

# Classification With K-Nearest Neighbors

---

GA DAT3

# Supervised vs. Unsupervised Learning

	continuous	categorical
supervised	regression	
classification		
unsupervised	reduction dimension	clustering

# Supervised vs. Unsupervised Learning

	continuous	categorical
supervised	regression	
classification		
unsupervised	reduction dimension	clustering

# Supervised Learning

150  
observations  
( $n = 150$ )



Fisher's *Iris* Data

Sepal length ⇅	Sepal width ⇅	Petal length ⇅	Petal width ⇅	Species ⇅
5.1	3.5	1.4	0.2	<i>I. setosa</i>
4.9	3.0	1.4	0.2	<i>I. setosa</i>
4.7	3.2	1.3	0.2	<i>I. setosa</i>
4.6	3.1	1.5	0.2	<i>I. setosa</i>
5.0	3.6	1.4	0.2	<i>I. setosa</i>
5.4	3.9	1.7	0.4	<i>I. setosa</i>
4.6	3.4	1.4	0.3	<i>I. setosa</i>
5.0	3.4	1.5	0.2	<i>I. setosa</i>

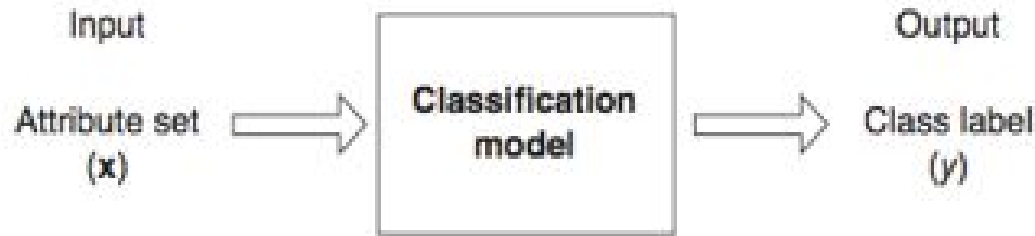
  
response

  
4 predictors ( $p = 4$ )

# Classification Problems

Q: How does a classification problem work?

A: Data in, predicted labels out.

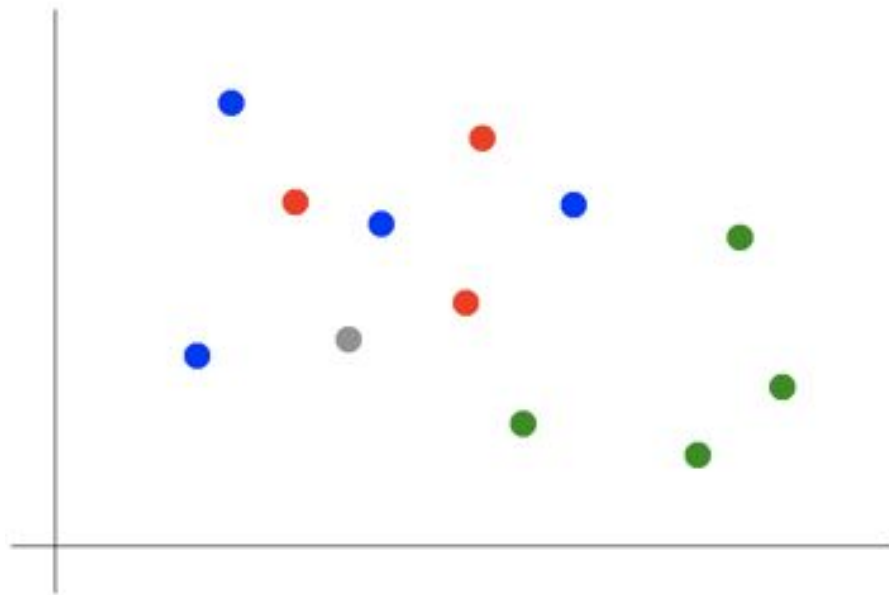


**Figure 4.2.** Classification as the task of mapping an input attribute set  $x$  into its class label  $y$ .

# Classification With KNN

Suppose we want to predict the color of the gray dot.

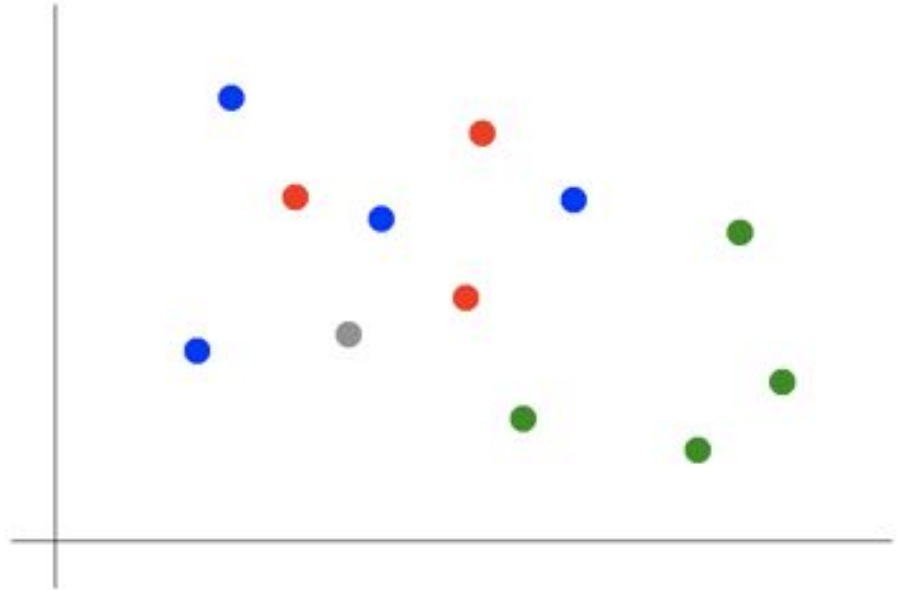
QUESTION:  
What are the  
predictors?  
What is the  
response?



# Classification With KNN

Suppose we want to predict the color of the gray dot.

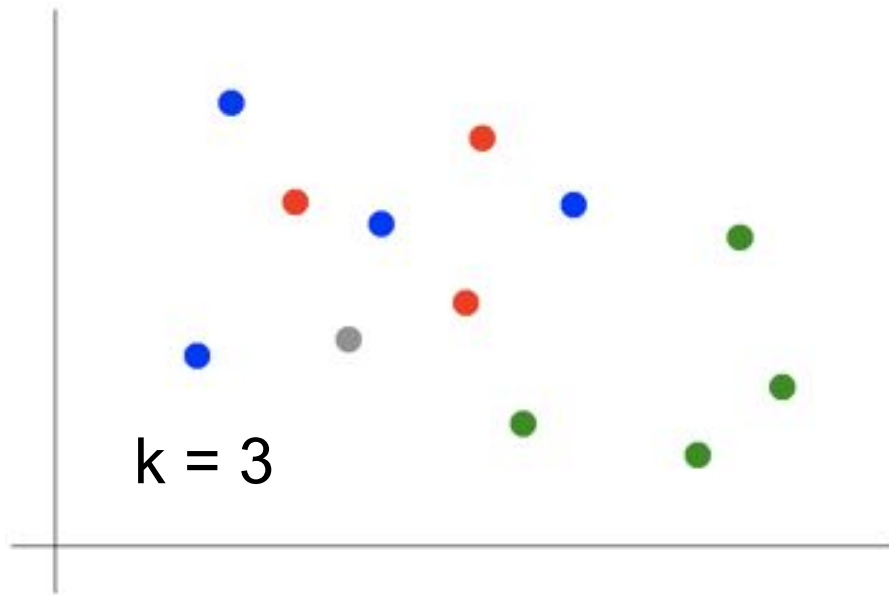
1) Pick a value for  $k$ .



# Classification With KNN

Suppose we want to predict the color of the gray dot.

1) Pick a value for  $k$ .





# Classification With KNN

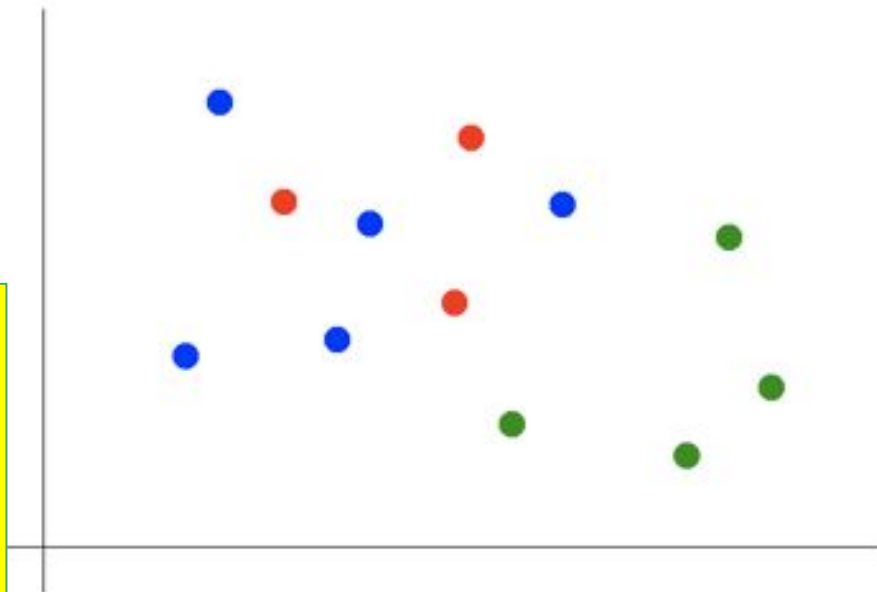
Suppose we want to predict the color of the gray dot.

1) Pick a value for  $k$ .

2) Find colors of  $k$   
nearest neighbors.

3) Assign the most  
common color  
to the gray dot.

NOTE:  
Our definition  
of  
“nearest”  
implicitly uses  
the Euclidean  
distance  
function.



# Classification With KNN

## Advantages of KNN:

- Simple to understand and explain
- Model training phase is fast
- Non-parametric (does not presume a “form” of the “decision boundary”)

## Disadvantages of KNN:

- Prediction phase can be slow when  $n$  is large
- Sensitive to irrelevant features

Q??

---