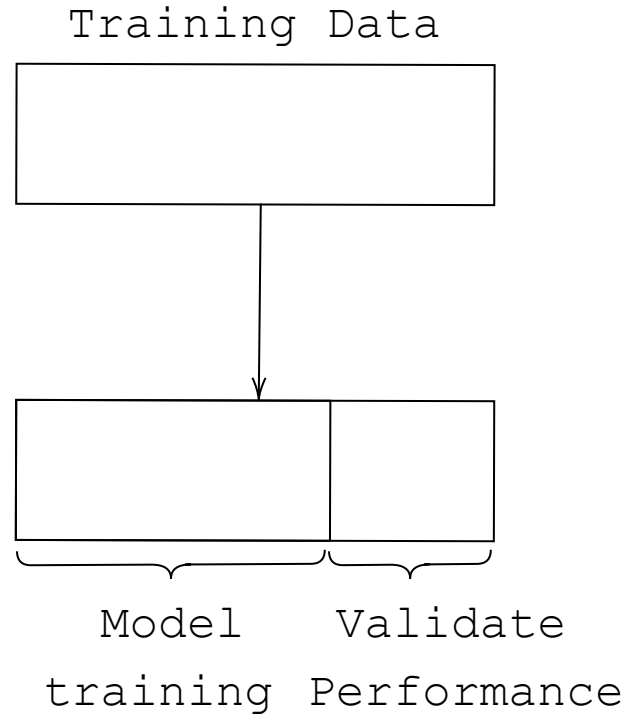
Test Data



Train set and Test set



Train set and Test set



Try out different
models and use the
best one on Test data

Thank you

