K Nearest Neighbor

Cesar Acosta Ph.D.

Department of Industrial and Systems Engineering
University of Southern California

K-Nearest Neighbors

K-Nearest neighbors for classification predicts the category using the k-closest observations

K-Nearest Neighbors – Example (K=1)

.

Y	X
0	13
1	32
0	19
0	14
1	28
1	22

K-Nearest Neighbors – Example (K=1)

.

Y	X		distance
0	13		
1	32		19
0	19		6
0	14		1
1	28		15
1	22		9

K-Nearest Neighbors – Example (K=1)

.

Y	X	Yhat	distance	
0	13			
1	32		19	
0	19		6	
0	14	0	1	min
1	28		15	
1	22		9	

K-Nearest Neighbors – Example (K=1)

.

Y	X		distance
0	13		19
1	32		
0	19		13
0	14		18
1	28		4
1	22		10

K-Nearest Neighbors – Example (K=1)

.

Y	X	Yhat	distance	
0	13		19	
1	32			
0	19		13	
0	14	0	18	
1	28	1	4	min
1	22		10	

K-Nearest Neighbors – Example (K=1)

.

Y	X1	X2
0	13	29
1	32	8
0	19	21
0	12	33
1	28	14
1	22	12

K-Nearest Neighbors – Example (K=1)

.

Y	X1	X2		distance
0	13	29		
1	32	8		28.32
0	19	21		10.00
0	12	33		4.12
1	28	14		21.21
1	22	12		19.24

K-Nearest Neighbors – Example (K=1)

.

Y	X1	X2		distance		
0	13	29				
1	32	8		28.32		(13,29) to (32,8)
0	19	21		10.00		(13,29) to (19,21)
0	12	33		4.12		(13,29) to (12,33)
1	28	14		21.21		(13,29) to (28,14)
1	22	12		19.24		(13,29) to (22,12)

K-Nearest Neighbors – Example (K=1)

.

Y	X1	X2	Yhat	distance		
0	13	29				
1	32	8		28.32		(13,29) to (32,8)
0	19	21		10.00		(13,29) to (19,21)
0	12	33	0	4.12	min	(13,29) to (12,33)
1	28	14		21.21		(13,29) to (28,14)
1	22	12		19.24		(13,29) to (22,12)

K-Nearest Neighbors – Example (K=1)

.

Y	X1	X2		distance
0	13	29		28.32
1	32	8		
0	19	21		18.38
0	12	33		32.02
1	28	14		7.21
1	22	12		10.77

K-Nearest Neighbors – Example (K=1)

.

Y	X1	X2		distance	
0	13	29		28.32	(32,8) to (13,29)
1	32	8			
0	19	21		18.38	(32,8) to (19,21)
0	12	33		32.02	(32,8) to (12,33)
1	28	14		7.21	(32,8) to (28,14)
1	22	12		10.77	(32,8) to (22,12)

K-Nearest Neighbors – Example (K=1)

.

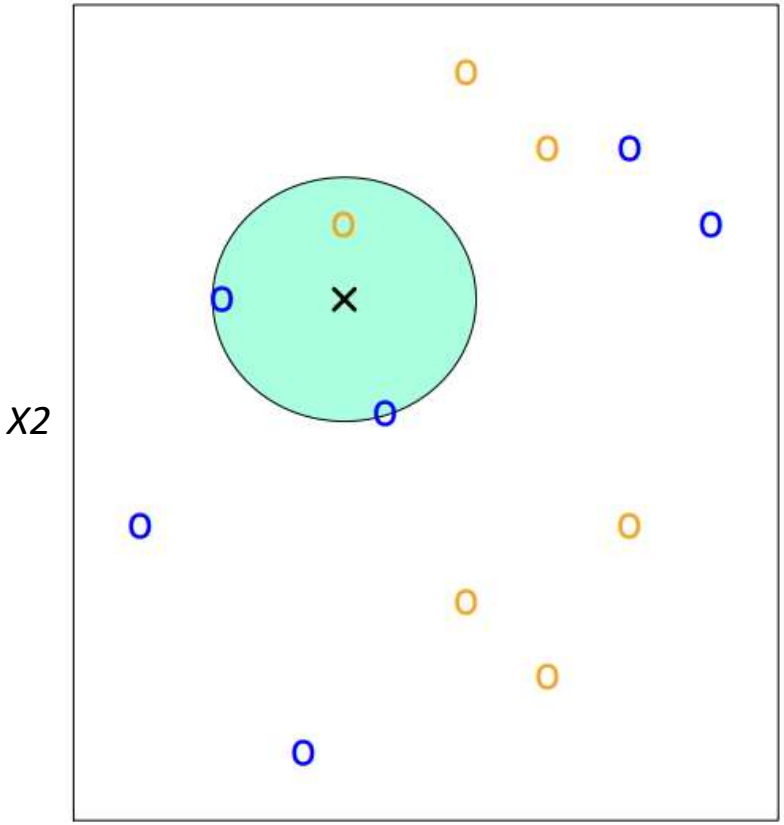
Y	X1	X2	Yhat	distance		
0	13	29		28.32		(32,8) to (13,29)
1	32	8				
0	19	21		18.38		(32,8) to (19,21)
0	12	33	0	32.02		(32,8) to (12,33)
1	28	14	1	7.21	min	(32,8) to (28,14)
1	22	12		10.77		(32,8) to (22,12)

K-Nearest Neighbors – Example (K=1)

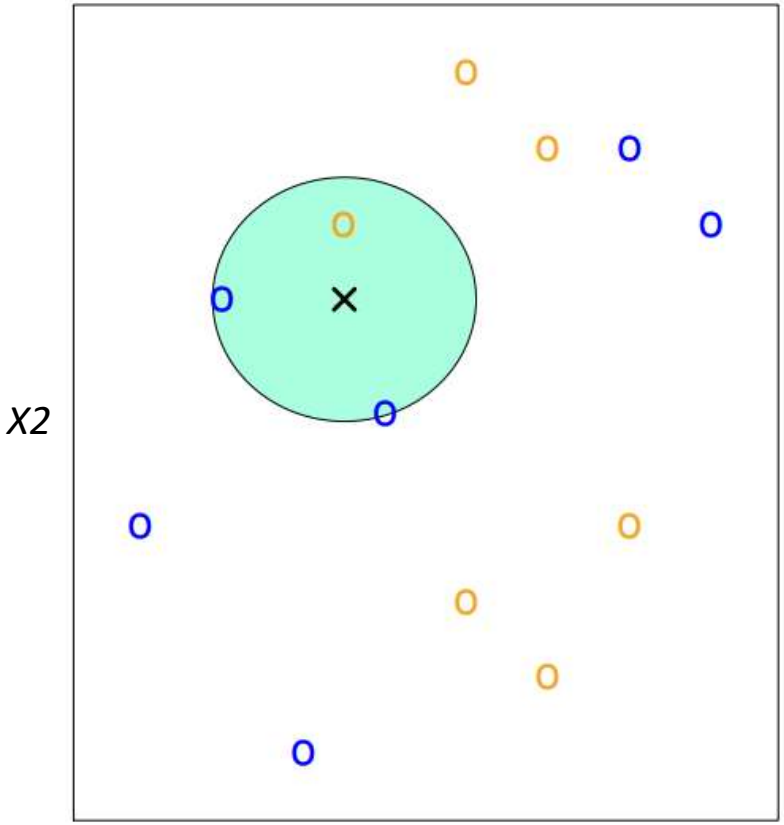
.

Y	X1	X2	Yhat	distance		
0	13	29		28.32		(32,8) to (13,29)
1	32	8				
0	19	21		18.38		(32,8) to (19,21)
0	12	33	0	32.02		(32,8) to (12,33)
1	28	14	1	7.21	min	(32,8) to (28,14)
1	22	12		10.77		(32,8) to (22,12)

K-Nearest Neighbors – Example (K=3)



K-Nearest Neighbors – Example (K=3)

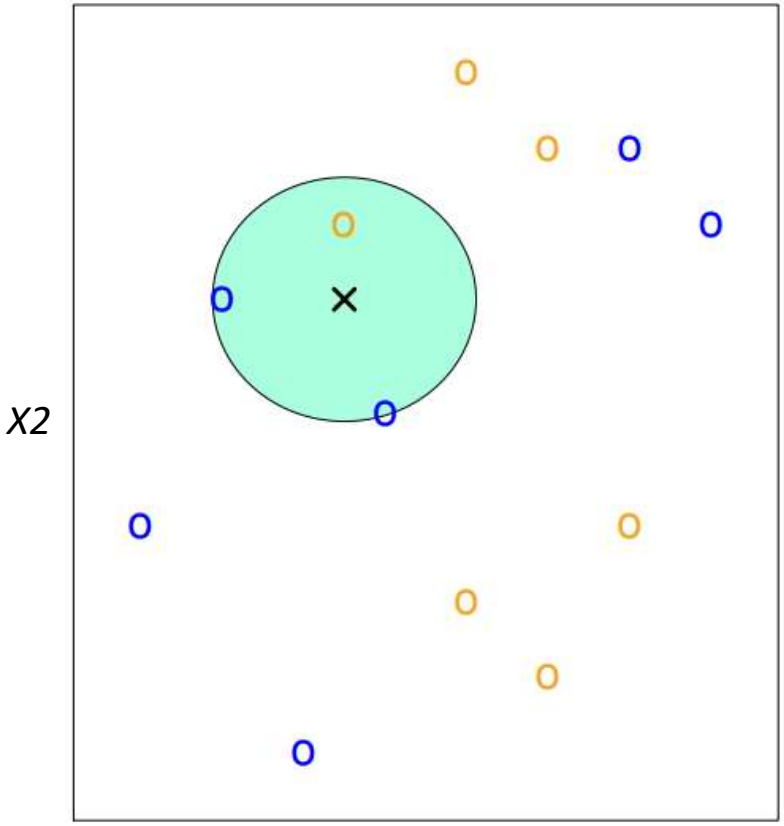


Identify k=3 observations closest to x

There are two from category 1 (blue) and one from category 0 (orange)

Hence x is predicted from category 1 (blue)

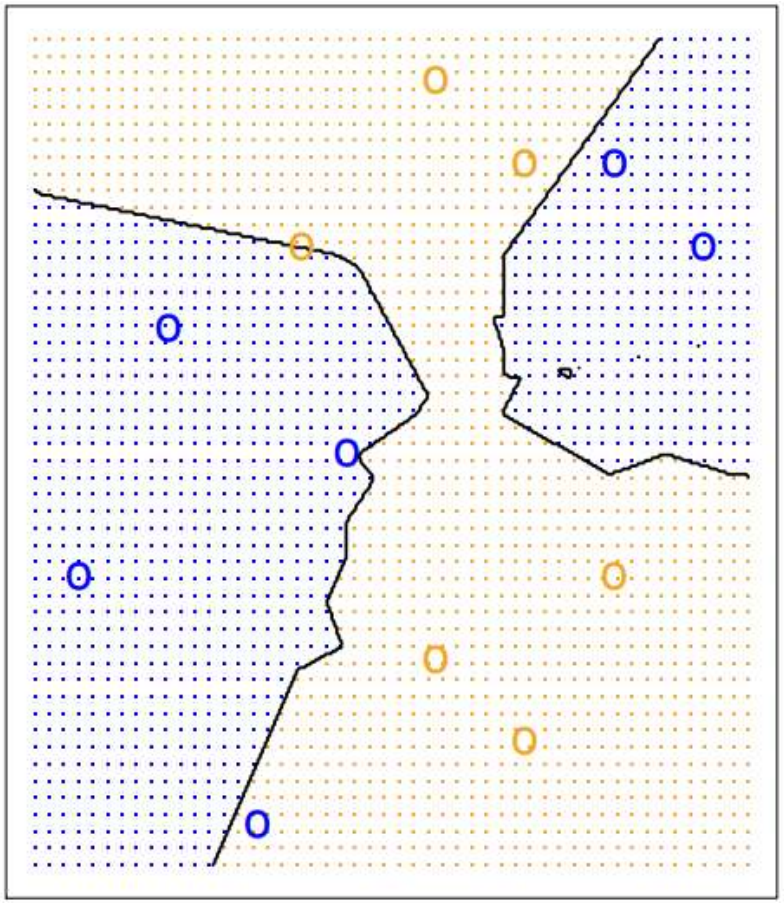
K-Nearest Neighbors – Example (K=3)



Identify k=3 observations closest to x
There are two from category 1 (blue) and one from category 0 (orange)
Hence x is predicted from category 1 (blue) with probability 2/3

K-Nearest Neighbors – Example (K=3)

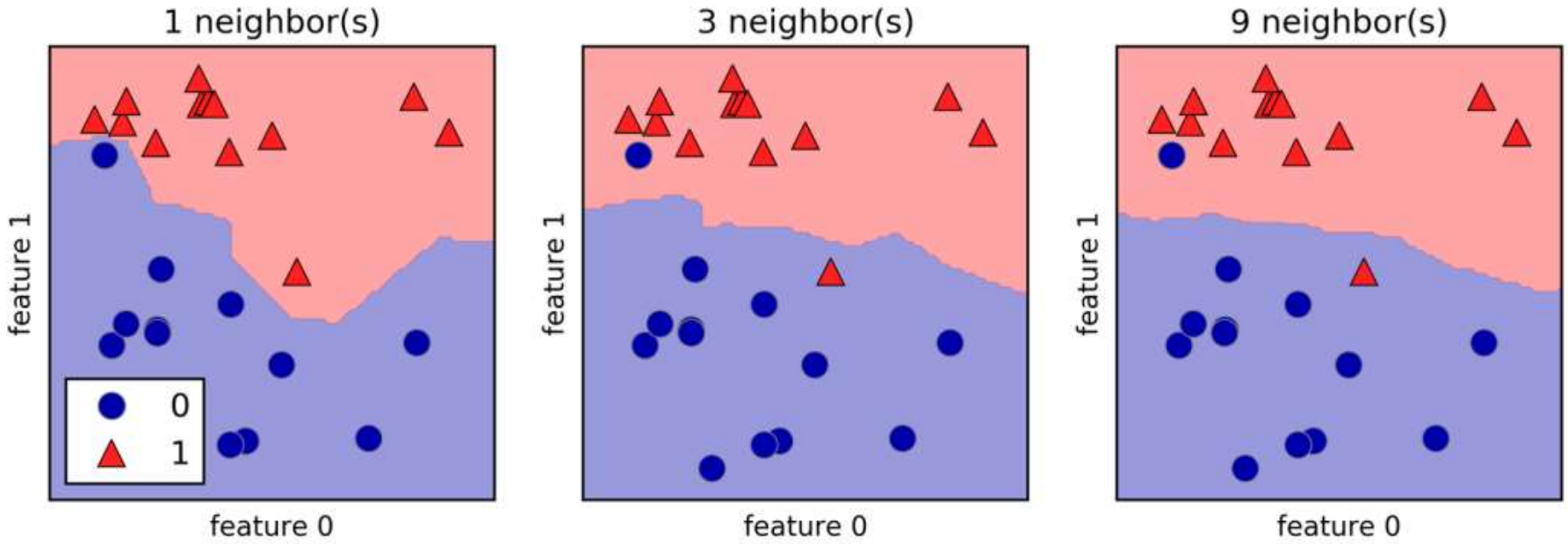
KNN (k=3) DECISION BOUNDARY



K-Nearest Neighbors – Assumptions

- All predictors are numeric
- KNN is a distance-based method (predictions may change with a different scale)
- large k results in a smooth decision boundary
- small k results in a more variable decision boundary

K-Nearest Neighbors, k = 1,3,9



Use Cross validation to select best value for *k*