Curve Fitting

Dr. Umesh Aundhakar

Objective:

- The purpose of performing curve fits.
- Different types of curve fits.
- Choosing an appropriate curve fit.

Purpose of Curve Fitting

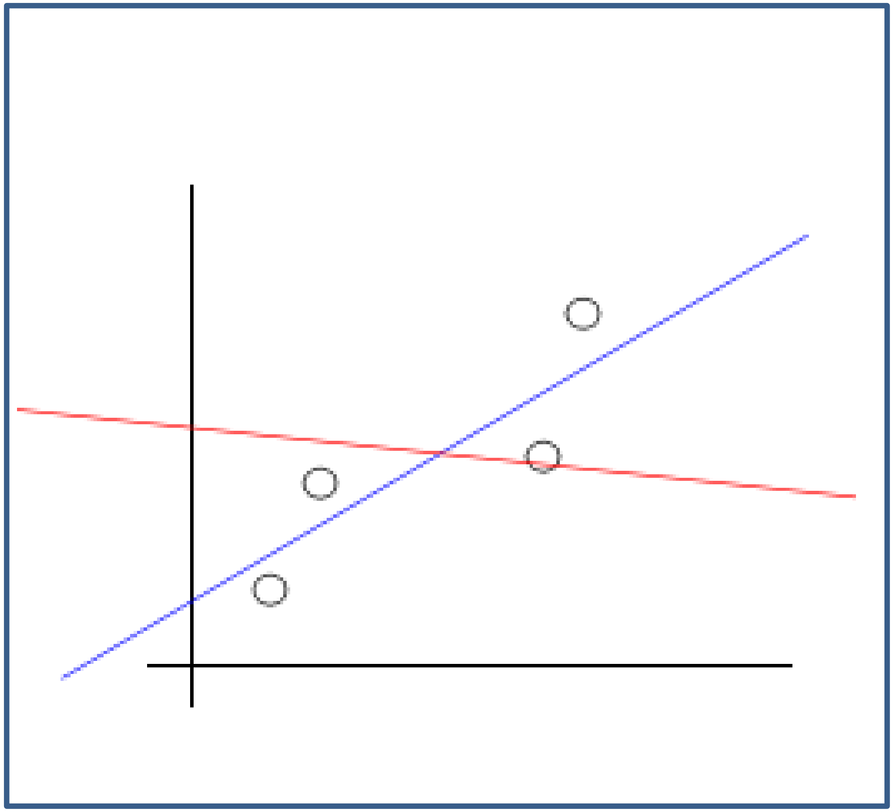
Curve fitting, also known as regression analysis. It is used to find the "best fit" **line** or **curve** for a series of data points.

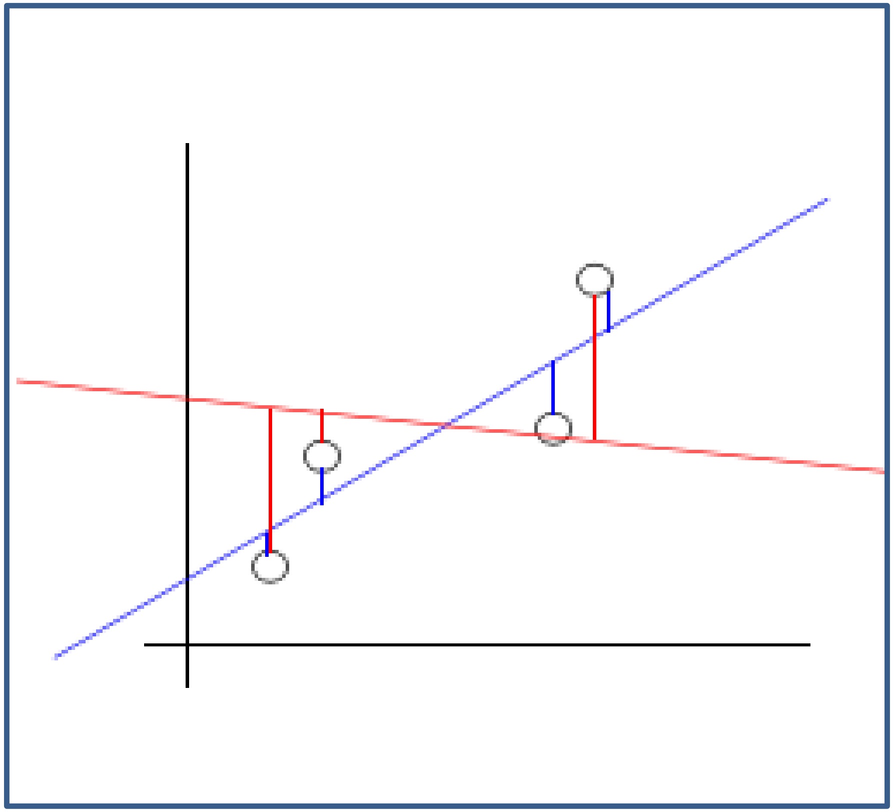
Curve fit will produce an equation that can be used to find points anywhere along the curve.

