NUMERICAL METHODS

$$U^{n+1} = U^n + \Delta t\, f(U^n)$$

$$\frac{\partial v}{\partial t} + V \cdot \nabla v = \nabla \cdot (k \nabla v) + g(v)$$

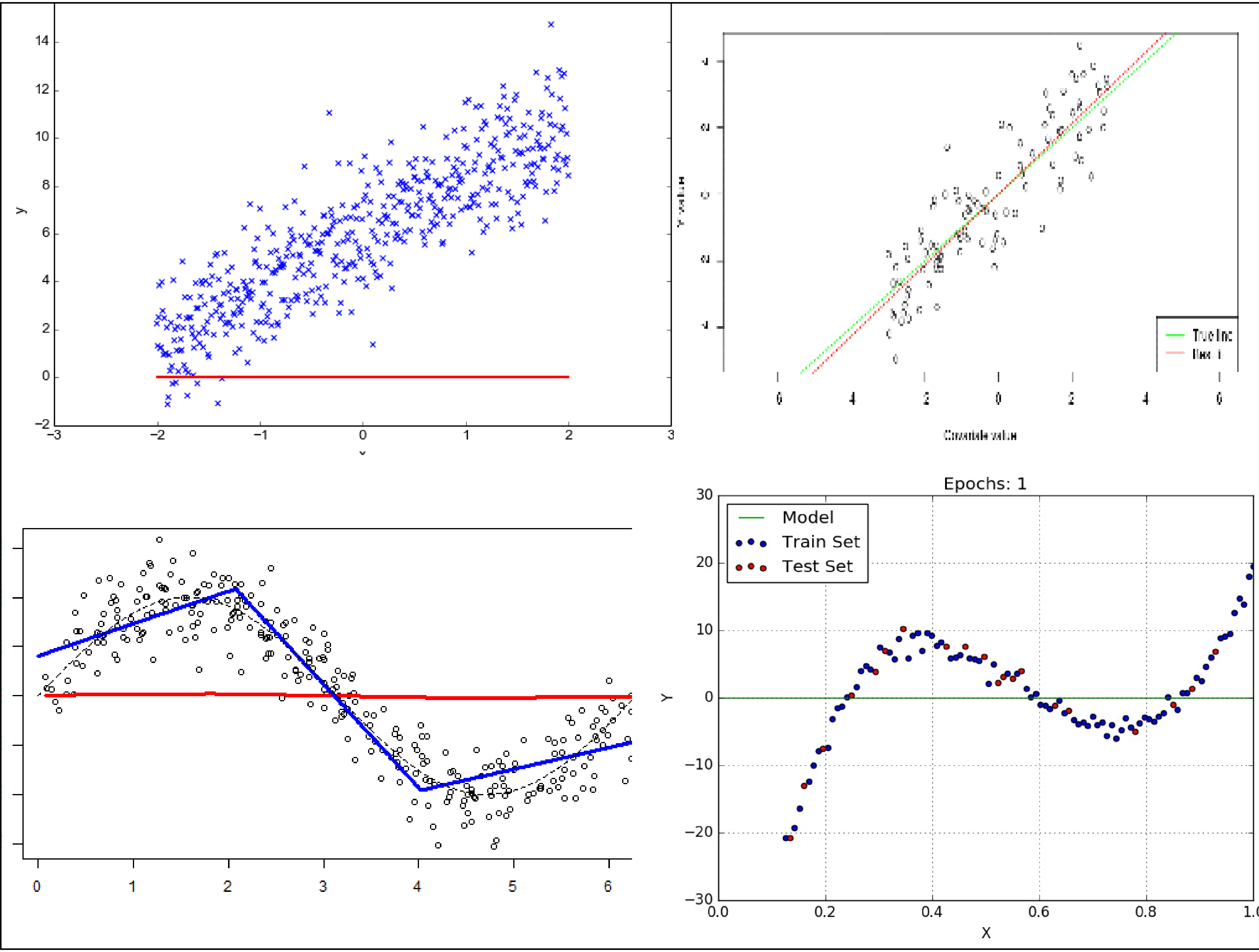
$$(\lambda + \mu)\nabla(\nabla \cdot u) + \mu \nabla^2 u = \alpha(3\lambda + 2\mu)\nabla T - \rho b$$

Lecture 8

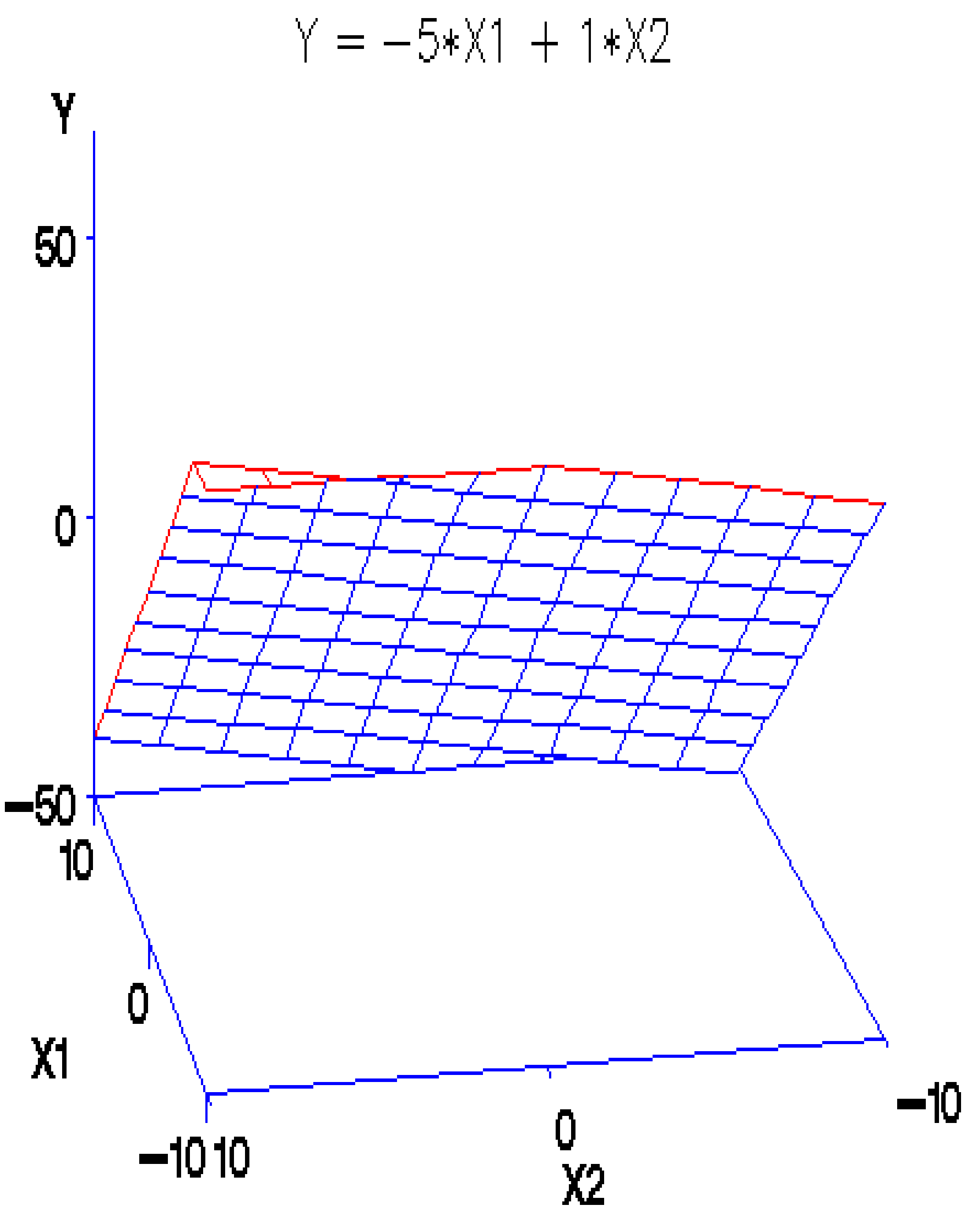
$$\rho\left(\frac{\partial u}{\partial t} + V \cdot \nabla u\right) = -\frac{\partial p}{\partial x} + \frac{\partial \tau_{xx}}{\partial x} + \frac{\partial \tau_{yx}}{\partial y} + \frac{\partial \tau_{zx}}{\partial z} + \rho f_x$$

