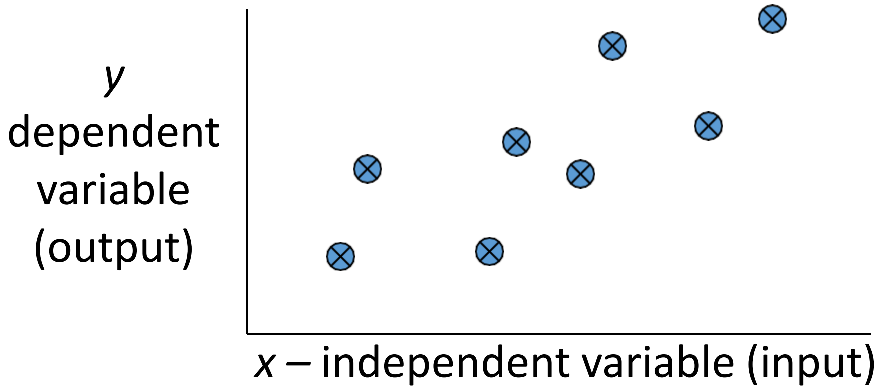
Regression

- In regression the output is continuous
- Many models could be used – Simplest is linear regression
  - Fit data with the best hyper-plane which "goes through" the points

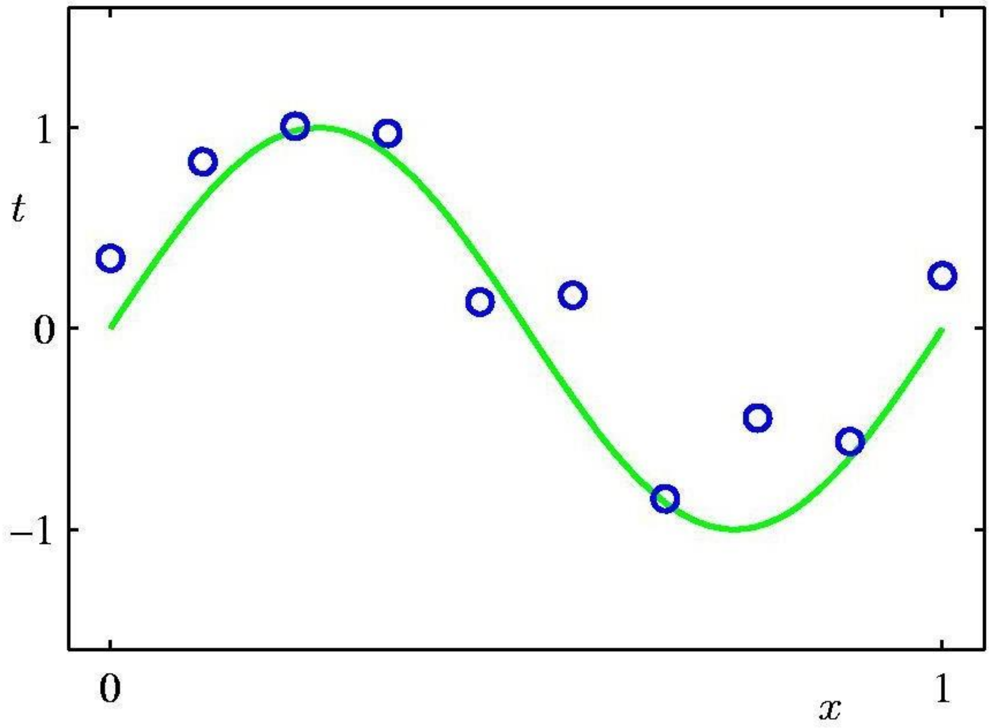


A Simple Regression Problem

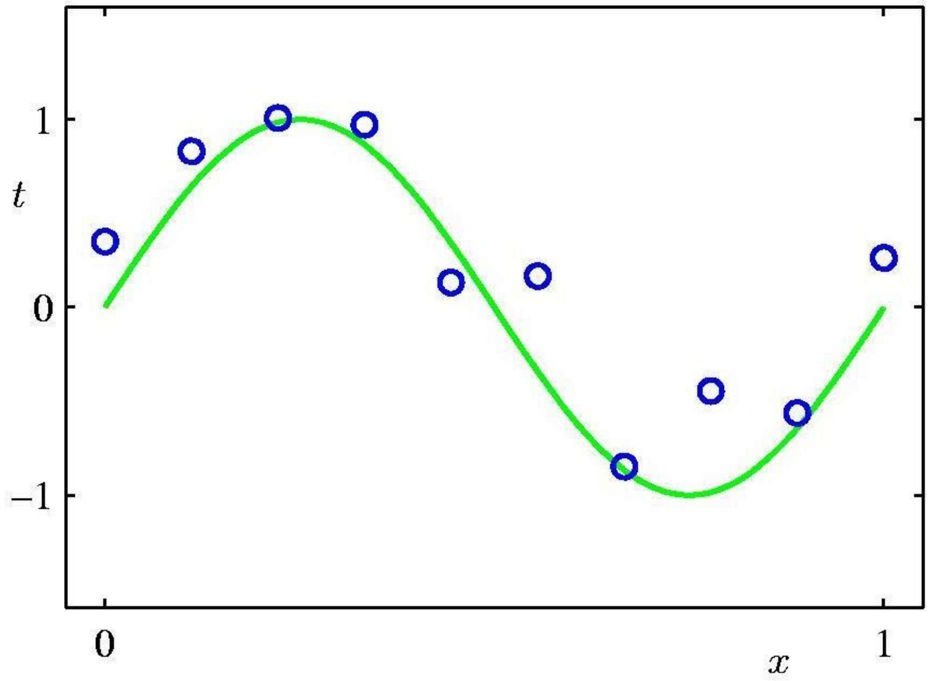
- We observe a real-valued input variable x and we wish to use this observation to predict the value of a real-valued target variable t.
- We use synthetically generated data from the function $\sin(2\pi x)$ with random noise included in the target values.
  - A small level of random noise having a Gaussian distribution
- We have a training set comprising N observations of x, written $\mathbf{x} \equiv (x_1, \ldots, x_N)^T$, together with corresponding observations of the values of t, denoted $\mathbf{t} \equiv (t_1, \ldots, t_N)^T$.
- Our goal is to predict the value of t for some new value of x,

Polynomial Curve Fitting

- A training data set of N = 10 points, (blue circles),
- The green curve shows the actual function $\sin(2\pi x)$ used to generate the data.
- Our goal is to predict the value of t for some new value of x, without knowledge of the green curve.



Polynomial Curve Fitting

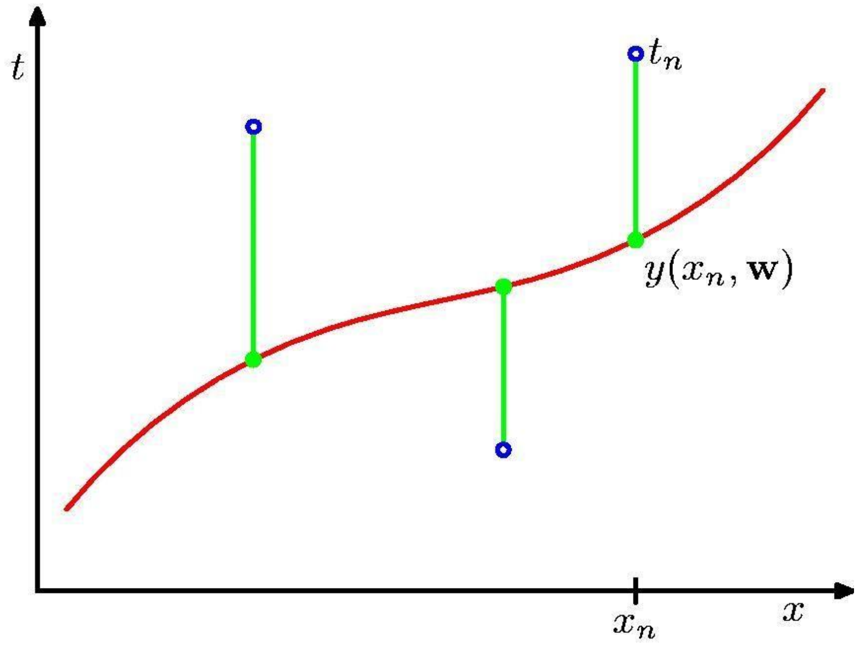


- We try to fit the data using a polynomial function of the form

$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j$$

Polynomial Curve Fitting

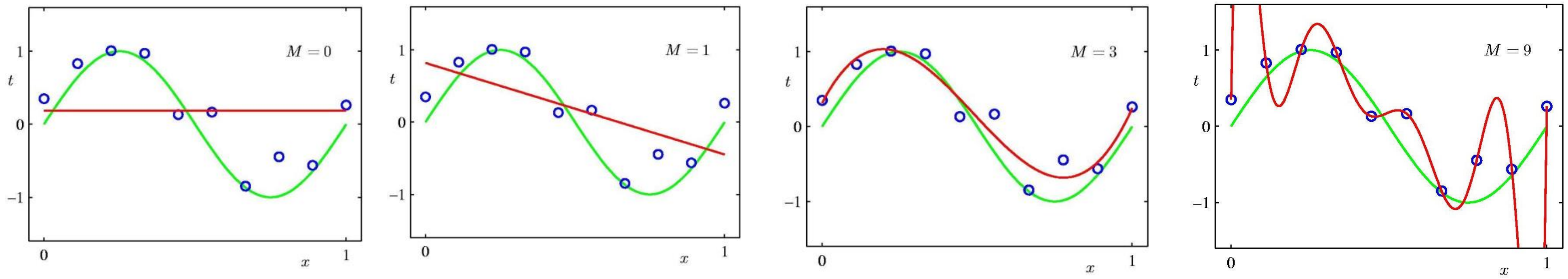
- The values of the coefficients will be determined by fitting the polynomial to the training data.
- This can be done by minimizing an error function that measures the misfit between the function y(x,w), for any given value of w, and the training set data points.
- Error Function: the sum of the squares of the errors between the predictions $y(x_n,w)$ for each data point x_n and the corresponding target values t_n.



$$E(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n,\mathbf{w}) - t_n\}^2$$

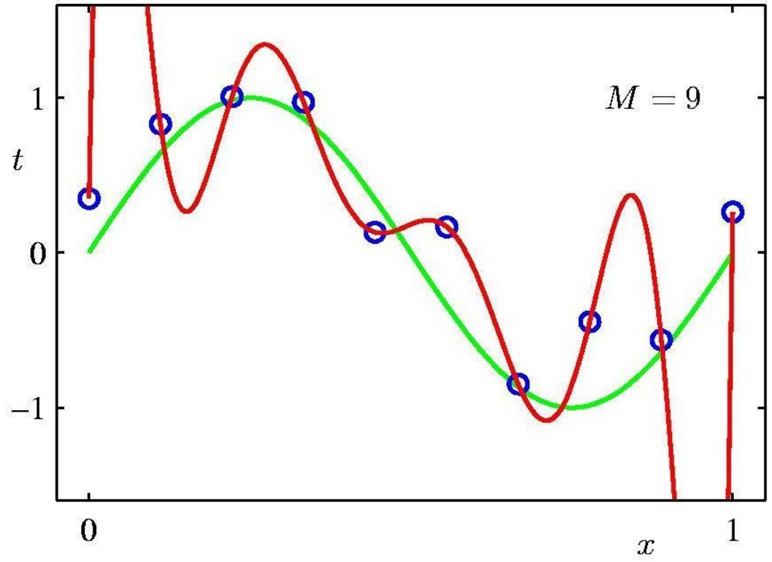
Polynomial Curve Fitting

- The 0^{th} order (M=0) and first order (M=1) polynomials give rather poor fits to the data and consequently rather poor representations of the function sin(2πx).
- The third order (M=3) polynomial seems to give the best fit to the function sin(2πx) of the examples.
- When we go to a much higher order polynomial (M=9), we obtain an excellent fit to the training data.
  - In fact, the polynomial passes exactly through each data point and E(w*) = 0.



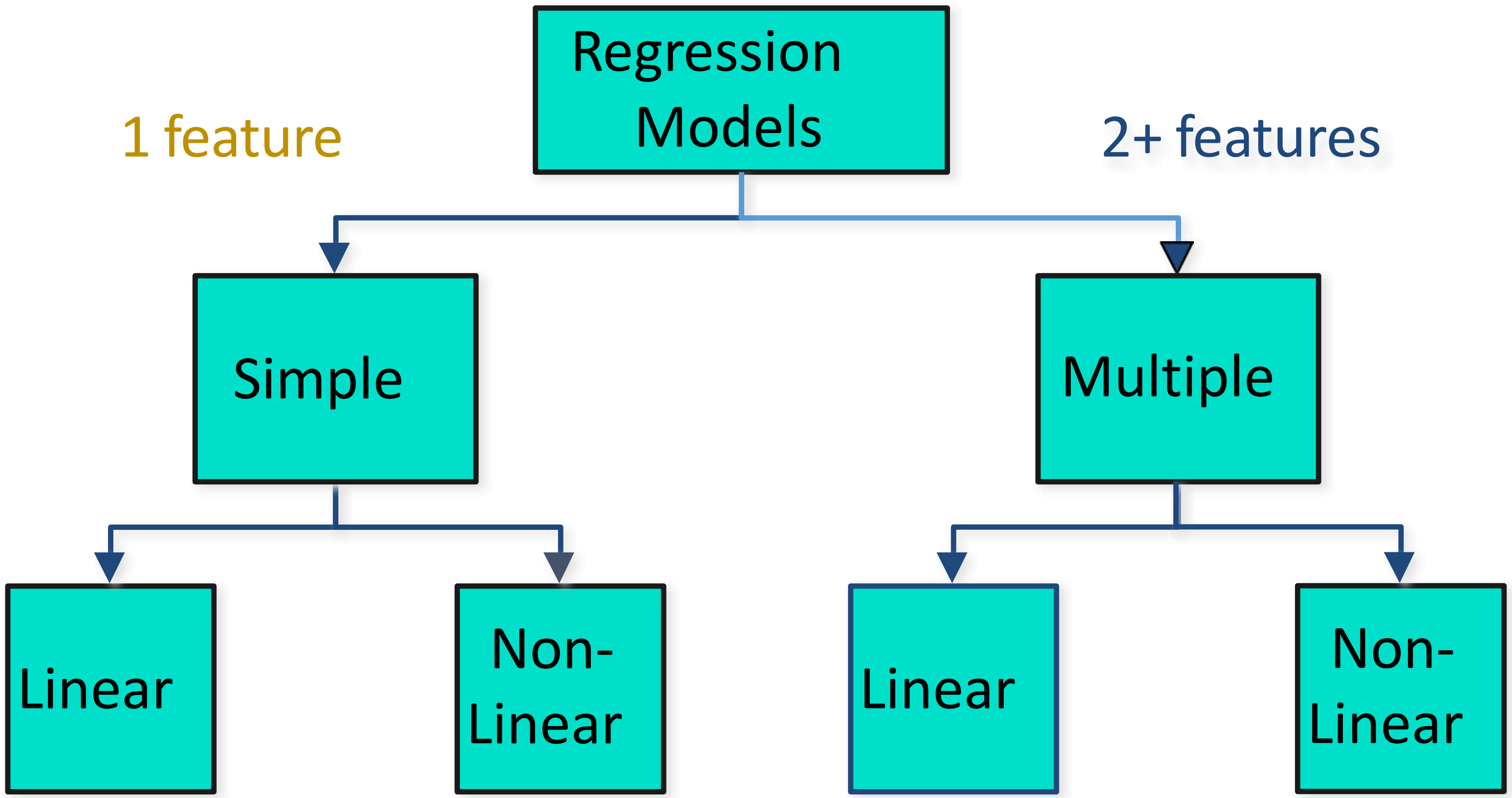
Polynomial Curve Fitting

- We obtain an excellent fit to the training data with 9th order.
- However, the fitted curve oscillates wildly and gives a very poor representation of the function sin(2πx).



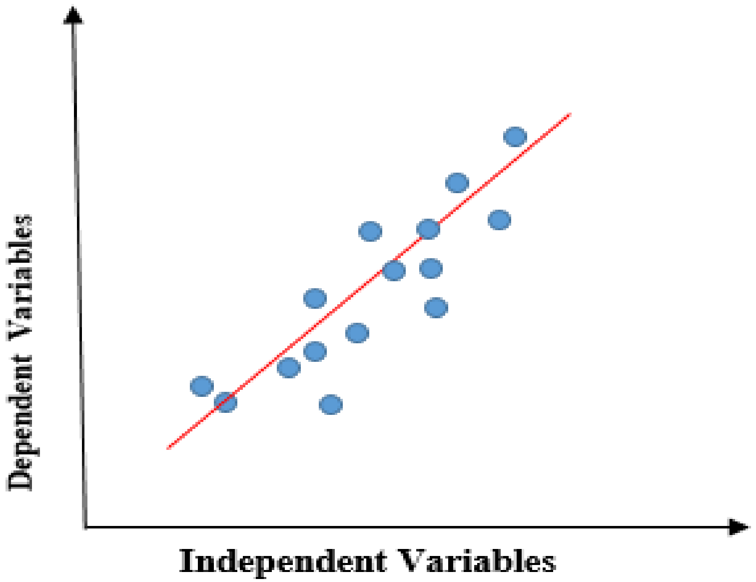
- This behaviour is known as **over-fitting**

Types of Regression Models



Liner Regression

- Linear regression is a quiet and simple statistical regression method used for predictive analysis and shows the relationship between the continuous variables.
- Linear regression shows the linear relationship between the independent variable (X-axis) and the dependent variable (Y-axis), consequently called linear regression.
- *If there is a single input variable (x), such linear regression is called **simple linear regression**. And if there is more than one input variable, such linear regression is called **multiple linear regression.***
- The linear regression model gives a sloped straight line describing the relationship within the variables.



- The above graph presents the linear relationship between the dependent variable and independent variables.
- When the value of x (**independent variable**) increases, the value of y (**dependent variable**) is likewise increasing. The red line is referred to as the best fit straight line.
- *To calculate best-fit line linear regression uses a traditional slope-intercept form.*

Data Model in simple Linear Regression

- Data is modelled using a straight line with continuous variable.
- Relationship between variables is a linear function.
- Linear equation representation with single feature is given:

Population y-intercept

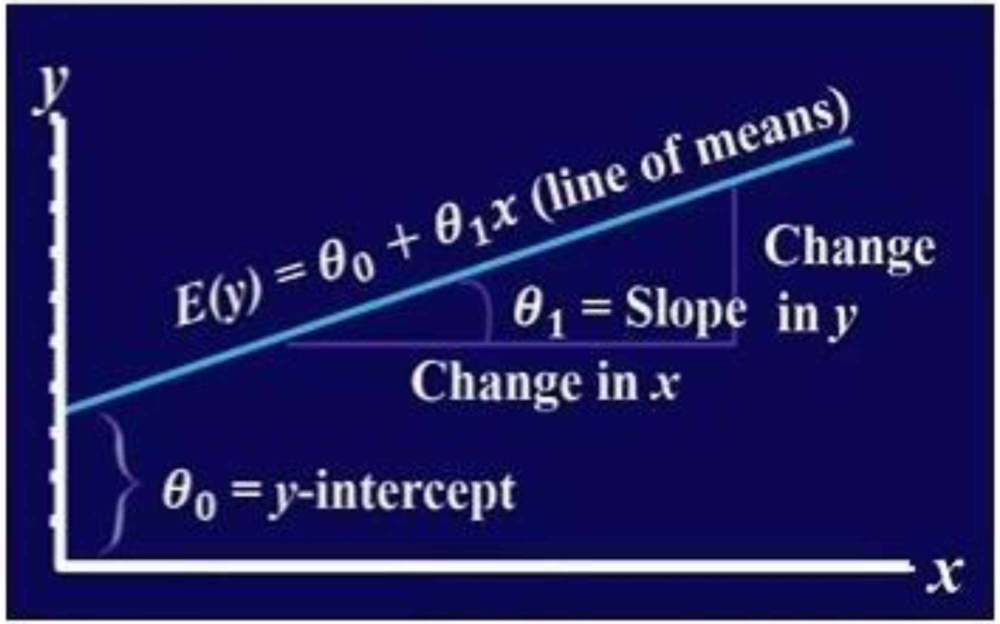
Population Slope

Random Error

$$y = \theta_0 + \theta_1 x + \varepsilon$$

Dependent (Response) Variable

Independent (Explanatory) Variable



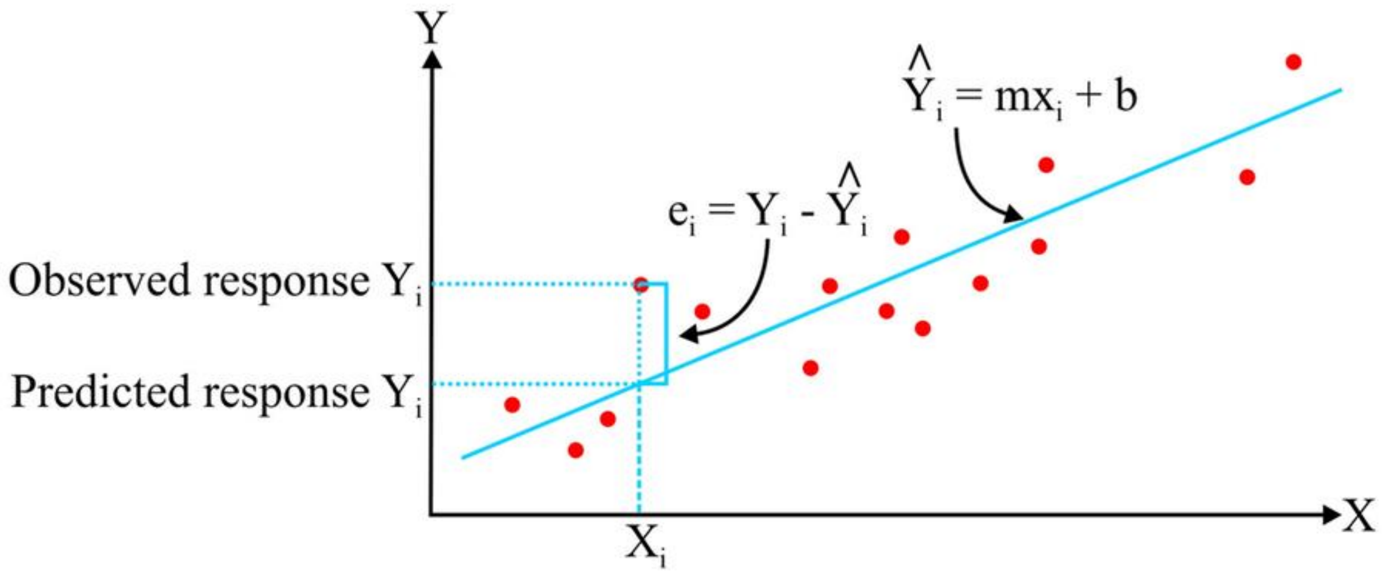
- It is reduced into

$$y' = h_\theta(x) = \theta^T.X$$

Least Squares method – Linear regression

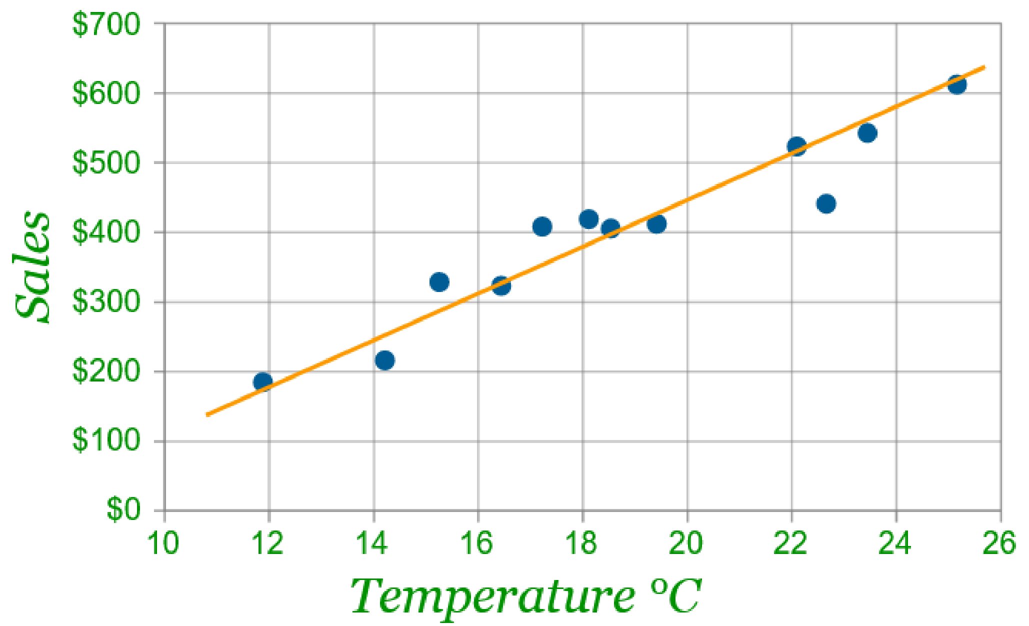
- least squares is a statistical method for determining the best fit line for given data in the form of an equation such as y=mx+b.
- This method is frequently used in data fitting, where the best fit result is supposed to reduce the sum of squared errors, which is defined as the difference between observed and fitted values.
- The sum of squared errors helps identify variation in observable data.
- The method of least squares analysis begins with a set of data points to be plotted on the graph of the XY plane.
- The goal of this method is to minimise the sum of squared errors as much as possible.
- least squares regression line best fits a linear relationship between two variables by minimising the vertical distance between the data points and the regression line

What is Least Squares method?



Line of Best Fit

- Imagine you have some points, and want to have a **line** that best fits them like this:



The Line

Our aim is to calculate the values **m** (slope) and **b** (y-intercept) in the equation of a line :

y = mx + b

Where:

- **y** = how far up
- **x** = how far along
- **m** = Slope or Gradient (how steep the line is)
- **b** = the Y Intercept (where the line crosses the Y axis)

Steps

Step 1: For each (x,y) point calculate x^2 and xy

Step 2: Sum all x, y, x^2 and xy, which gives us Σx, Σy, Σx^2 and Σxy

Step 3: Calculate Slope **m**:

$$\mathbf{m} = \frac{N\,\Sigma(xy) - \Sigma x\,\Sigma y}{N\,\Sigma(x^2) - (\Sigma x)^2}$$

(N is the number of points.)

Step 4: Calculate Intercept **b**:

$$\mathbf{b} = \frac{\Sigma y - m\,\Sigma x}{N}$$

Step 5: Assemble the equation of a line

$$y = mx + b$$

Example

Sam found how many **hours of sunshine** vs how many **ice creams** were sold at the shop from Monday to Friday.

"x" Hours of Sunshine	"y" Ice Creams Sold
2	4
3	5
5	7
7	10
9	15

- find the best **m** (slope) and **b** (y-intercept) that suits that data y = mx + b

Step 1: For each (x,y) calculate x^2 and xy:

x	y	x^2	xy
2	4	4	8
3	5	9	15
5	7	25	35
7	10	49	70
9	15	81	135

Step 2: calculate Σx, Σy, Σx^2 and Σxy

x	y	x^2	xy
2	4	4	8
3	5	9	15
5	7	25	35
7	10	49	70
9	15	81	135
Σx: 26	**Σy: 41**	**Σx^2: 168**	**Σxy: 263**

Step 3: Calculate Slope m

$$\mathbf{m} = \frac{N\,\Sigma(xy) - \Sigma x\,\Sigma y}{N\,\Sigma(x^2) - (\Sigma x)^2}$$

$$= \frac{5 \text{ x } 263 - 26 \text{ x } 41}{5 \text{ x } 168 - 26^2}$$

$$= \frac{1315 - 1066}{840 - 676}$$

$$= \frac{249}{164} = 1.5183\ldots$$

Step 4: Calculate Intercept b

$$\mathbf{b} = \frac{\Sigma y - m\,\Sigma x}{N}$$

$$= \frac{41 - 1.5183 \text{ x } 26}{5}$$

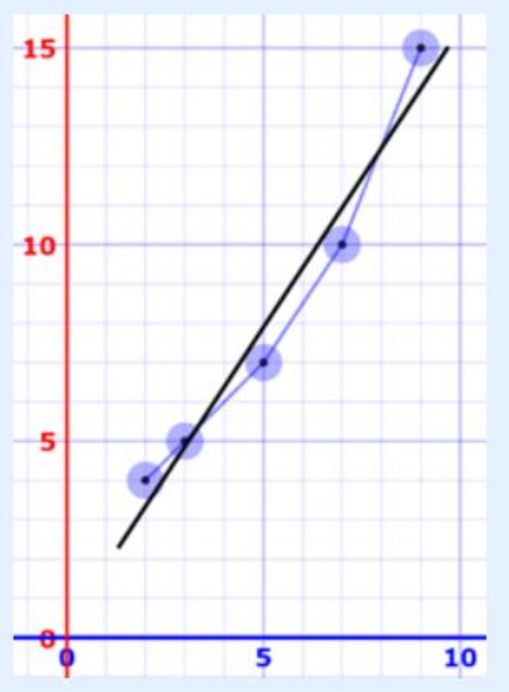
$$= 0.3049\ldots$$

Step 5: Assemble the equation of a line

$$y = mx + b$$

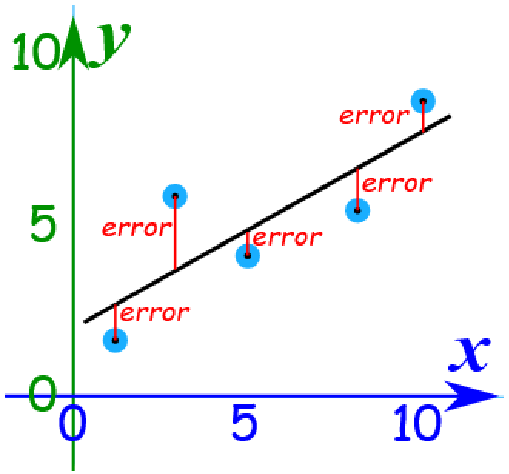
$$y = 1.518x + 0.305$$

x	y	y = 1.518x + 0.305	error
2	4	3.34	−0.66
3	5	4.86	−0.14
5	7	7.89	0.89
7	10	10.93	0.93
9	15	13.97	−1.03



How does it work?

- It works by making the total of the **square of the errors** as small as possible (that is why it is called "least squares"):



Example 2

- Ram's parents are interested to guess the 6th std gpa from the gpa's scores from 1-5 classes.

Class	GPA
1	8.6
2	7.2
3	9.4
4	7.7
5	8.0

- Predict 6th class GPA of Ram?