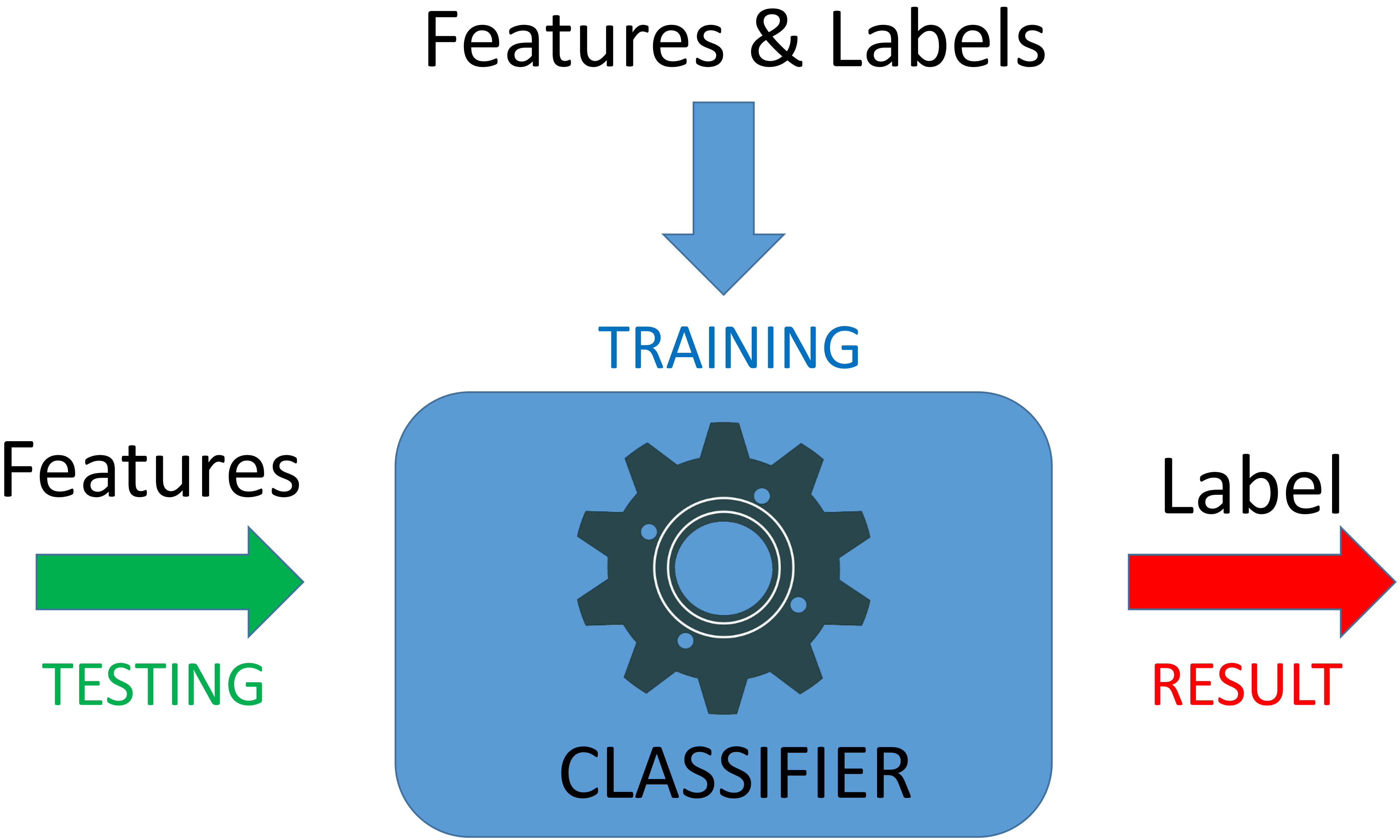
WEEK 02 - MACHINE LEARNING INTRO

LINEAR REGREESSION & IMAGE PRE-PROCESSING

Thakshila Dasun
BSc. Hons in Mechanical Engineering
(Mechatronics Specialization)
CIMA, UK
Academy of Innovative Education

AIE
ACADEMY OF INNOVATIVE EDUCATION

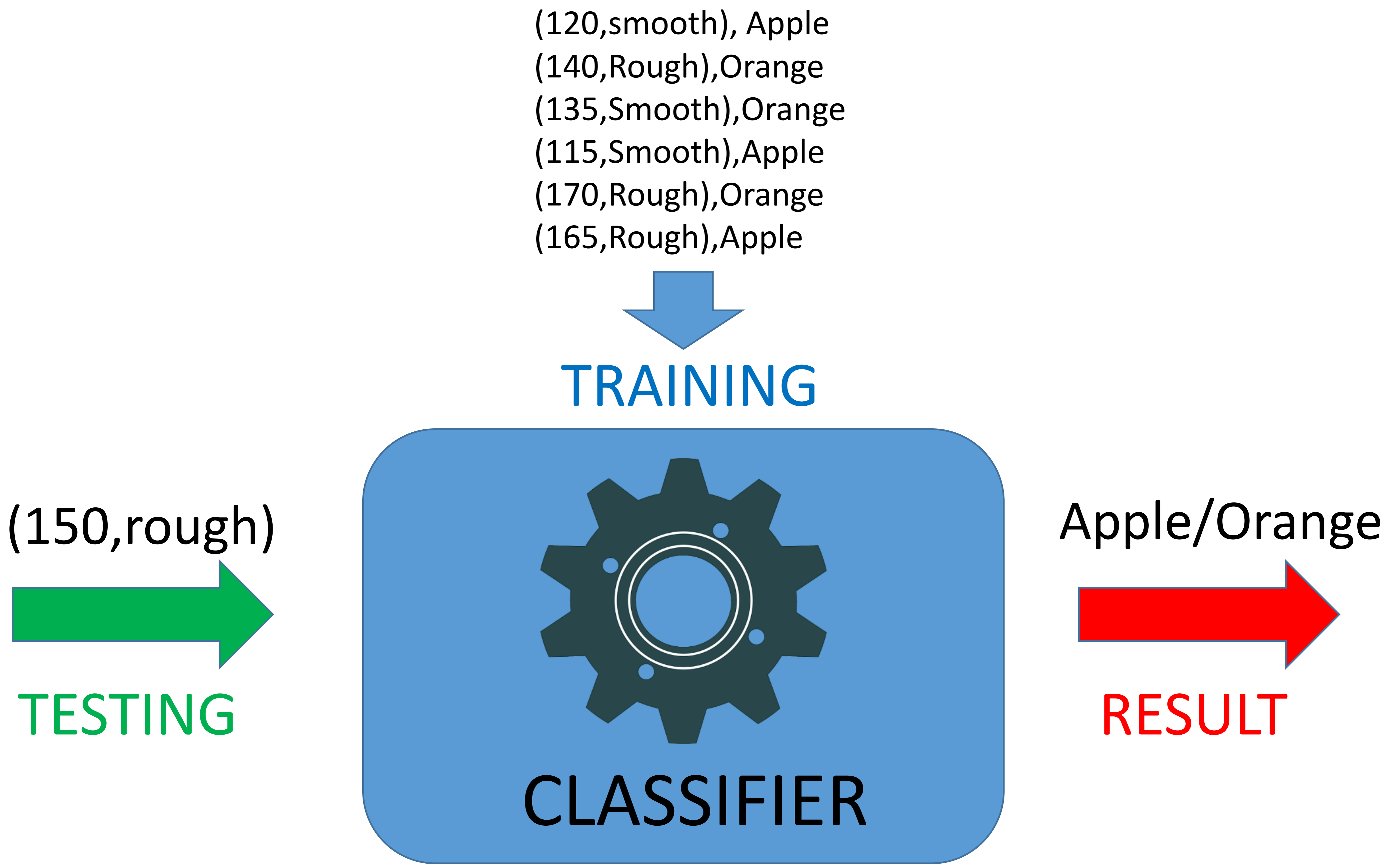

Features & Labels
TRAINING
Features
TESTING
CLASSIFIER
Label
RESULT

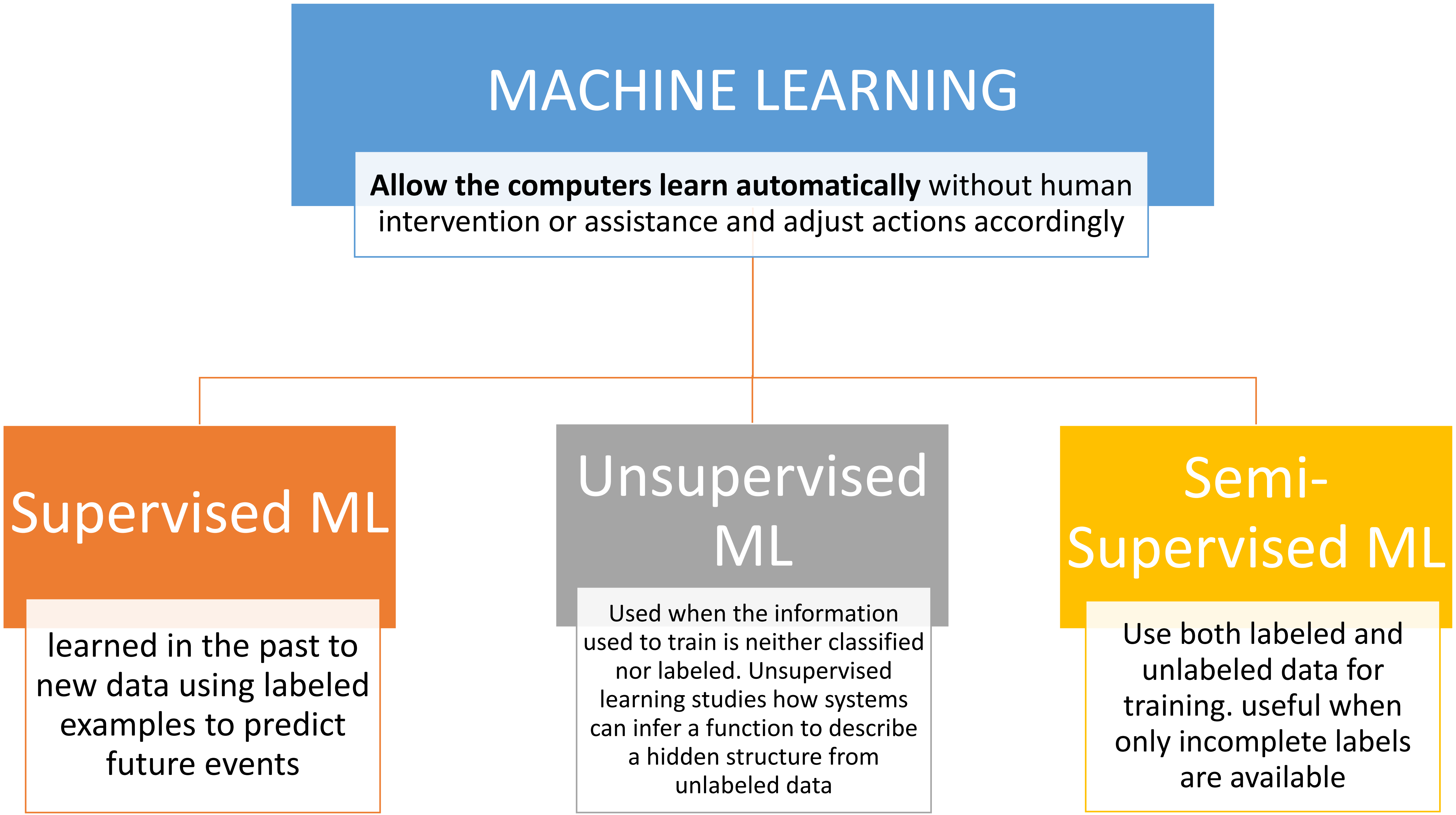
What are the Critical Features ?

1. Color
2. Weight
3. Texture
4. Radius
5. Volume

Prepare a dataset from good features

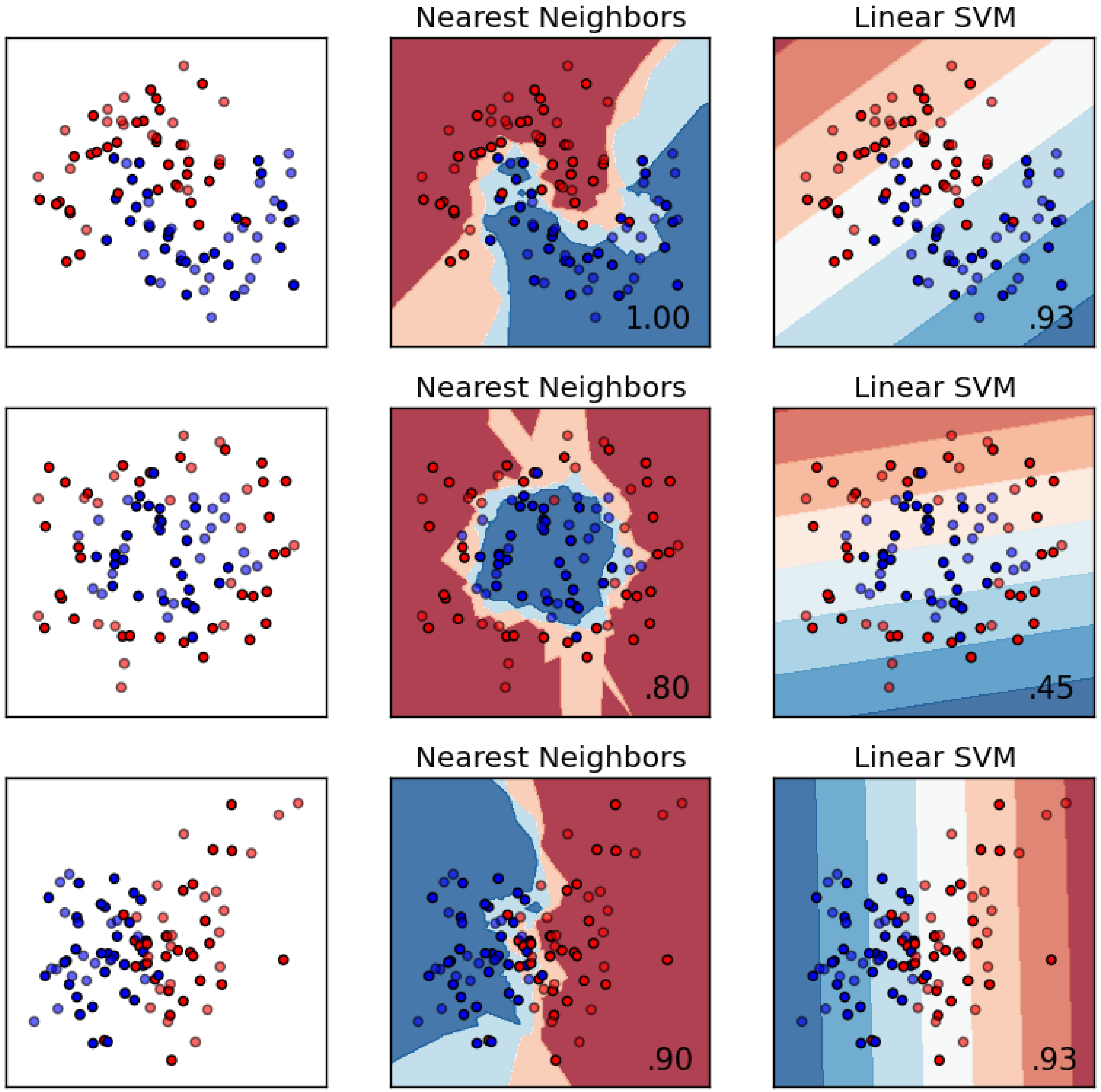
FEATURES		LABEL
Weight (g)	Texture	
120	Smooth	Apple
140	Rough	Orange
135	Smooth	Orange
115	Smooth	Apple
170	Rough	Orange
165	Rough	Apple