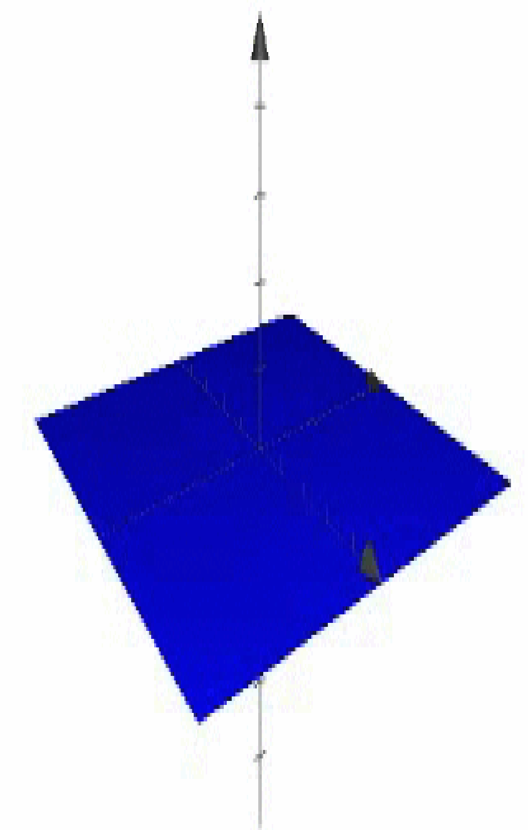
$$\nabla^2 u = f$$

Dr. P V Ramana



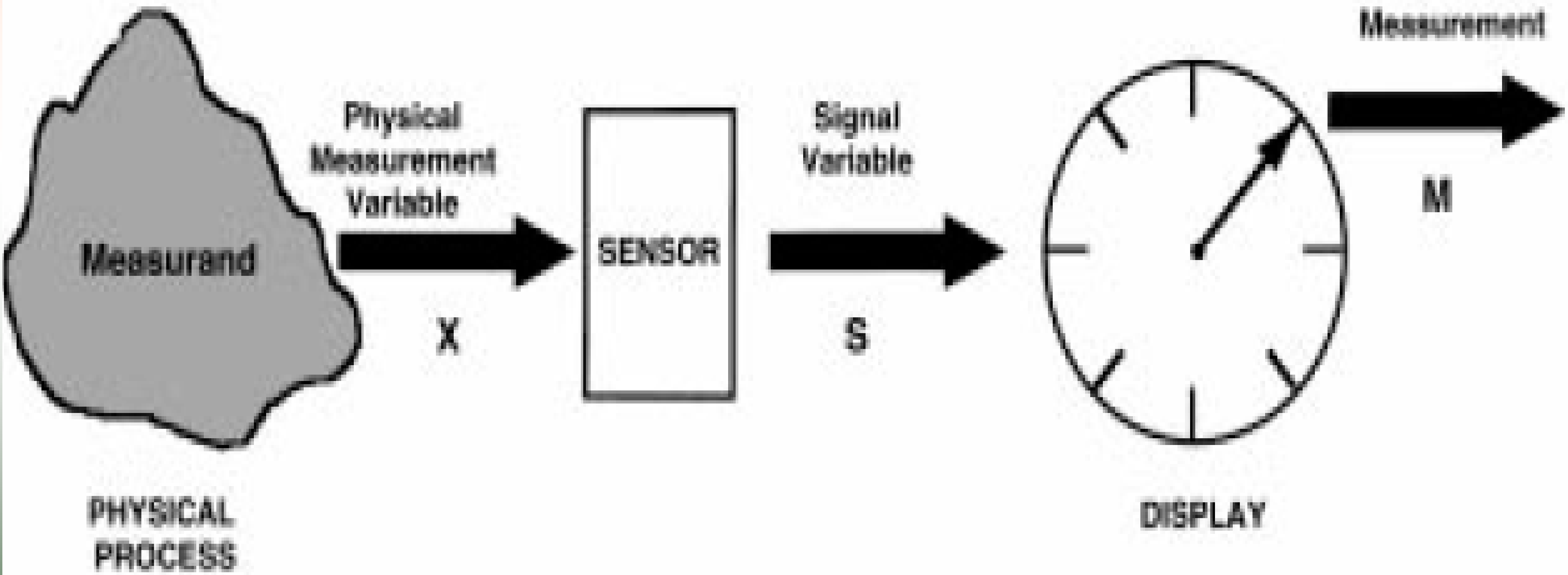
y
x
Epochs: 1
Model
Train Set
Test Set
X
Y


Y = −5*X1 + 1*X2
Y
50
0
−50
10
0
X1
−10 10
0
X2
−10



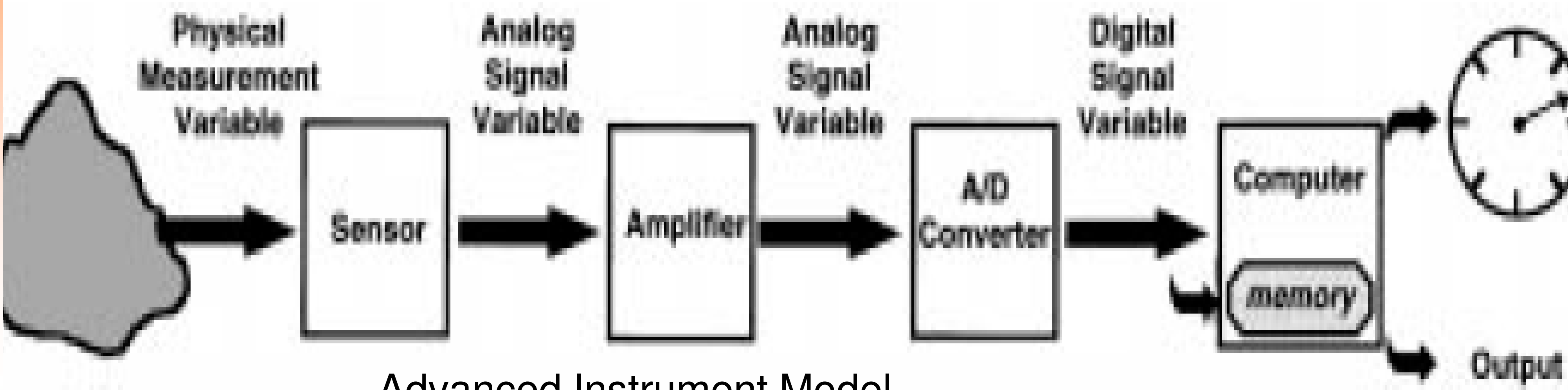
Instrument & Calibration

An *instrument* is a device that transforms a *physical variable* of interest (the *measurand*) into a form that is suitable for recording (the *measurement*).



Simple Instrument Model

Instrument & Calibration



Advanced Instrument Model

Sensors

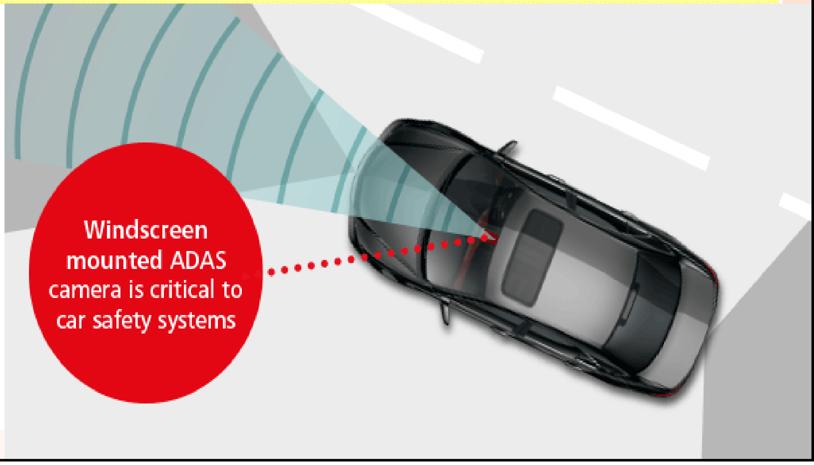
- **Sensors convert physical variables to signal variables.**
- **Sensors are often transducers : They are devices that convert input energy of one form into output energy of another form.**
- **Sensors can be categorized into three broad classes depending on how they interact with the environment they are measuring.**
- ***Passive sensors:* do not add energy as part of the measurement process but may remove energy in their operation.**
- ***Active sensors :* add energy to the measurement environment as part of the measurement process.**
- ***Hybrid sensors :* Mixed energies.**

Calibration

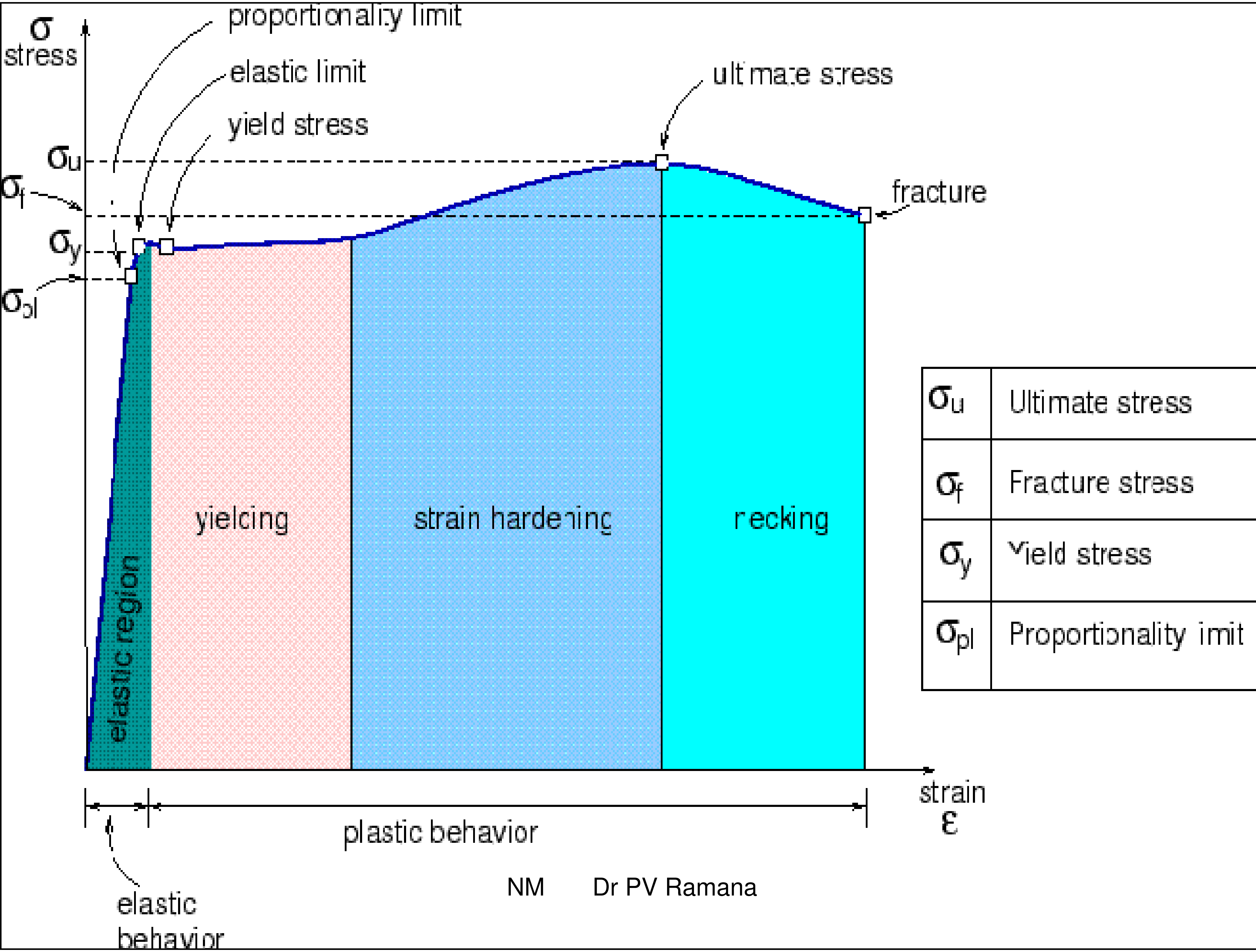
- The process of development of a relationship between the physical measurement variable input and the signal variable (output) for a specific sensor is known as the *calibration* of the sensor.