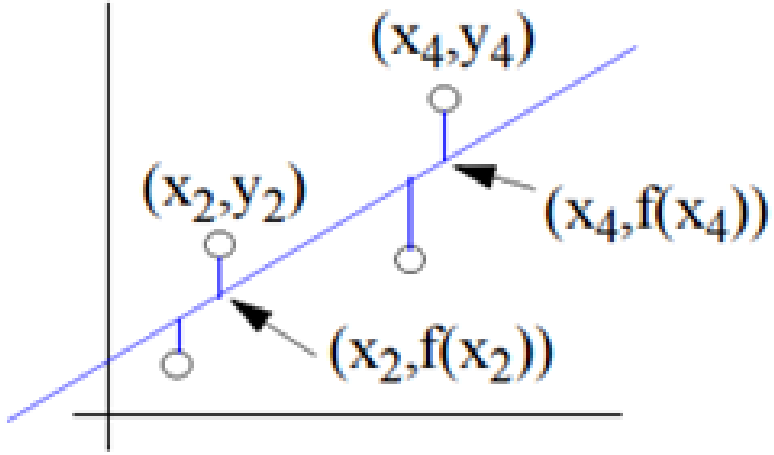
Sometimes, you may just want to use a curve fit to smooth the data and improve the appearance of your plot.

- Straight line fitting
- Parabola fitting
- Exponential curve fitting

Method of least Square







$$err = \sum (d_i)^2 = (y_1 - f(x_1))^2 + (y_2 - f(x_2))^2 + (y_3 - f(x_3))^2 + (y_4 - f(x_4))^2$$

$$\text{minimize} \quad err = \sum_{i=1}^{\#\text{ data points} = n} (y_i - (ax_i + b))^2$$

The ‘best’ line has minimum error between line and data points. This is called the least squares approach, since we minimize the square of the error.