(120,smooth), Apple
(140,Rough),Orange
(135,Smooth),Orange
(115,Smooth),Apple
(170,Rough),Orange
(165,Rough),Apple
TRAINING
(150,rough)
TESTING
CLASSIFIER
Apple/Orange
RESULT


MACHINE LEARNING
Allow the computers learn automatically without human intervention or assistance and adjust actions accordingly
Supervised ML
learned in the past to new data using labeled examples to predict future events
Unsupervised ML
Used when the information used to train is neither classified nor labeled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabeled data
Semi-Supervised ML
Use both labeled and unlabeled data for training. useful when only incomplete labels are available

Classification & Regression

2 Types of Predictions in Machine Learning, Qualitative and Quantitative

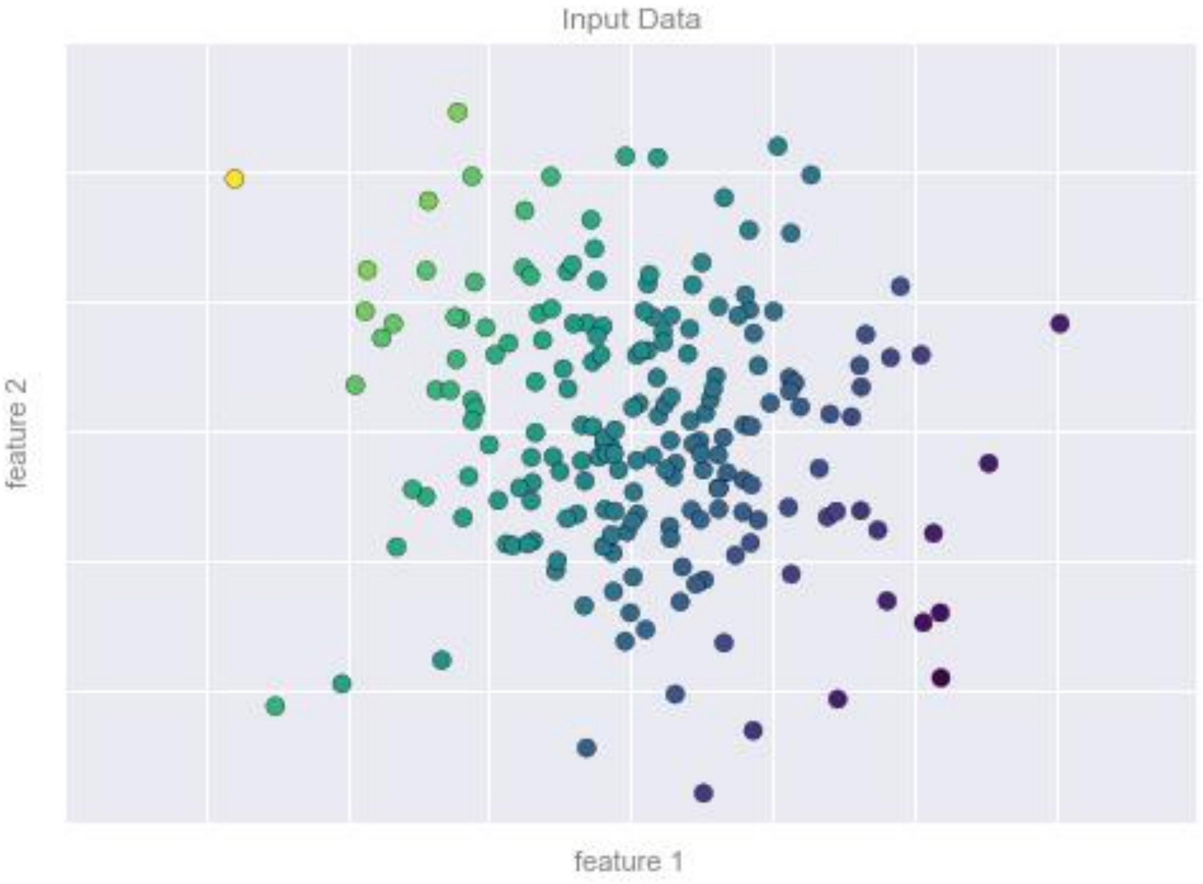


1. Classification: Predicting discrete labels

- “Classification" indicates that the data has discrete class label.
- Classification predictive modeling is the task of approximating a mapping function (f) from input variables (X) to discrete output variables (y) or classes.
- The output variables are often called labels or categories. The mapping function predicts the class or category for a given observation

2. Regression: Predicting continuous labels

- Regression predictive modeling is the task of approximating a mapping function (f) from input variables (X) to a continuous output variable (y).
- A continuous output variable is a real-value, such as an integer or floating point value. These are often quantities, such as amounts and sizes.
- For example, a house may be predicted to sell for a specific dollar value, perhaps in the range of $100,000 to $200,000.



Algorithms (1)

- Classification Algorithms (Supervised):
  1. Support vector machines
  2. Nearest Neighbors
  3. Decision Trees and Random Forests
  4. Gaussian naive Bayes

- Classification Algorithms (unsupervised):
  1. K-Means Clustering.
  2. Means Shift Clustering
  3. Gaussian Mixture Models
  4. Spectral Clustering

Algorithms (2)

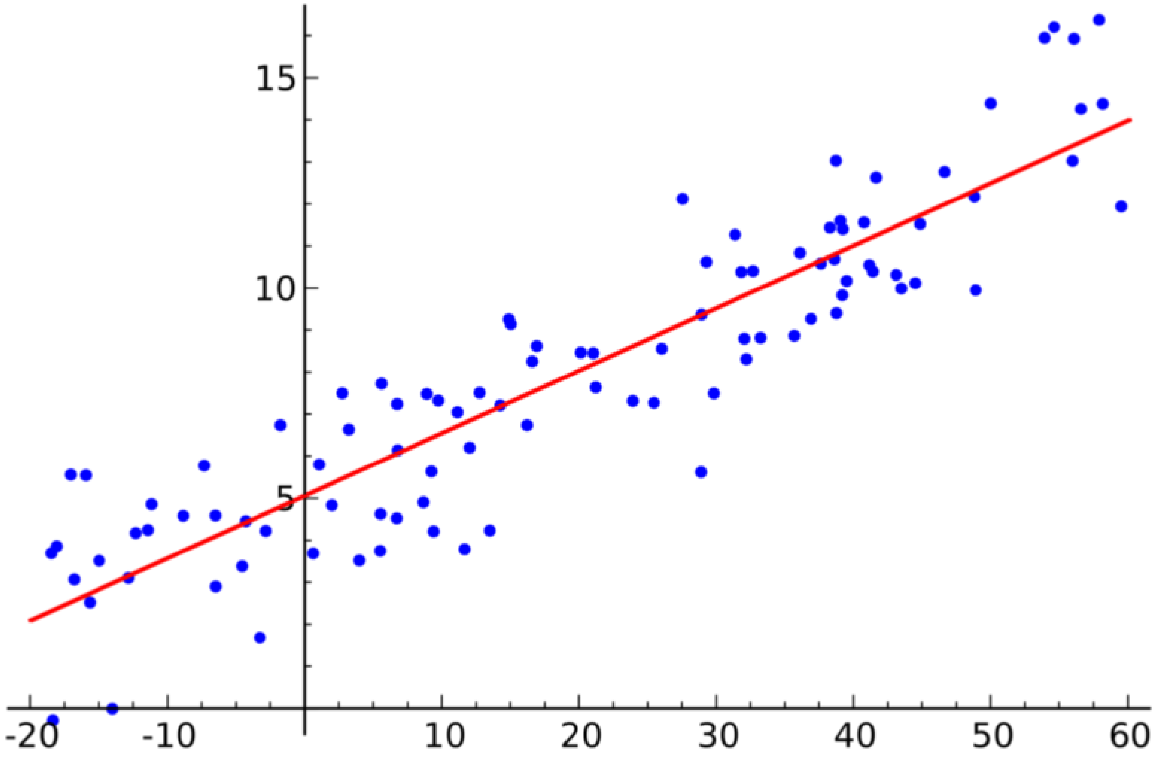
- Dimensionality Reduction Algorithms
    1. Manifold Learning.
    2. Principal Component Analysis

- Regression Algorithms
    1. Linear regression
    2. Support Vector Machines
    3. Random forest regression

- High Level Classifiers (Deep Learning)
    1. Neural Networks

LINEAR REGRESSION (1)

- Regression is a method of modelling a target value based on independent predictors.
- This method is mostly used for forecasting and finding out cause and effect relationship between variables.
- Regression techniques mostly differ based on the number of independent variables and the type of relationship between the independent and dependent variables.



LINEAR REGRESSION (2)

- Simple linear regression is a type of regression analysis where the number of independent variables is one and there is a linear relationship between the independent(x) and dependent(y) variable.

$$y = m*x + c$$

$$m = \frac{\overline{x} \cdot \overline{y} - \overline{xy}}{(\overline{x})^2 - \overline{x^2}}$$

$$b = \overline{y} - m\overline{x}$$

R Squared and Coefficient of Determination Theory (1)

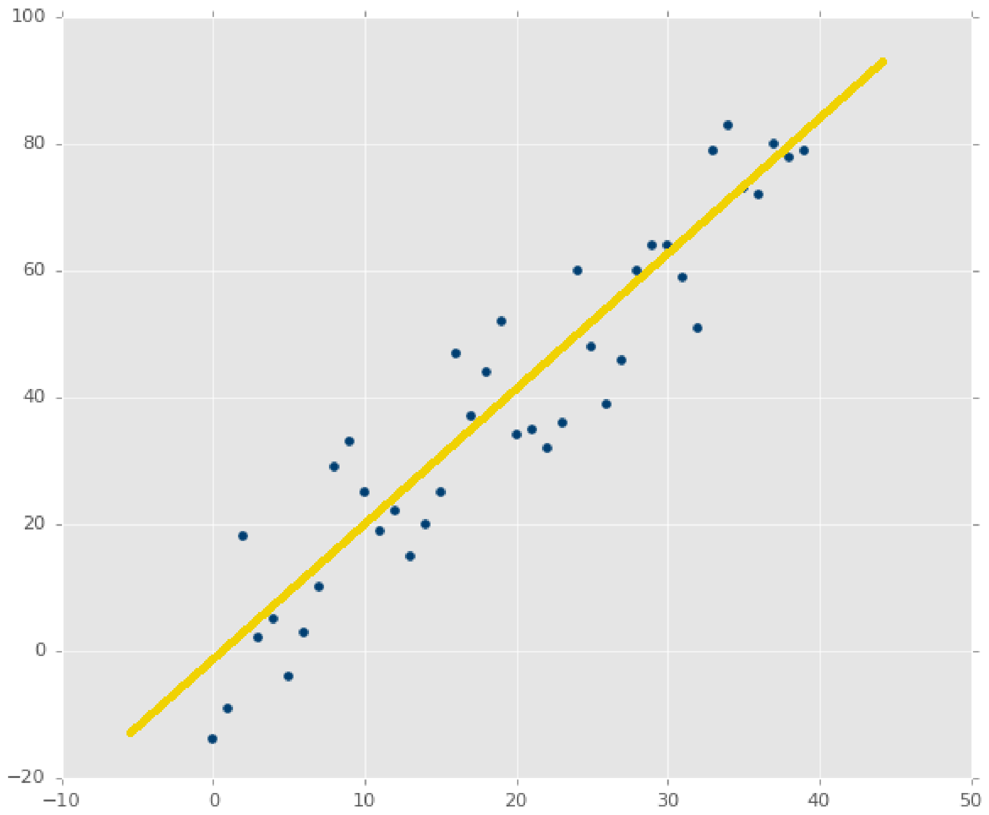


Figure 1

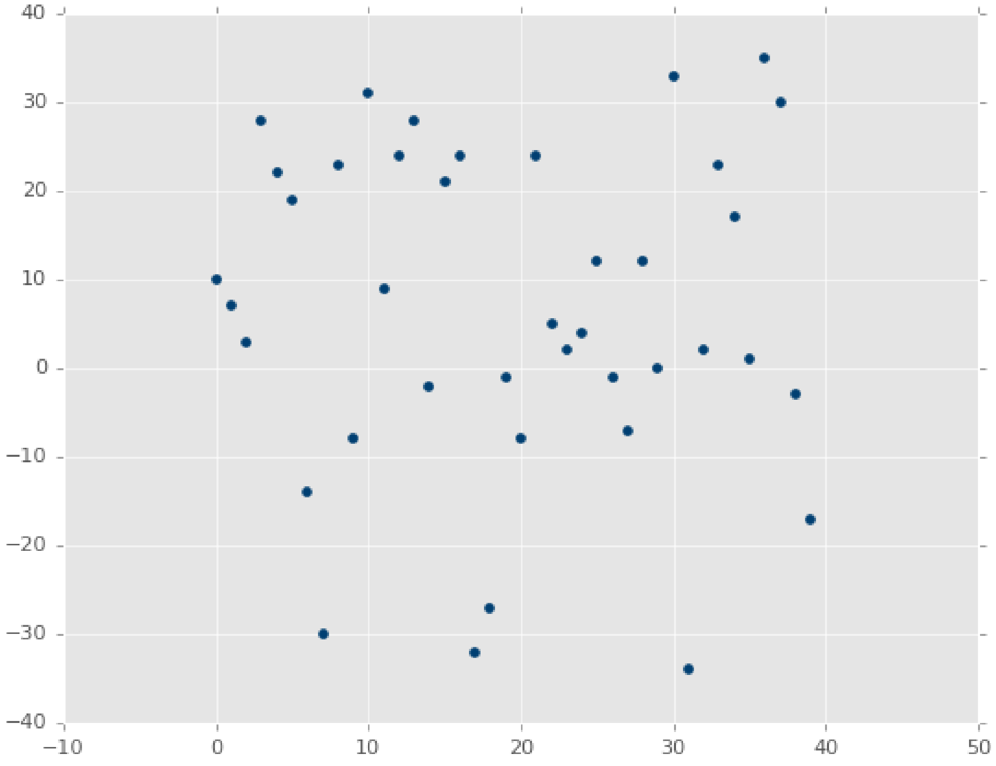


Figure 2

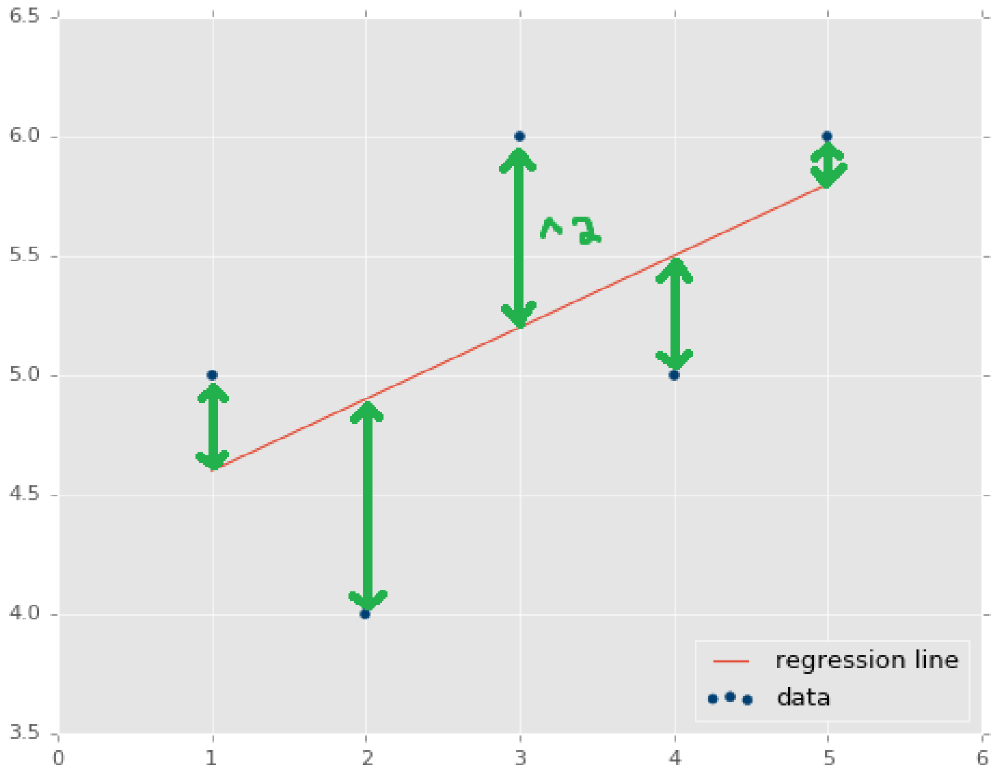
R Squared and Coefficient of Determination Theory (2)

- In the second image, there is a best fit line, though even the best fitting line is still going to be useless,
- And we'd like to know that before we spend precious computational power on it.
- The standard way to check for errors is by using squared errors. You will hear this method either called R squared or the coefficient of determination.

How to Compute Coefficient of Determination

- The distance between the regression line's y values, and the data's y values is the error, then we square that. The line's squared error is either a mean or a sum of this, we'll simply sum it.

$$r^2 = 1 - \frac{SE\hat{y}}{SE_{\overline{y}}}$$



Predicting Accuracy of the model by Coefficient of Determination

- The equation is essentially 1 minus the division of the squared error of the regression line and the squared error of the mean y line.
- The goal is to have the r squared value, otherwise called the coefficient of determination, as close to 1 as possible.

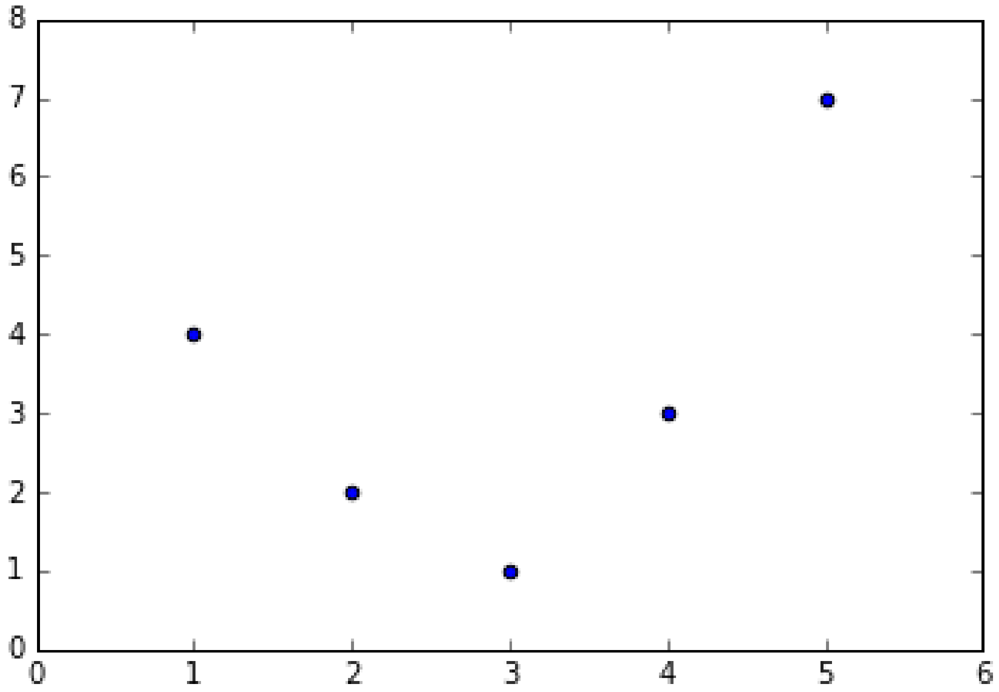
One Real Example

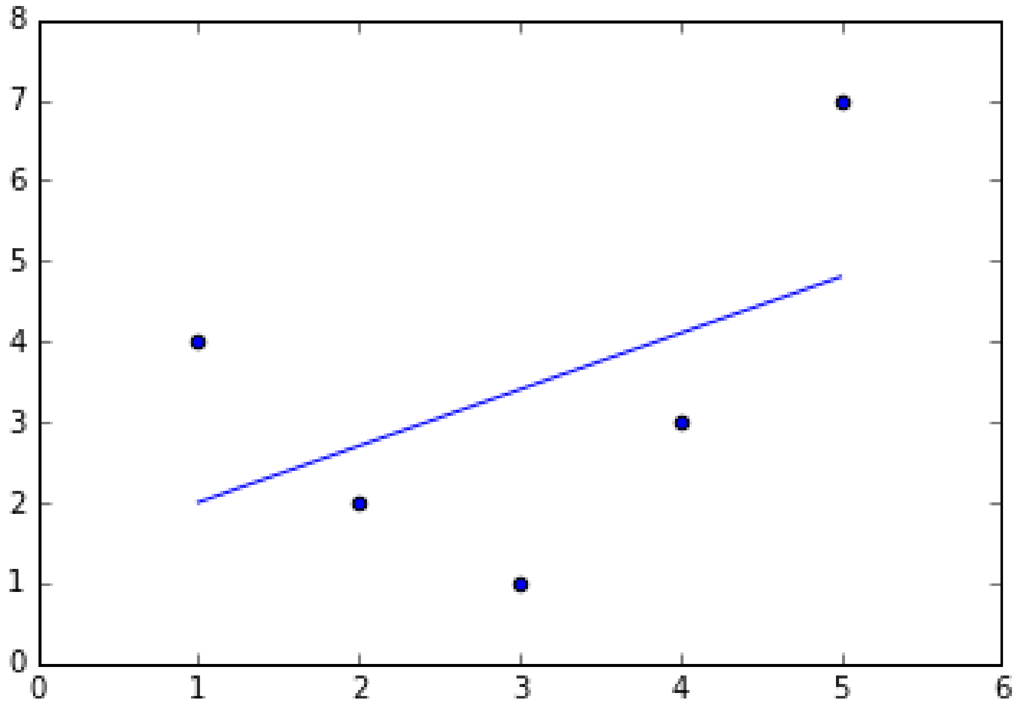
Table 8.3: Trend Adjusted Exponential Smoothing Forecasts with Smoothing Constant of 0.5 and Smoothing Constant for Trend of 0.3

Period	Month	Demand	F_t ($\alpha = 0.5$)	T_t ($\beta = 0.3$)	AF_t
1	January	38	38.00	-	-
2	February	40	38.00	0.00	38.00
3	March	41	39.00	0.30	39.30
4	April	37	40.00	0.62	40.62
5	May	45	38.50	0.03	38.53
6	June	50	41.75	1.02	42.77
7	July	43	45.88	1.97	47.85
8	August	47	44.44	0.95	45.39
9	September	56	45.72	1.05	46.77
10	October	52	50.86	2.28	53.14
11	November	56	51.43	1.77	53.20
12	December	58	53.71	1.92	55.63
13	January	-	55.86	1.88	57.74

Dealing with higher dimensional polynomial functions

- This data clearly cannot be well described by a straight line,
- Still, we can fit a line to the data using LinearRegression and get the optimal result:





Feature Pipelines

- With any of the preceding examples, it can quickly become tedious to do the transformations by hand, especially if you wish to string together multiple steps. For example, we might want a processing pipeline that looks something like this:

  1. Impute missing values using the mean
  2. Transform features to quadratic
  3. Fit a linear regression