



**DAWSON**  
C O L L E G E

K-Nearest Neighbors (KNN)

Garry Ka Lok CHU

# Artificial Intelligence

## Module: K-Nearest

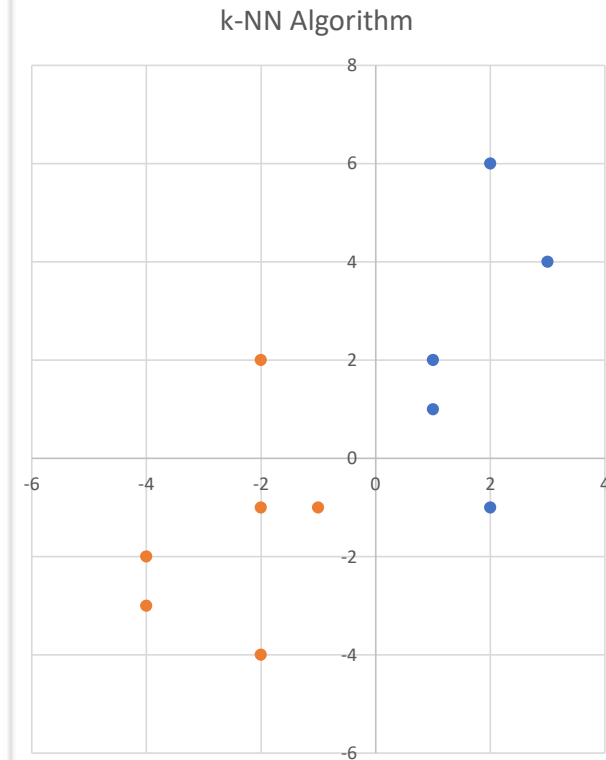
### Neighbors (KNN)

Classification is one of the tasks performed by artificial intelligence. Given a set of training data with values accompanied by known labels, the system can classify any new data using the algorithm of KNN.

KNN								k =	3	Test point	0	2
x	y	Category	Distance <sup>2</sup>	Distance	Distance	Rank	NN?					
1	2	A									A total count	
2	-1	A										
2	6	A										
1	1	A									B total count	
3	4	A										
-2	-4	B										
-4	-2	B										
-2	-1	B										
-1	-1	B										
-2	2	B										
-4	-3	B										

In this example, 11 points are included in the dataset. The objective is to determine which category (0, 2) belongs to.

First, determine the value of k. In this case, k = 3 is chosen.

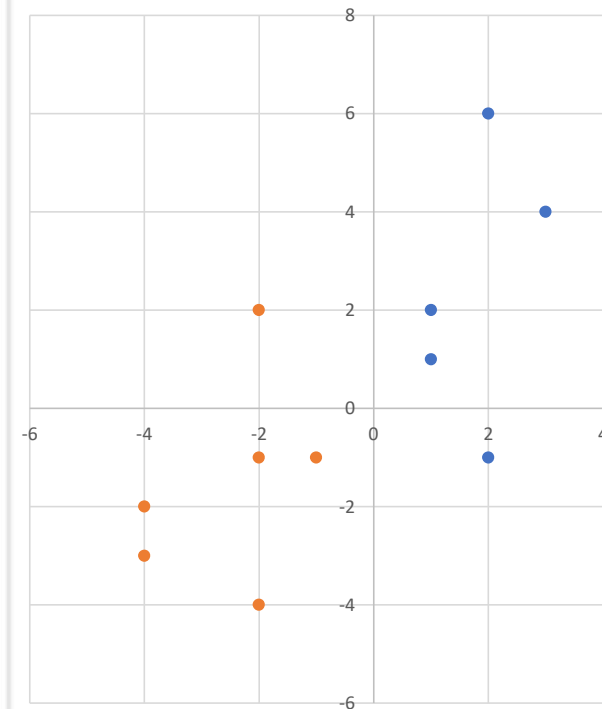


KNN				k =	3			Test point	0	2
x	y	Category	Distance^2	Distance	Distance	Rank	NN?			
1	2	A	1	1	1.0000				A total count	
2	-1	A	13	$\sqrt{13}$	3.6056					
2	6	A	20	$2\sqrt{5}$	4.4721					
1	1	A	2	$\sqrt{2}$	1.4142				B total count	
3	4	A	13	$\sqrt{13}$	3.6056					
-2	-4	B	40	$2\sqrt{10}$	6.3246					
-4	-2	B	32	$4\sqrt{2}$	5.6569					
-2	-1	B	13	$\sqrt{13}$	3.6056					
-1	-1	B	10	$\sqrt{10}$	3.1623					
-2	2	B	4	2	2.0000					
-4	-3	B	41	$\sqrt{41}$	6.4031					

Calculate the distance between the test point and each point in the training dataset.

$$distance = \sqrt{(0 - 2)^2 + (2 - 6)^2} = \sqrt{20} = 2\sqrt{5}$$

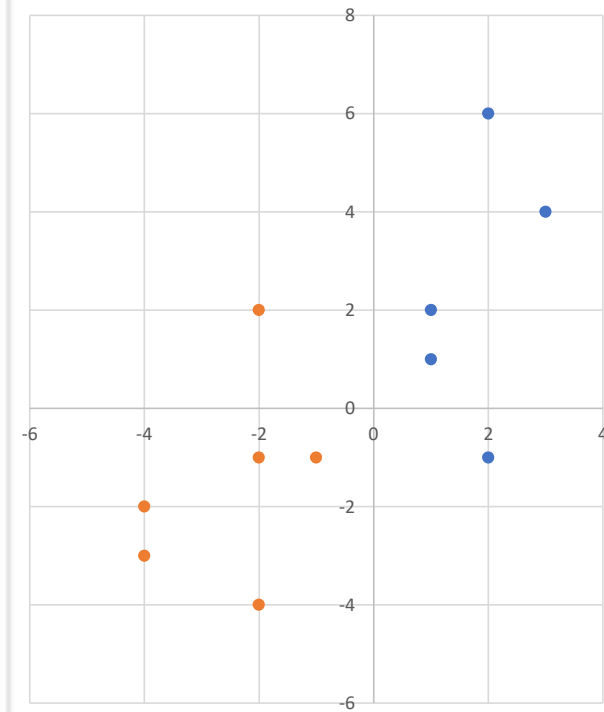
k-NN Algorithm



KNN								k =	3	Test point	0	2
x	y	Category	Distance^2	Distance	Distance	Rank	NN?					
1	2	A	1	1	1.0000	1						
2	-1	A	13	$\sqrt{13}$	3.6056	5					A total count	
2	6	A	20	$2\sqrt{5}$	4.4721	8						
1	1	A	2	$\sqrt{2}$	1.4142	2						B total count
3	4	A	13	$\sqrt{13}$	3.6056	5						
-2	-4	B	40	$2\sqrt{10}$	6.3246	10						
-4	-2	B	32	$4\sqrt{2}$	5.6569	9						
-2	-1	B	13	$\sqrt{13}$	3.6056	5						
-1	-1	B	10	$\sqrt{10}$	3.1623	4						
-2	2	B	4	2	2.0000	3						
-4	-3	B	41	$\sqrt{41}$	6.4031	11						

Rank the distances in increasing order.

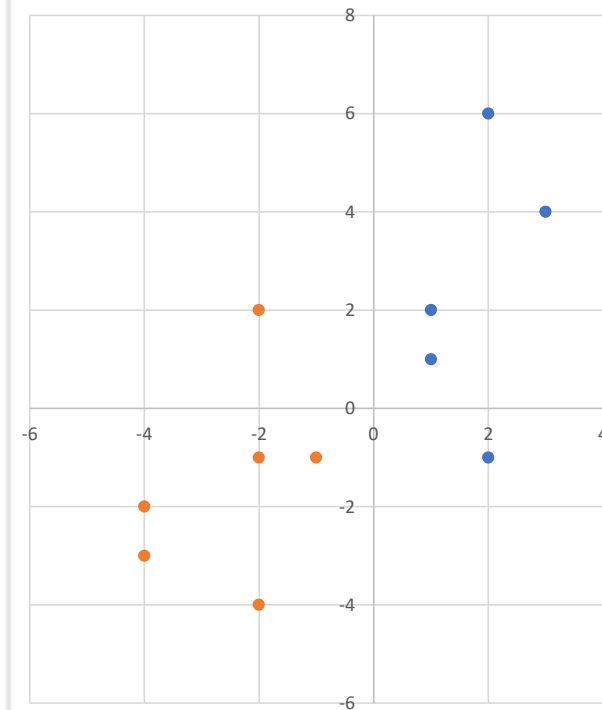
k-NN Algorithm



KNN				k =	3			Test point	0	2
x	y	Category	Distance^2	Distance	Distance	Rank	NN?			
1	2	A	1	1	1.0000	1	Y		A total count	
2	-1	A	13	$\sqrt{13}$	3.6056	5	N			
2	6	A	20	$2\sqrt{5}$	4.4721	8	N			
1	1	A	2	$\sqrt{2}$	1.4142	2	Y		B total count	
3	4	A	13	$\sqrt{13}$	3.6056	5	N			
-2	-4	B	40	$2\sqrt{10}$	6.3246	10	N			
-4	-2	B	32	$4\sqrt{2}$	5.6569	9	N			
-2	-1	B	13	$\sqrt{13}$	3.6056	5	N			
-1	-1	B	10	$\sqrt{10}$	3.1623	4	N			
-2	2	B	4	2	2.0000	3	Y			
-4	-3	B	41	$\sqrt{41}$	6.4031	11	N			

Determine K-nearest neighbors based on k minimum distances.

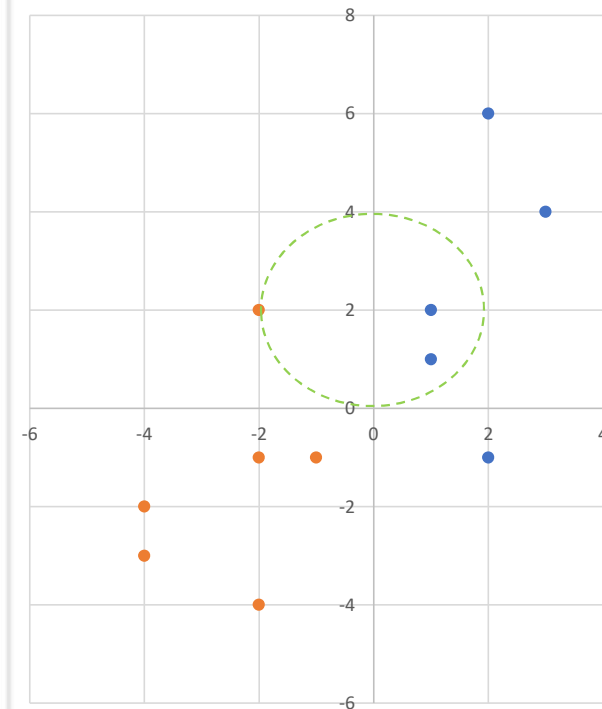
k-NN Algorithm



KNN								k =	3	Test point	0	2
x	y	Category	Distance^2	Distance	Distance	Rank	NN?					
1	2	A	1	1	1.0000	1	Y			A total count	2	
2	-1	A	13	$\sqrt{13}$	3.6056	5	N					
2	6	A	20	$2\sqrt{5}$	4.4721	8	N					
1	1	A	2	$\sqrt{2}$	1.4142	2	Y			B total count	1	
3	4	A	13	$\sqrt{13}$	3.6056	5	N					
-2	-4	B	40	$2\sqrt{10}$	6.3246	10	N					
-4	-2	B	32	$4\sqrt{2}$	5.6569	9	N					
-2	-1	B	13	$\sqrt{13}$	3.6056	5	N					
-1	-1	B	10	$\sqrt{10}$	3.1623	4	N					
-2	2	B	4	2	2.0000	3	Y					
-4	-3	B	41	$\sqrt{41}$	6.4031	11	N					

Use the known labels to determine which category the test point belongs based on simple majority.

k-NN Algorithm

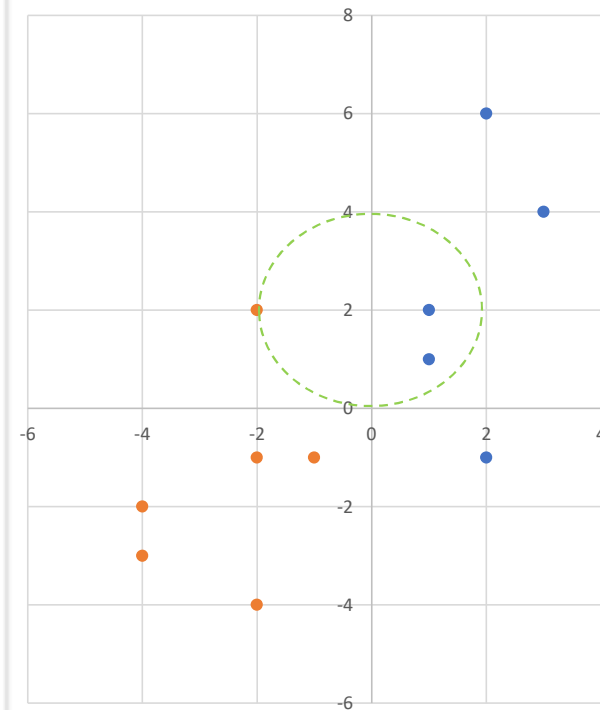


KNN				k =	3			Test point	0	2
x	y	Category	Distance^2	Distance	Distance	Rank	NN?			
1	2	A	1	1	1.0000	1	Y		A total count	2
2	-1	A	13	$\sqrt{13}$	3.6056	5	N			
2	6	A	20	$2\sqrt{5}$	4.4721	8	N			
1	1	A	2	$\sqrt{2}$	1.4142	2	Y		B total count	1
3	4	A	13	$\sqrt{13}$	3.6056	5	N			
-2	-4	B	40	$2\sqrt{10}$	6.3246	10	N			
-4	-2	B	32	$4\sqrt{2}$	5.6569	9	N			
-2	-1	B	13	$\sqrt{13}$	3.6056	5	N			
-1	-1	B	10	$\sqrt{10}$	3.1623	4	N			
-2	2	B	4	2	2.0000	3	Y			
-4	-3	B	41	$\sqrt{41}$	6.4031	11	N			

**Conclusion:**

The test point belongs to Category A.

k-NN Algorithm



## Applications:

- Data Compression
- Finance and economics
- Health Science such as Genetics



## Advanced Questions:

- Although KNN is a simple algorithm to apply, which value of  $k$  should be chosen as the optimal value?
- Other than Euclidean distance, may we use other type of distance?

## Further Readings:

- *Data Breast Cancer Prediction with K-Nearest Neighbor Algorithm using Different Distance Measurements (2018) by Victoria Rodriquez, Karan Sharma & Dana Walker*