Straight line fitting

- Fit a linear equation,

$$y = a + bx$$

- Normal equations are,

$$na + b\sum x = \sum y$$

$$a\sum x + b\sum x^2 = \sum xy$$

Parabola fitting

- Fit a Quadratic equation,

$$y = a + bx + cx^2$$

- Normal equations are,

$$na + b\sum x + c\sum x^2 = \sum y$$

$$a\sum x + b\sum x^2 + c\sum x^3 = \sum xy$$

$$a\sum x^2 + b\sum x^3 + c\sum x^4 = \sum x^2 y$$

Exponential Line fitting

- Fit a linear equation,

$$y = ab^x$$

- Normal equations are,

$$nA + B\sum x = \sum Y$$

$$A\sum x + B\sum x^2 = \sum xY$$

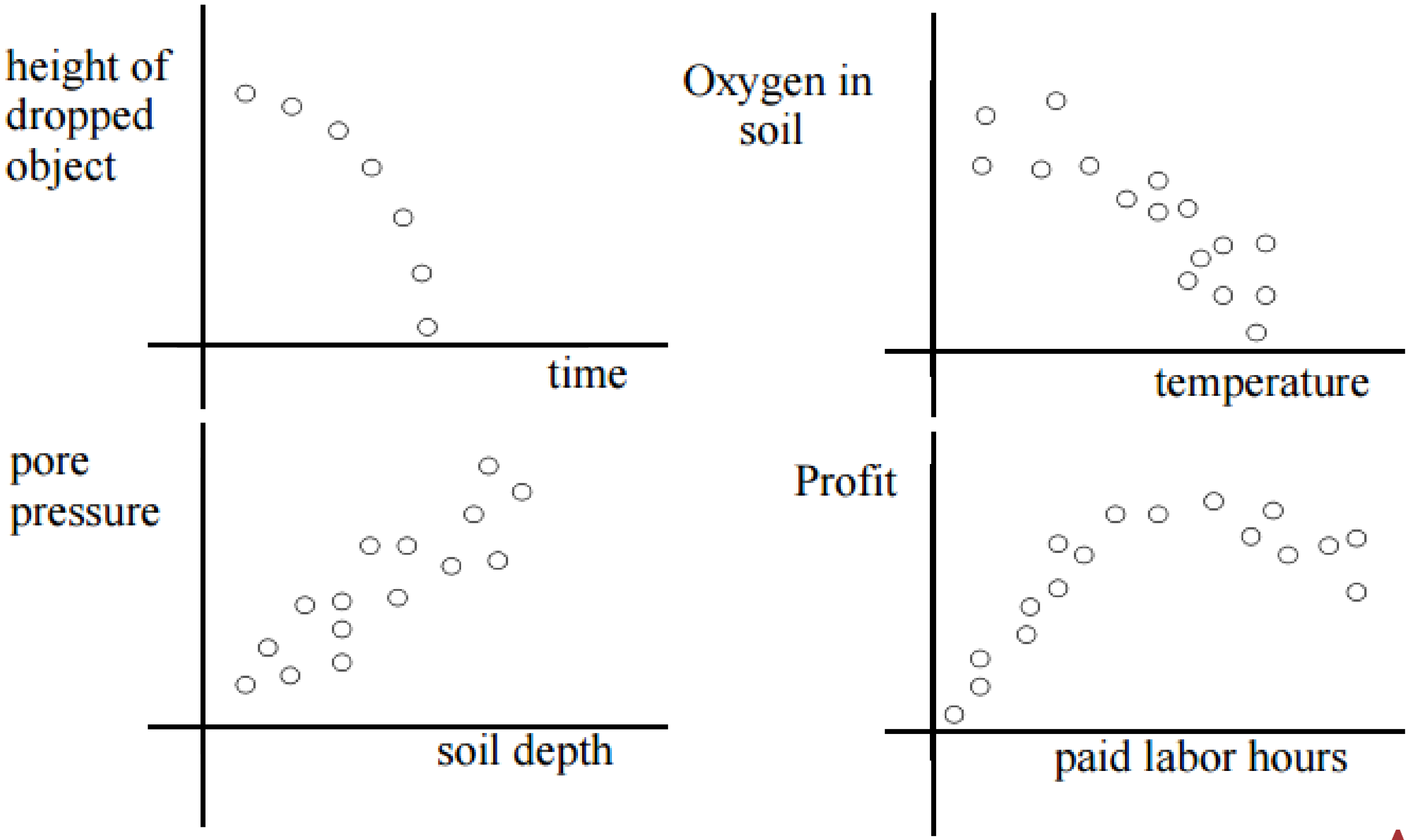
Where, $A = \log a$

$B = \log b$

$Y = \log y$

Aegis
SCHOOL OF BUSINESS
SCHOOL OF DATA SCIENCE
SCHOOL OF TELECOMMUNICATION

Is a straight line suitable for each of these cases ???



How one can decide the fitted curve is good fit to the data set?

What is R-squared?

The goodness of fit is shown as an R^2-value. A value of $R^2=1.0$ indicates a perfect fit, whereas $R^2=0.0$ indicates that the fitted curve might be unsuitable for the given data.

R^2 :

$$R2 = \frac{SSR}{SST} = 1.0 - \frac{SSE}{SST} \qquad 0 \le R2 \le 1$$

where

$$SSR = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2$$ (the regression sum of squares)

$$SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$ (the residual or error sum of squares)

$$SST = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$ (the total sum of squares, SST= SSE+SSR)

and $\hat{Y}_i$ represents the i^{th} fitted of the dependent variable Y.

Graphical Explanation: