



Lecture 35

Classification

Announcements

Prediction

Guessing the Value of an Attribute

- Based on incomplete information
 - One way of making predictions:
 - To predict an outcome for an individual,
 - find others who are like that individual
 - and whose outcomes you know.
 - Use those outcomes as the basis of your prediction.
 - Two Types of Prediction
 - Classification = Categorical; Regression = Numeric
-

Prediction Example: Spam or Not?

You made a Wells Fargo payment - wells Fargo.com You recently submitted a payment The ...

BUSINESS TRUST - -- I have a legal business proposal for you worth \$23,000,000. If you kn...

Hi - Today???!!!! What a wonderful day! Congrats again! I am definitely not doing s...

Michael Kors Handbags Up To 84% Plus Free Shipping! - Shop Handbags Online & In Store...

Machine Learning Algorithm

- A mathematical model
 - calculated based on sample data ("training data")
 - that makes predictions or decisions without being explicitly programmed to perform the task
-

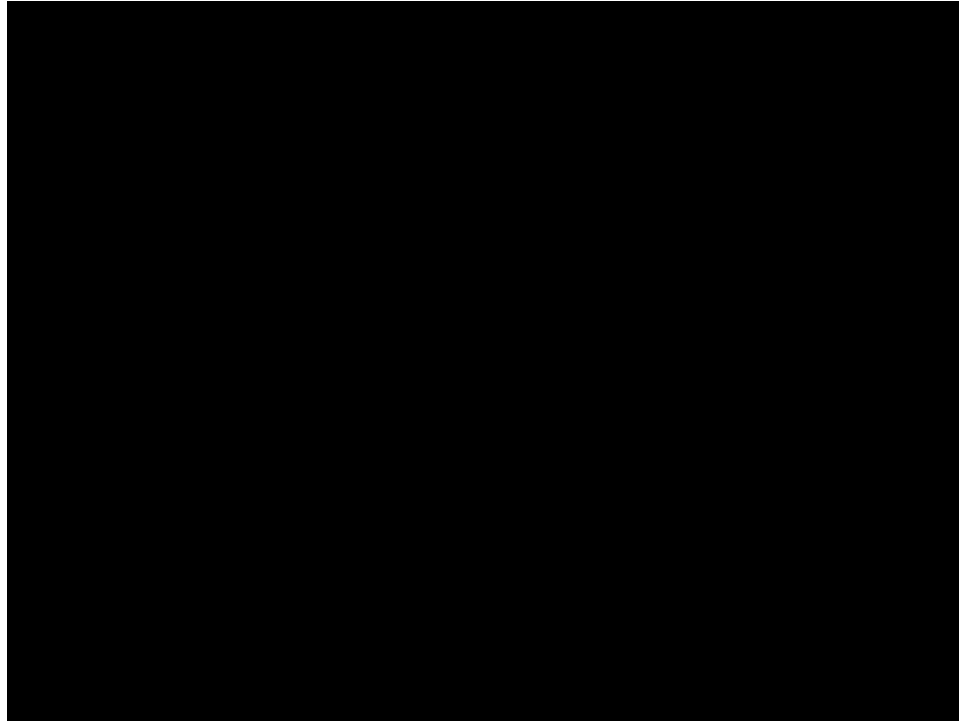
Classification

Classification Examples

will be automatically deleted. [Delete all spam messages now](#)

I have a legal business proposal for you worth \$23,000,000....

Classification Examples



Classification Examples

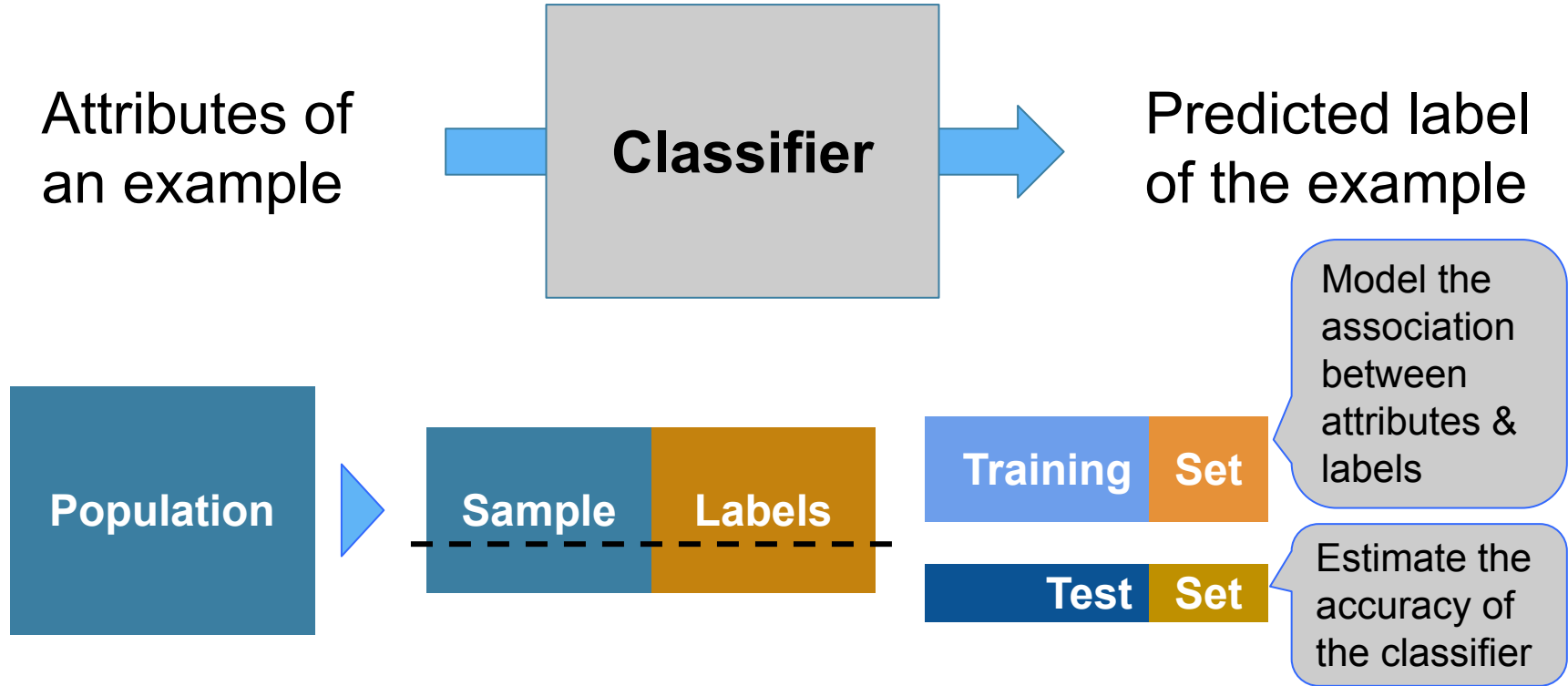
Top picks for you



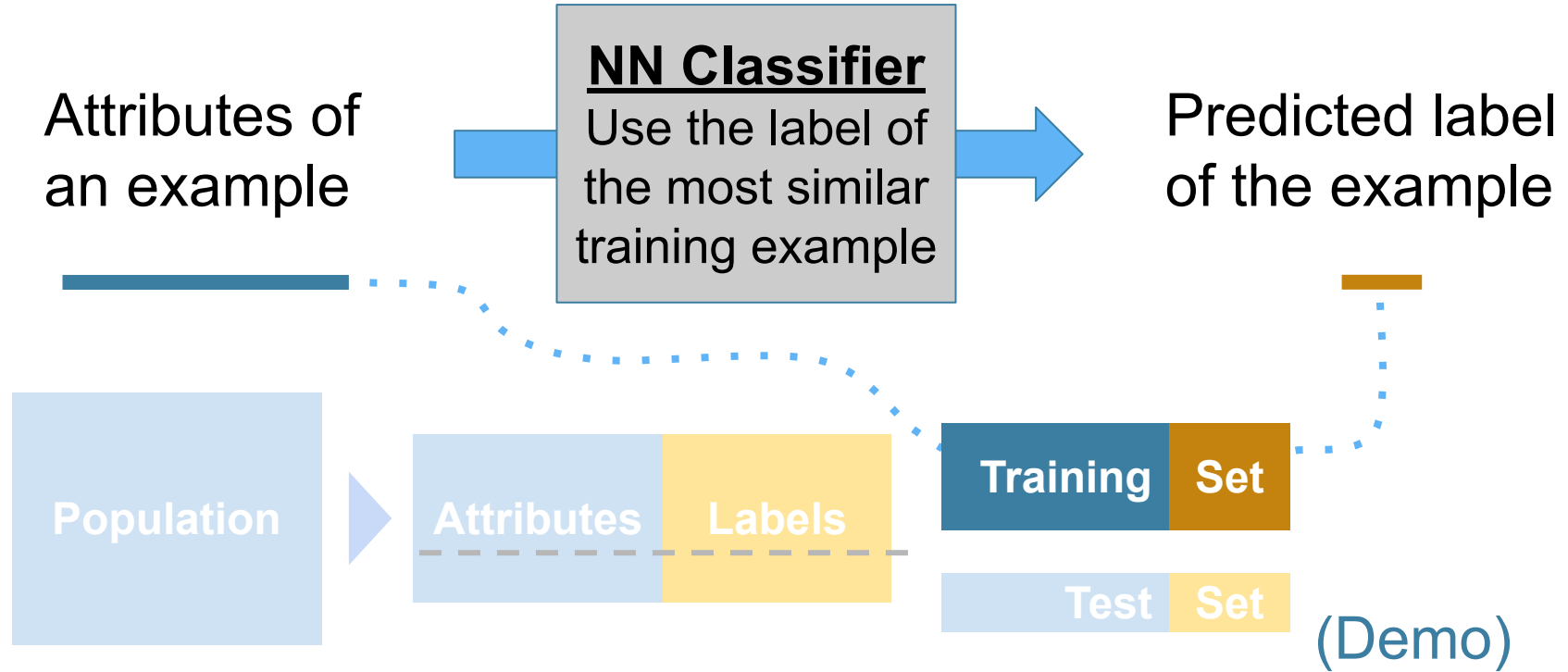
(Demo)

Classifiers

Training a Classifier



Nearest Neighbor Classifier



Rows

Rows of Tables

Each row contains all the data for one individual

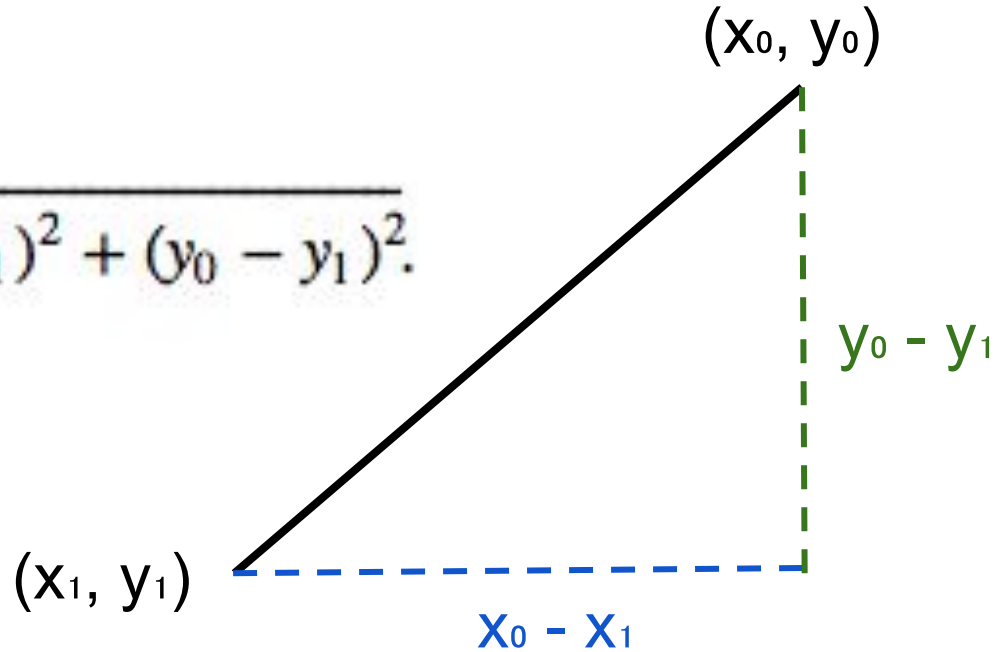
- `t.row(i)` evaluates to *i*th row of table `t`
 - `t.row(i).item(j)` is the value of column `j` in row `i`
 - If all values are numbers, then `np.array(t.row(i))` evaluates to an array of all the numbers in the row.
 - To consider each row individually, use

```
for row in t.rows:  
    ... row.item(j) ...
```
 - `t.exclude(i)` evaluates to the table `t` without its *i*th row
-

Distance

Pythagoras' Formula

$$D = \sqrt{(x_0 - x_1)^2 + (y_0 - y_1)^2}.$$



Distance Between Two Points

- Two attributes x and y :

$$D = \sqrt{(x_0 - x_1)^2 + (y_0 - y_1)^2}.$$

- Three attributes x , y , and z :

$$D = \sqrt{(x_0 - x_1)^2 + (y_0 - y_1)^2 + (z_0 - z_1)^2}$$

- and so on ...

(Demo)