Machine Learning – Steps in model building

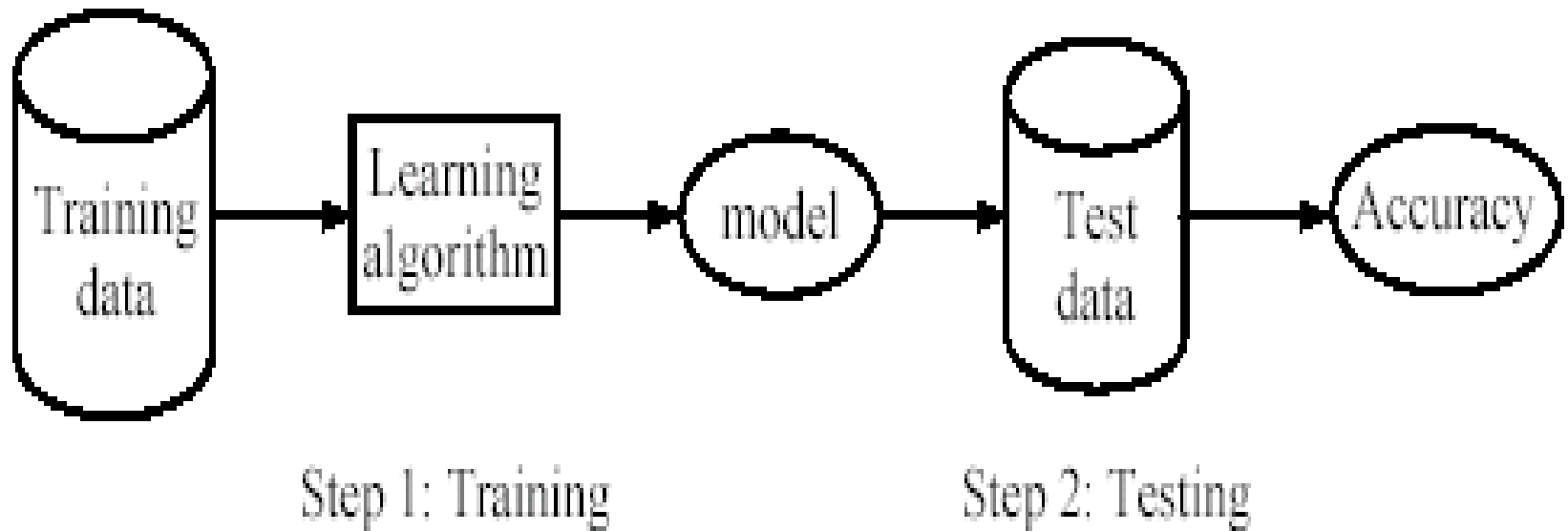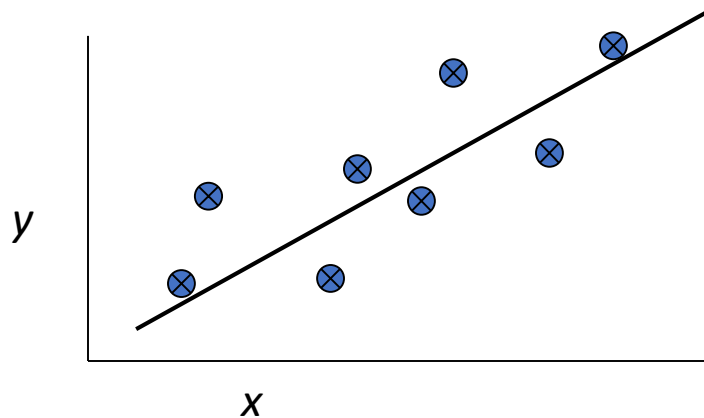Step 1: Training   Step 2: Testing

Regression

- In regression the output is continuous – like car MPG
  - Function Approximation
- Many models could be used – Simplest is linear regression
  - Fit data with the best hyper-plane which "goes through" the points
  - For each point the differences between the predicted point and the actual observation is the *residue*

How do we "learn" parameters

- For the 2-d problem (line) there are coefficients for the X attributes

- Regression Goal: To find the values for the coefficients which minimize the objective function value

Summary

- Linear Regression are tools for many simple situations
  - Fit the data with one linear or non-linear equation
- Use series of regression models to find which gives best results
- Results vary based on
  - Number of X values
  - Number of rows of data
  - Accuracy
- Regression models built
  - Linear Regression, Ridge, Lasso, LassoLars, BayesianRidge
  - More models exist
  - Add to the "for loop" for trying more models

Auto MPG Dataset

Y is a function of X

Target (Y)
MPG

X Values = 6

	MPG	Cylinders	Displacement	HP	Weight	Acceleration	Year	Origin	Car Name
1								D	D
2								R	R
…								O	O
								P	P
								P	P
397								E	E
398								D	D

398 = Total at start
6 = Drop from rows with missing values
====
392
75:25 Train/test split in ratio
294 = Training rows (X_train,y_train)
98 = Testing rows (X_test,y_test)

https://archive.ics.uci.edu/ml/datasets/auto+mpg