- **Typically, a sensor (or an entire instrument system) is calibrated by providing a known physical input to the system and recording the output.**

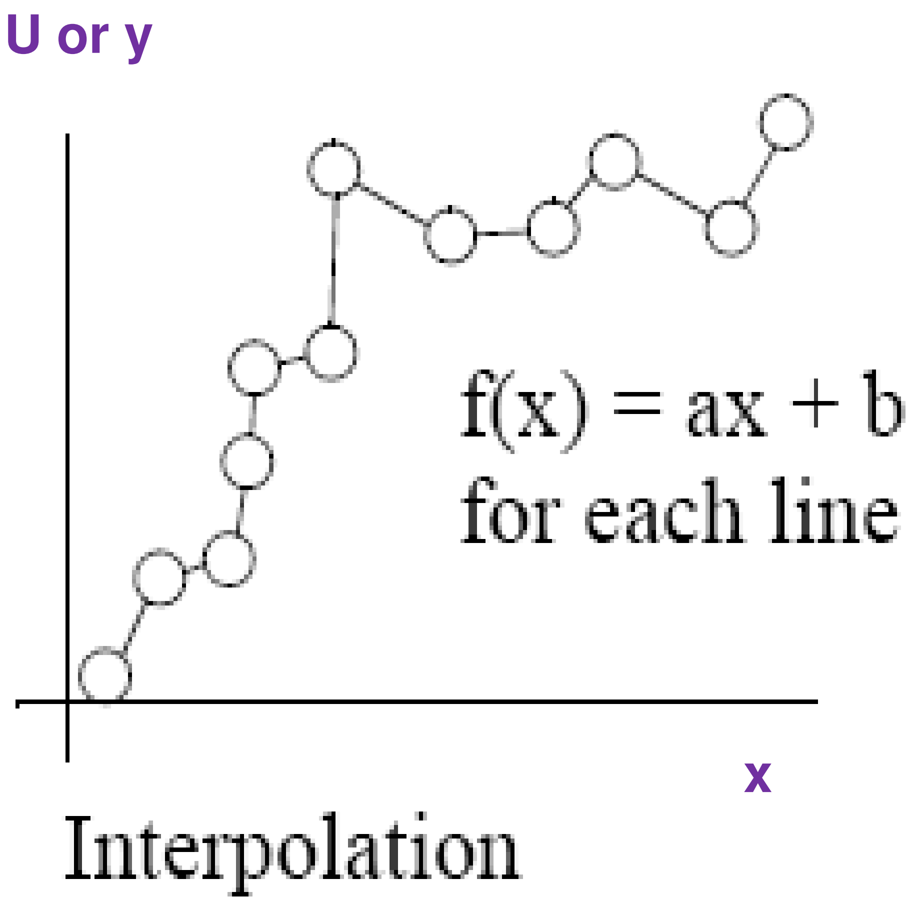
- **The data are plotted on a calibration curve.**

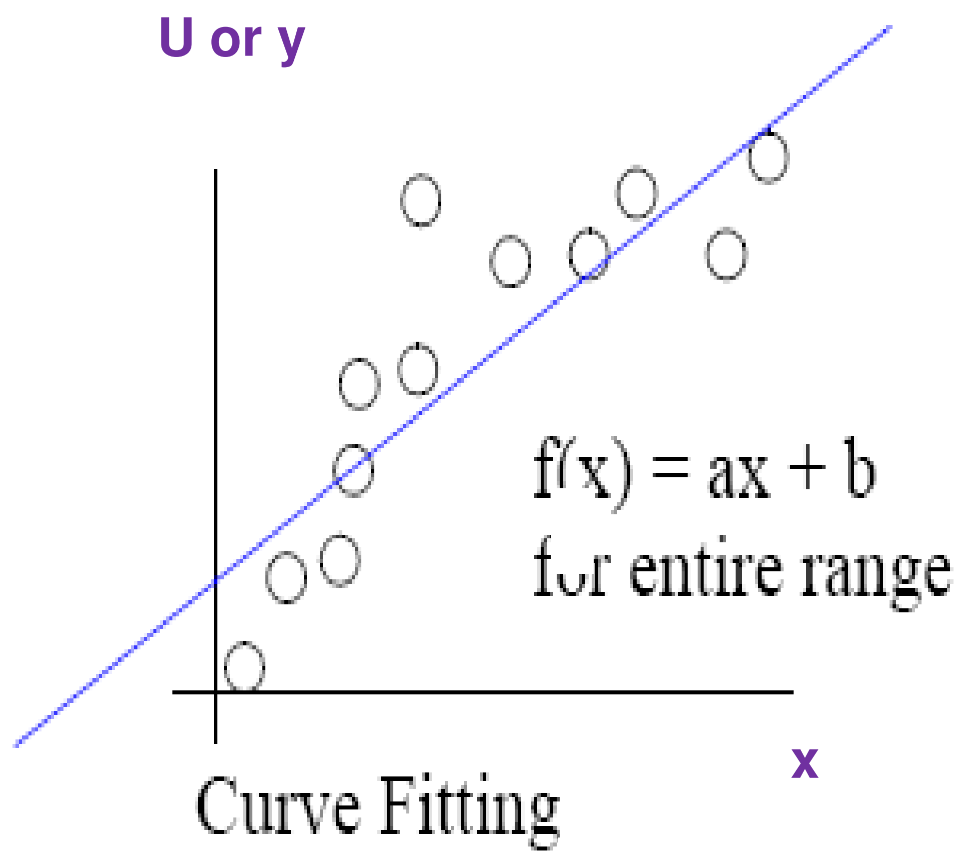


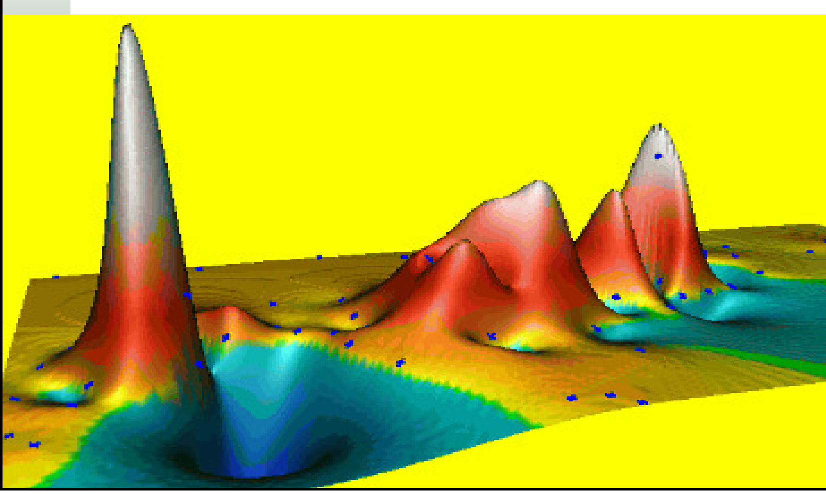

TIAN YUAN
PresenterMedia
E-STORE

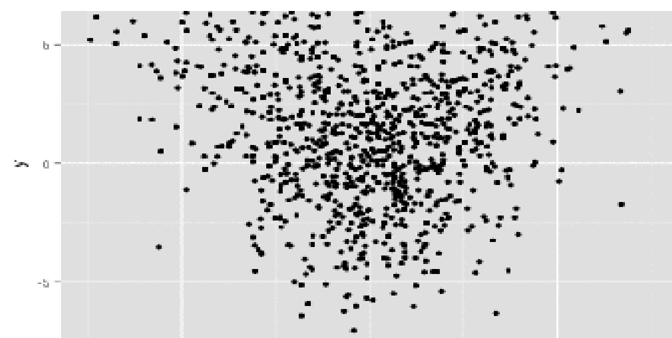

σ
stress
proportionality limit
elastic limit
yield stress
ultimate stress
σu
σf
σy
σpl
fracture
yielding
strain hardening
necking
elastic region
σu Ultimate stress
σf Fracture stress
σy Yield stress
σpl Proportionality limit
strain
ε
plastic behavior
elastic behavior

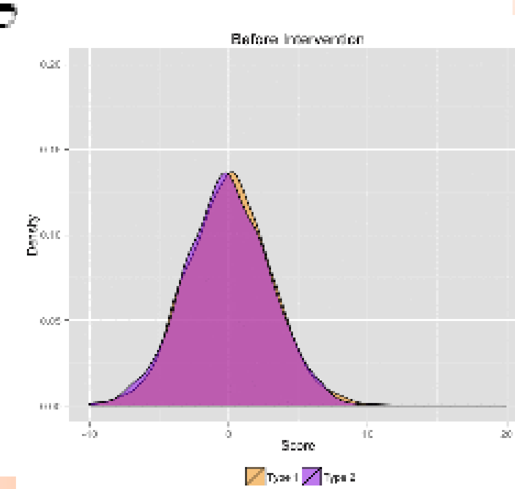
Interpolation Vs Curve Fitting





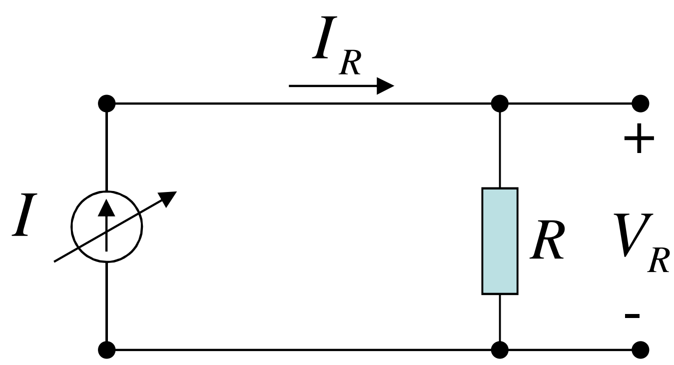


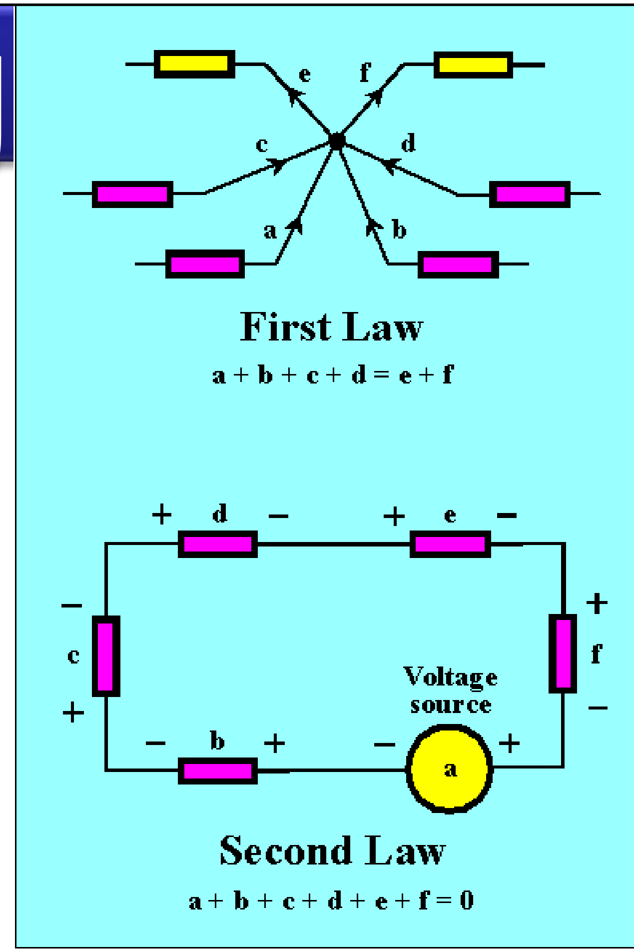




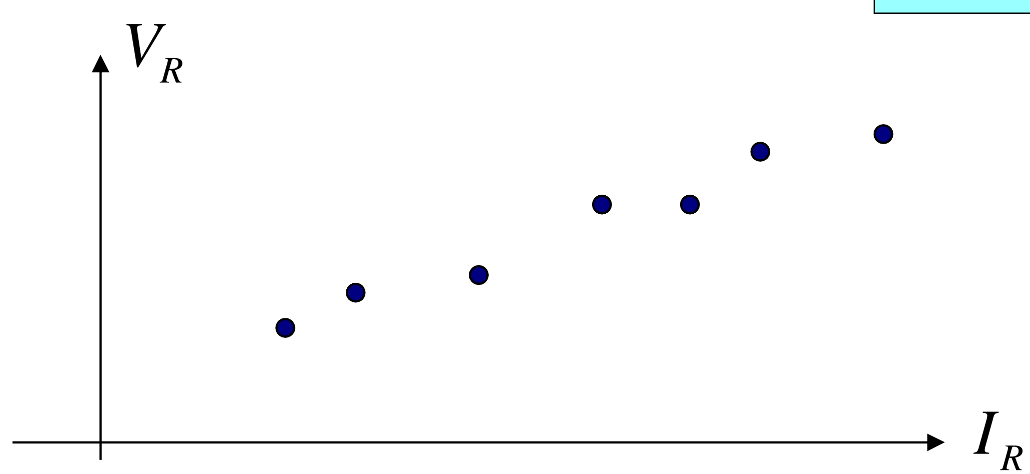
Curve Fitting

Introduction



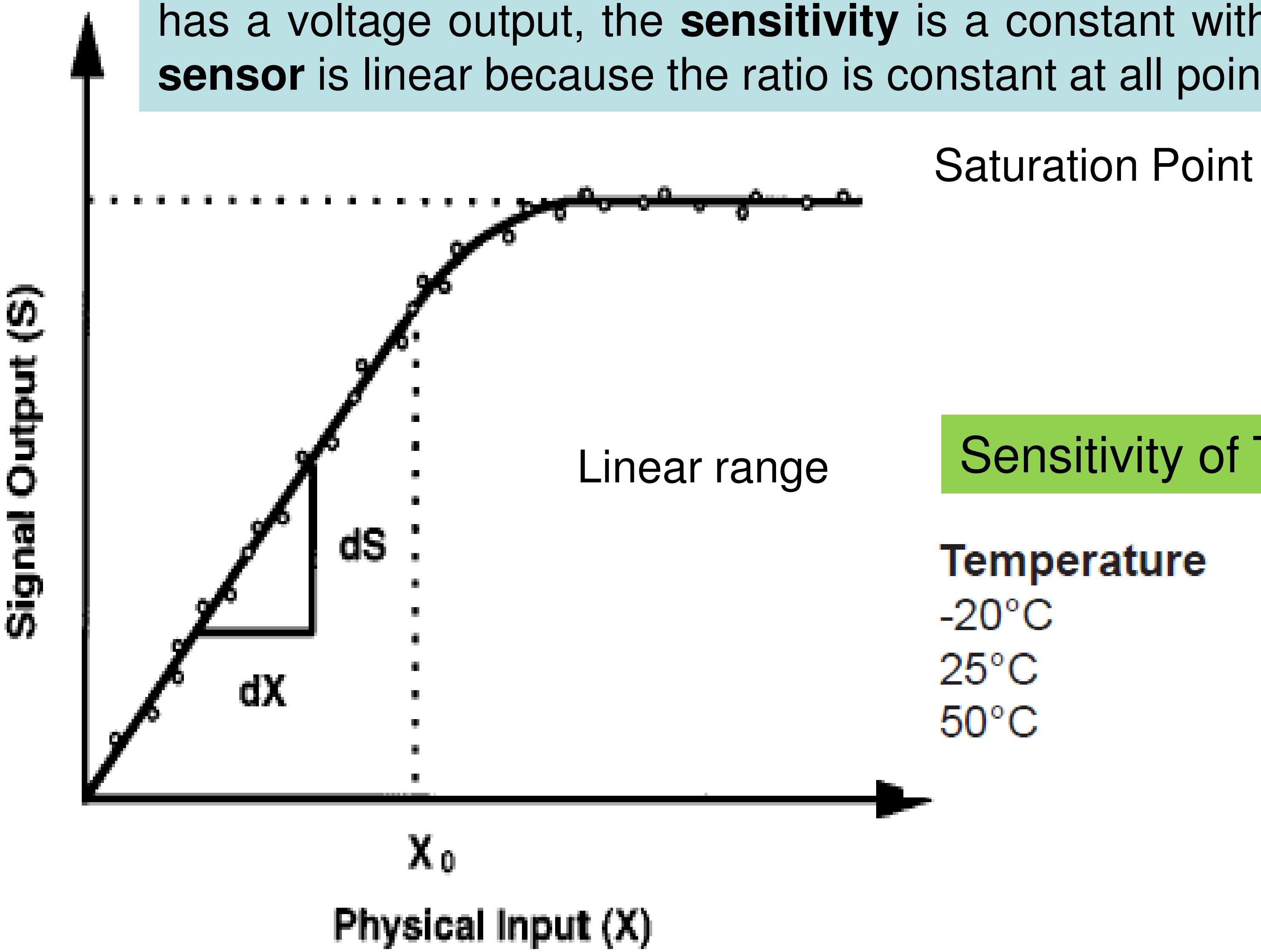


Experimental data:



Sensitivity of A Sensor

The **sensitivity** is then defined as the ratio between the output signal and measured property. For example, if a **sensor** measures temperature and has a voltage output, the **sensitivity** is a constant with the unit [V/K]; this **sensor** is linear because the ratio is constant at all points of measurement.



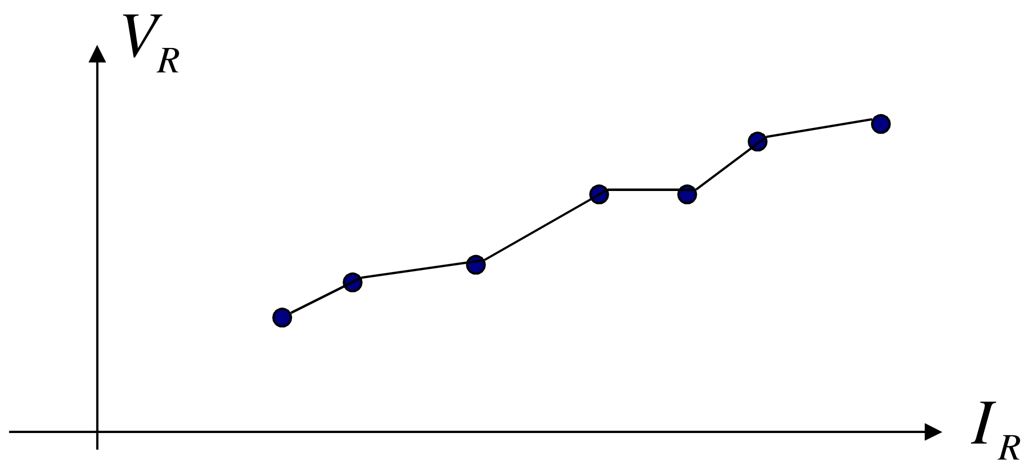
Sensitivity of Thermistor

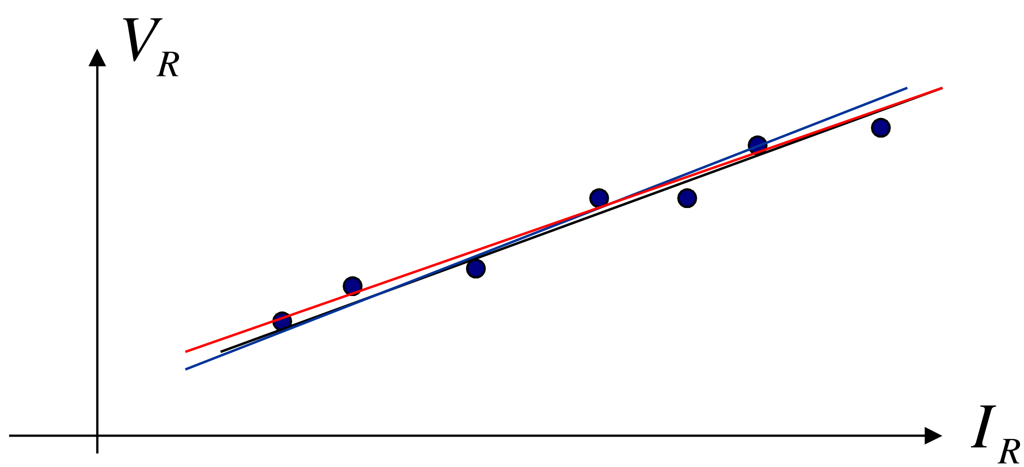
Temperature	Sensitivity
-20°C	5,600 Ω/°C
25°C	439 Ω/°C
50°C	137 Ω/°C

Curve Fitting

Introduction

Experimental data:



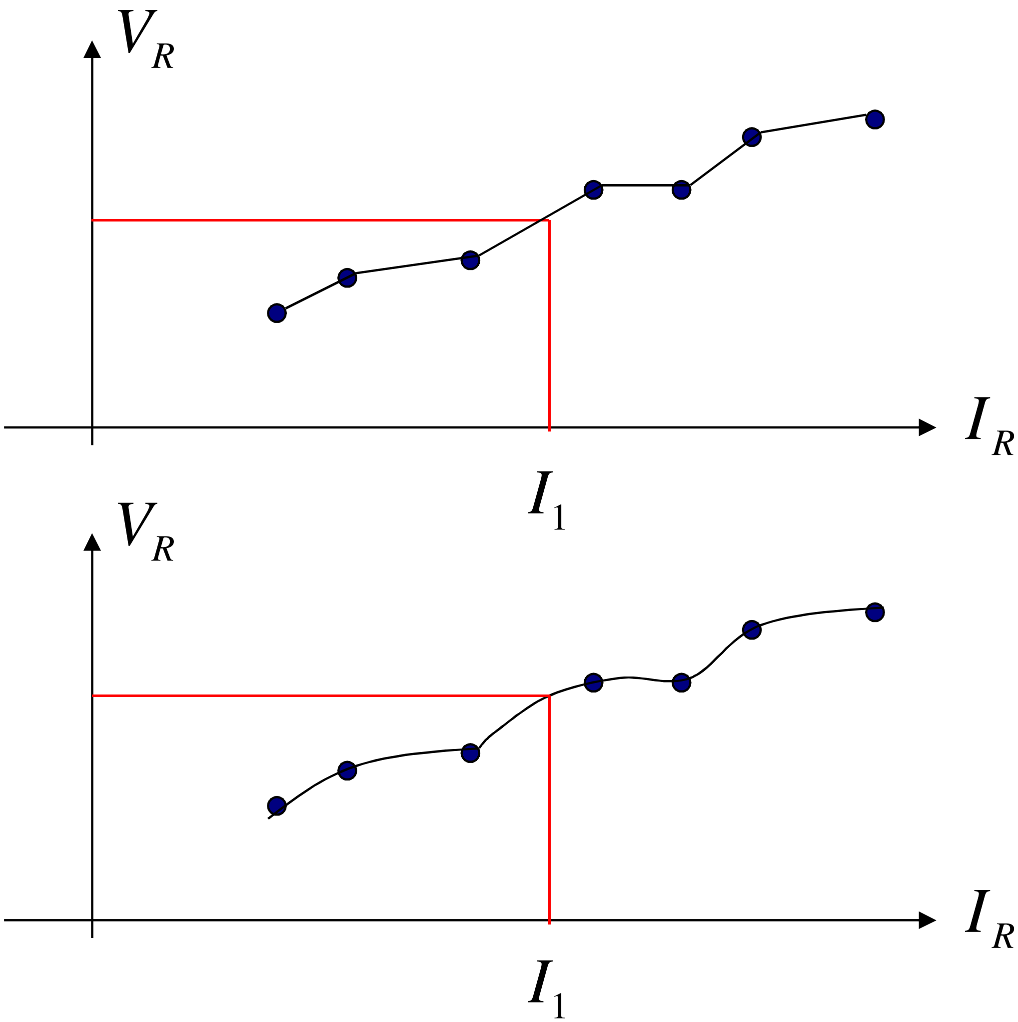


Least-squares regression

Interpolation

Introduction

Experimental data:



Linear Interpolation

Curvilinear Interpolation

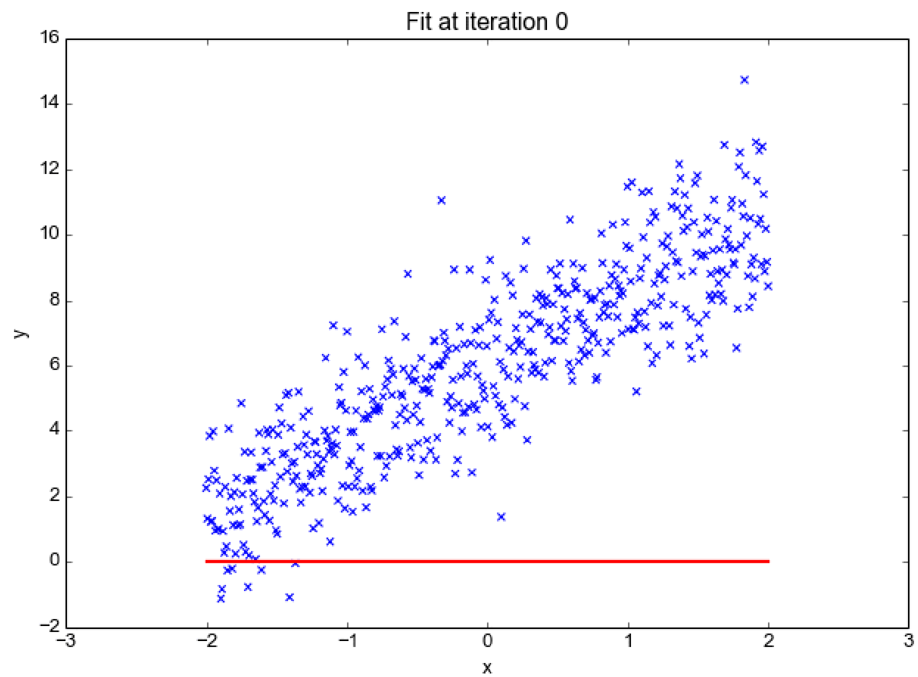
Curve Fitting

Importance of curve fitting

- **Trend analysis**

  - **- prediction -- interpolation**

  - **- forecasting -- extrapolation**



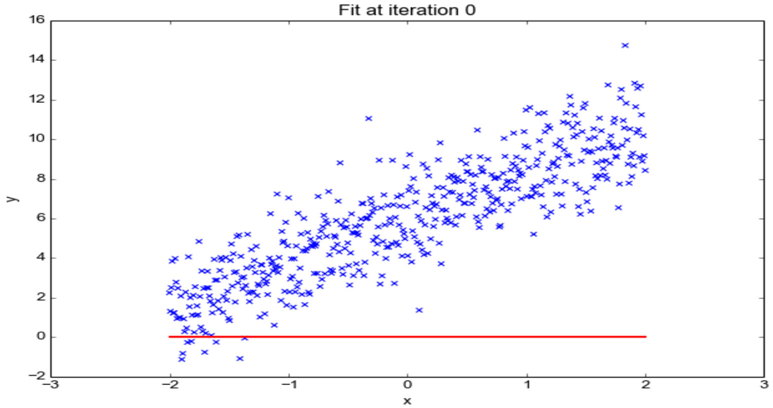
- **Hypothesis testing**

  - **- compare measured data with existing models**

Mathematical Background

- ***Arithmetic mean***. The sum of the individual data points (y_i) divided by the number of points (n).

$$\bar{y} = \frac{\sum y_i}{n}, \; i = 1, \ldots, n$$



- ***Standard deviation***. **The most common measure of a spread for a sample.**

$$S_y = \sqrt{\frac{S_t}{n-1}}, \quad S_t = \sum (y_i - \bar{y})^2$$

$$S_y = \frac{y_i - \bar{y}}{\sqrt{n-1}}$$

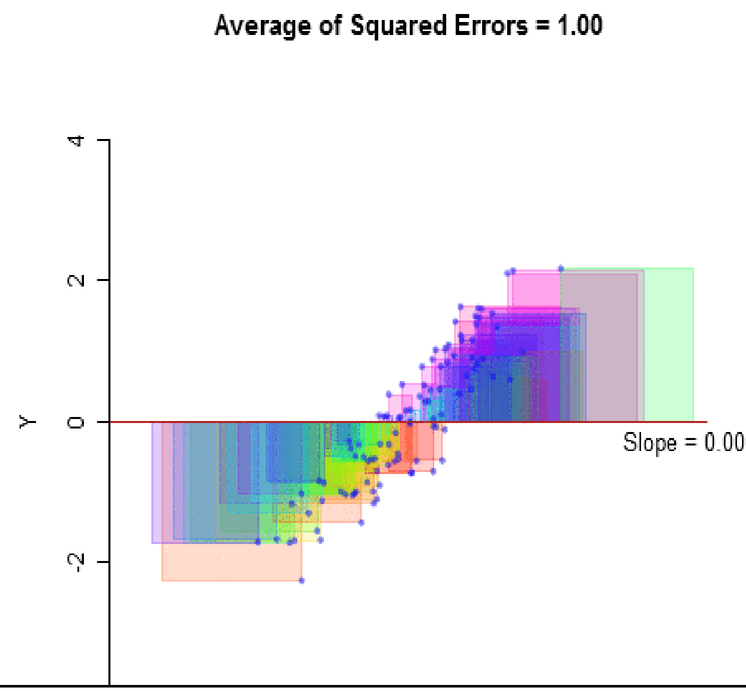
Mathematical Background

- ***Variance***. Representation of spread by the square of the standard deviation.

$$S_y^2 = \frac{\sum (y_i - \bar{y})^2}{n-1} \quad \text{or} \quad S_y^2 = \frac{\sum y_i^2 - \left(\sum y_i\right)^2 / n}{n-1}$$

- ***Coefficient of variation***. **Has the utility to quantify the spread of data.**

$$c.v. = \frac{S_y}{\bar{y}} 100\%$$



Curve Fitting

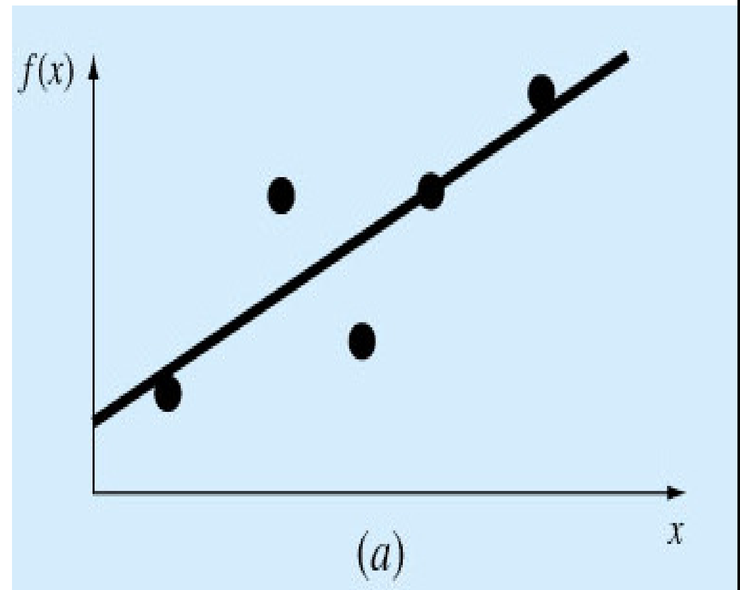
- Fit the best curve to a discrete data set and obtain estimates for other data points

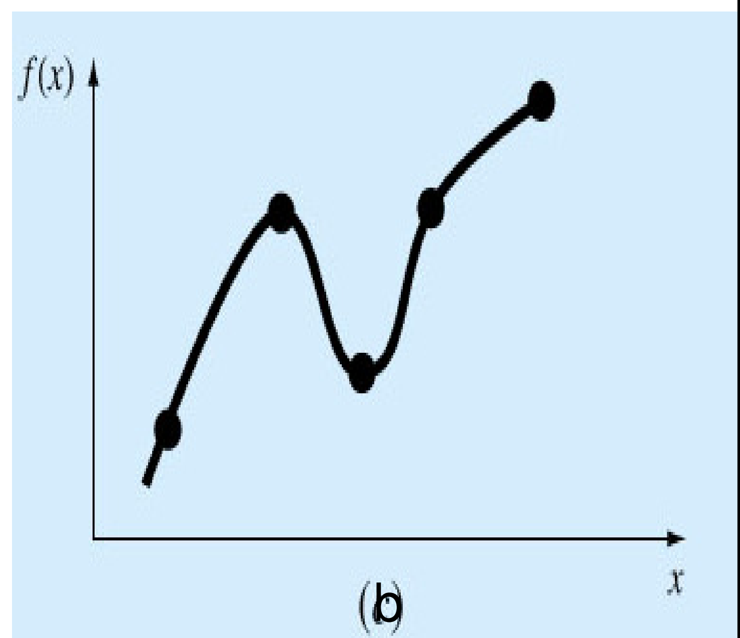
- Two general approaches:
  - *Data exhibit a significant degree of **scatter*** Find a single curve that represents the general trend of the data.
  - *Data is very **precise***. Pass a curve(s) exactly through each of the points.

- Two common applications in engineering:

Trend analysis. Predicting values of dependent variable: ***extrapolation*** beyond data points or ***interpolation*** between data points.

Hypothesis testing. Comparing existing mathematical model with measured data.





Simple Statistics

In sciences, if several measurements are made of a particular quantity, additional insight can be gained by summarizing the data in one or more well chosen statistics:

Arithmetic mean - The sum of the individual data points (y_i) divided by the number of points.

$$\bar{y} = \frac{\sum y_i}{n} \qquad i = 1, \ldots, n$$

Standard deviation – a common measure of spread for a sample

$$S_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n-1}}$$

or *variance*

$$S_y^2 = \frac{\sum (y_i - \bar{y})^2}{n-1}$$

Coefficient of variation –

$$c.v. = \frac{S_y}{\bar{y}} 100\%$$

quantifies the spread of data (similar to relative error)

Linear Regression

- Fitting a **straight line** to a set of paired observations: $(x_1, y_1), (x_2, y_2),\ldots,(x_n, y_n)$

y_i : **measured value**

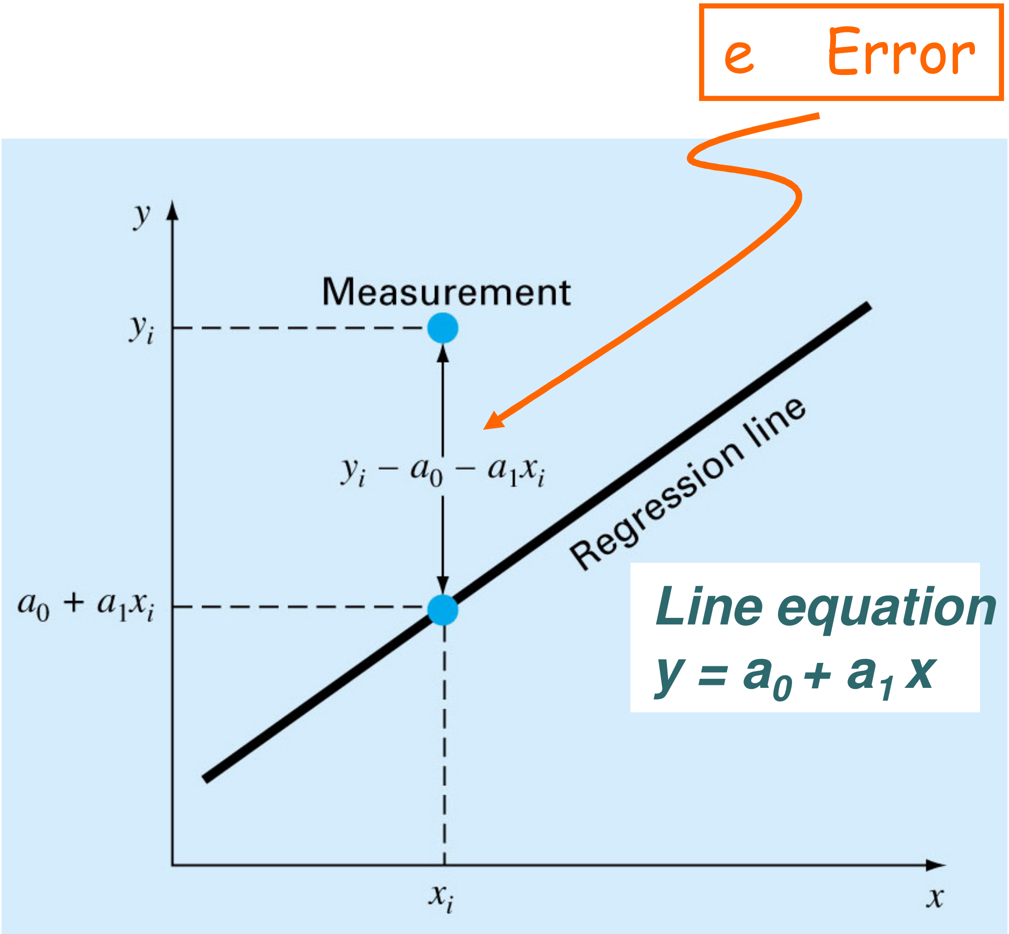
e : **error**

$$y_i = a_0 + a_1 x_i + e$$

$$e = y_i - a_0 - a_1 x_i$$

a_1 : slope

a_0 : intercept



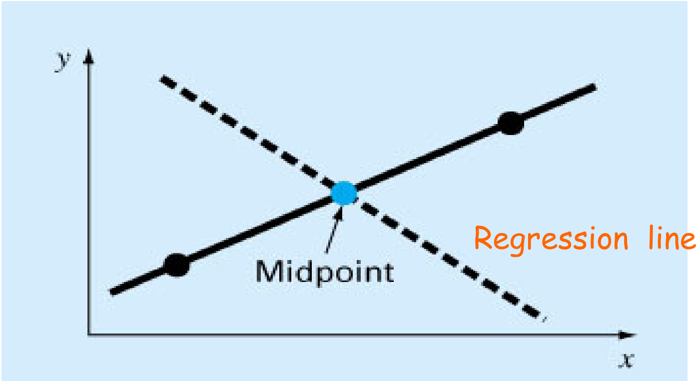
Choosing Criteria For a "Best Fit"

- **Minimize** the sum of the residual errors for all available data?

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (y_i - a_o - a_1 x_i)$$

Inadequate!

(see →→→)

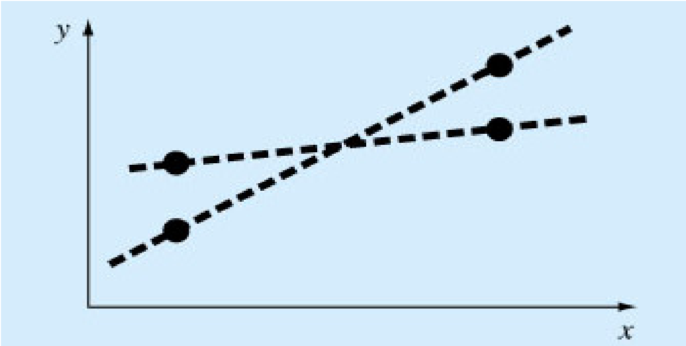


- Sum of the absolute values?

$$\sum_{i=1}^{n} |e_i| = \sum_{i=1}^{n} |y_i - a_0 - a_1 x_i|$$

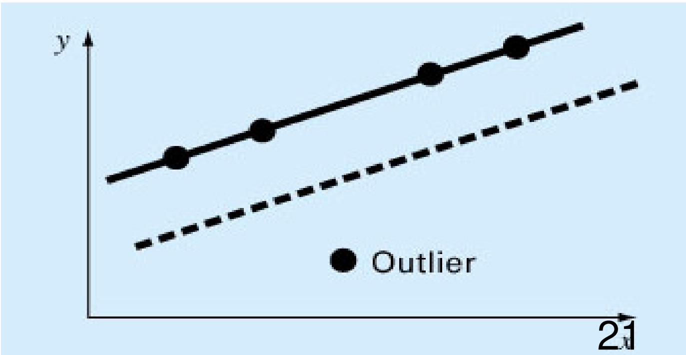
Inadequate!

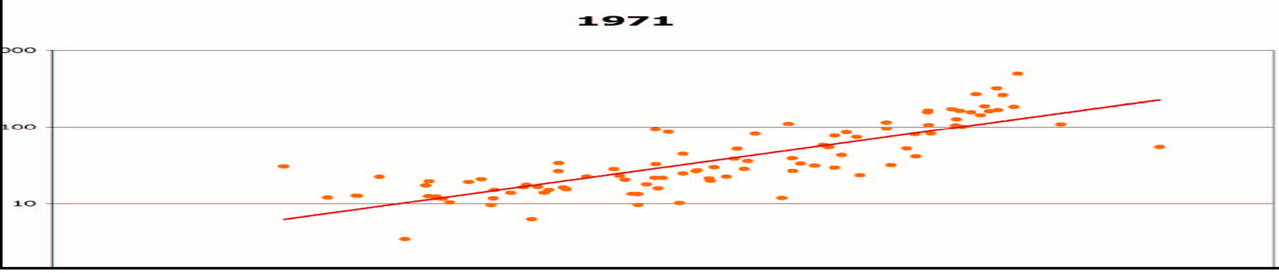
(see →→→)



- How about minimizing the distance that an individual point falls from the line?

This does not work either! see →→→



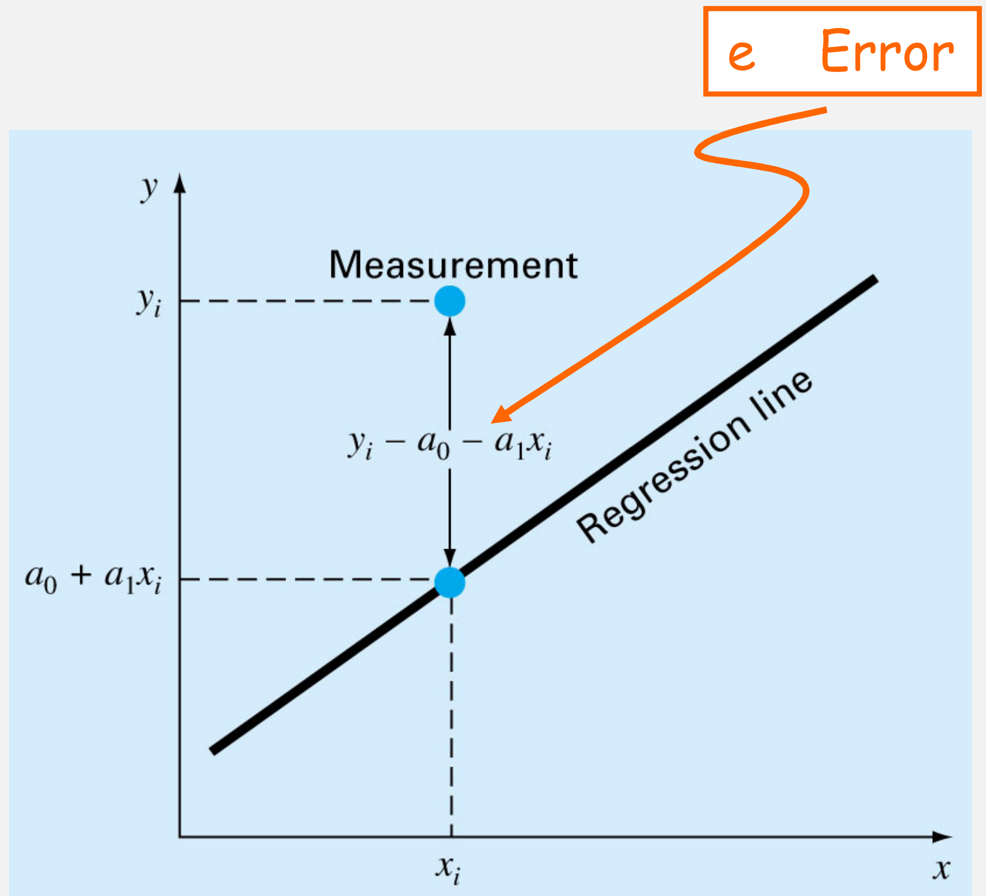


- **Best strategy is to *minimize* the sum of the squares of the residuals between the *measured*-y and the y *calculated* with the linear model:**

$$S_r = \sum_{i=1}^{n} e_i^2$$

$$= \sum_{i=1}^{n} (y_{i,measured} - y_{i,model})^2$$

$$S_r = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$



- **Yields a unique line for a given set of data**

- **Need to compute a_0 and a_1 such that S_r is minimized!**

Least-Squares Fit of a Straight Line

Minimize error : $$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

$$\frac{\partial S_r}{\partial a_o} = -2\sum (y_i - a_o - a_1 x_i) = 0 \qquad \Rightarrow \qquad \sum y_i - \sum a_0 - \sum a_1 x_i = 0$$

$$\frac{\partial S_r}{\partial a_1} = -2\sum \left[(y_i - a_o - a_1 x_i) x_i\right] = 0 \qquad \Rightarrow \qquad \sum y_i x_i - \sum a_0 x_i - \sum a_1 x_i^2 = 0$$

Since $\sum a_0 = n a_0$

$$\left.\begin{aligned} &(1) \qquad n a_0 + \left(\sum x_i\right) a_1 = \sum y_i \\ &(2) \quad \left(\sum x_i\right) a_0 + \left(\sum x_i^2\right) a_1 = \sum y_i x_i \end{aligned}\right\}$$

Normal equations which can be solved simultaneously

Least-Squares Fit of a Straight Line

$$\text{Minimize error}: \quad S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

Normal equations which can be solved simultaneously

$$\begin{aligned} n a_0 \quad + \quad \left(\sum x_i\right) a_1 &= \sum y_i \\ \left(\sum x_i\right) a_0 \quad + \quad \left(\sum x_i^2\right) a_1 &= \sum y_i x_i \end{aligned}$$

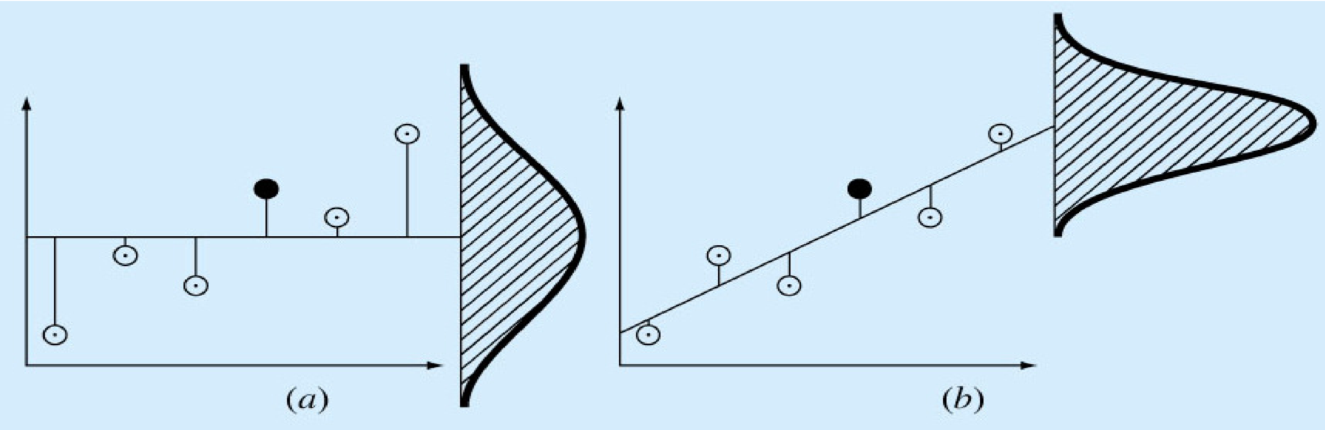
$$\begin{bmatrix} \mathrm{n} & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum y_i x_i \end{bmatrix}$$

$$a_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - \left(\sum x_i\right)^2}$$

using $\bar{y} = a_0 + a_1 \bar{x}$, a_0 can be expressed as $a_0 = \bar{y} - a_1 \bar{x}$

Mean values

"Goodness" of our fit



The spread of data

(a) around the mean

(b) around the best-fit line

Notice the improvement in the error due to *linear regression*

- **S_r = Sum of the squares of residuals around the regression line**
- **S_t = total sum of the squares around the mean**
- **$(S_t - S_r)$ quantifies the improvement or error reduction due to describing data in terms of *a straight line* rather than as *an average value*.**

r : *correlation coefficient*

r^2 : *coefficient of determination*

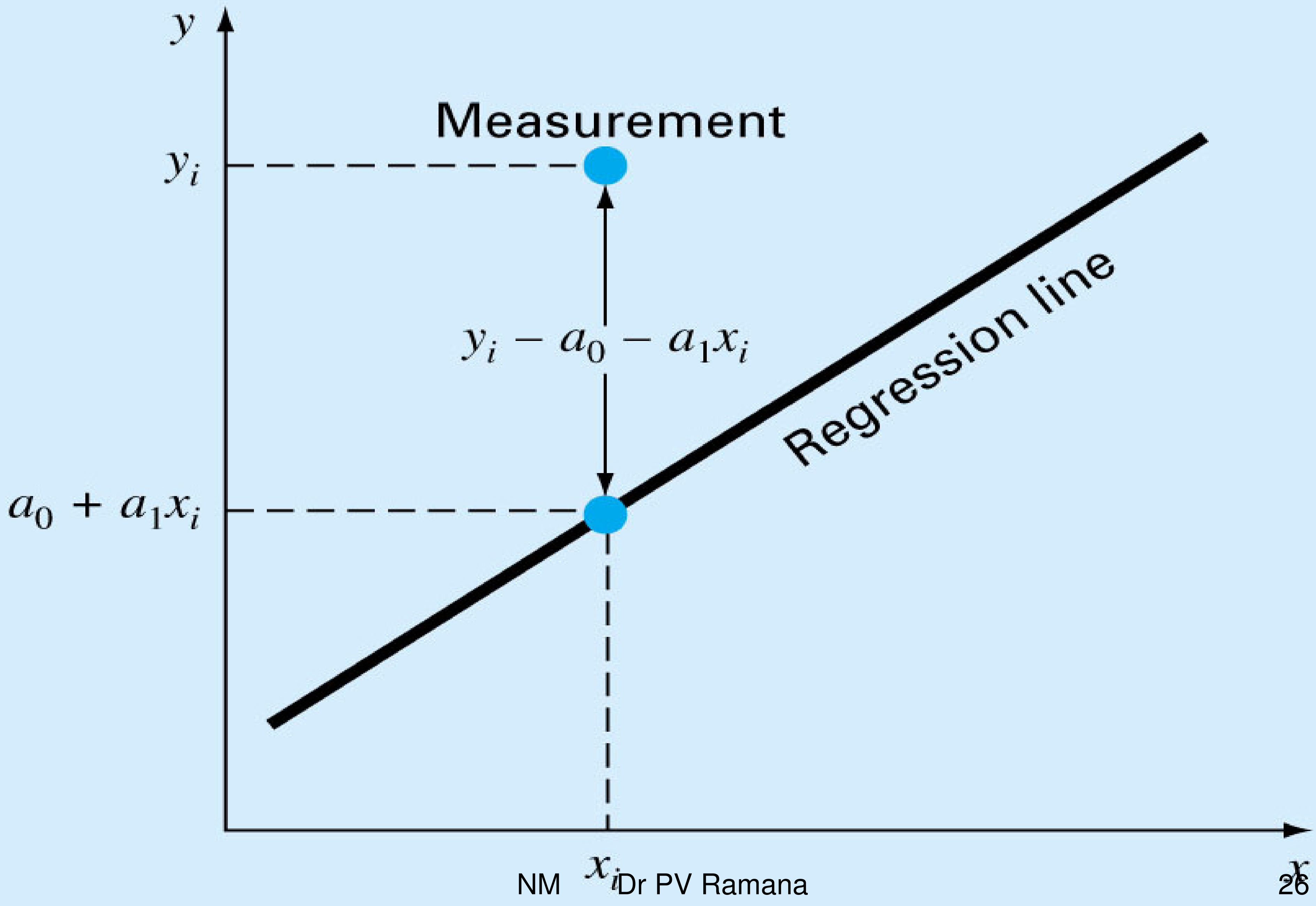
$$r^2 = \frac{S_t - S_r}{S_t}$$

$$S_t = \sum (y_i - \bar{y})^2$$

$$S_r = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

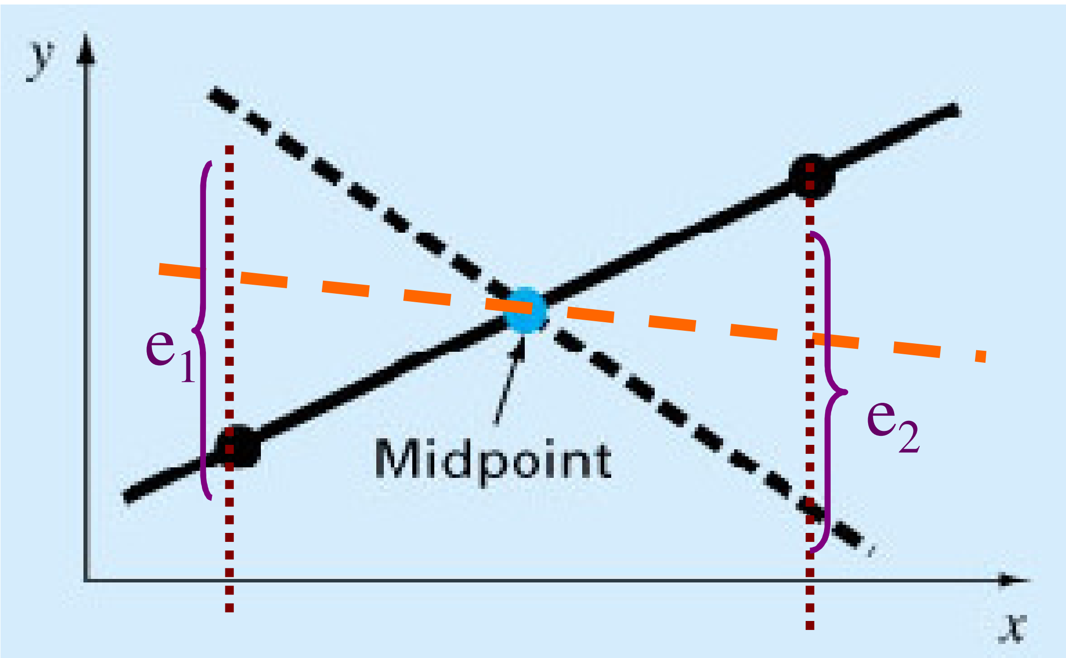
- **For a perfect fit $S_r = 0$ and $r = r^2 = 1$ signifies that the line explains 100 percent of the variability of the data.**
- **For $r = r^2 = 0$ ➔ $S_r = S_t$ ➔ the fit represents no improvement**

Linear Regression: Residual



Linear Regression: Criteria for a "Best" Fit

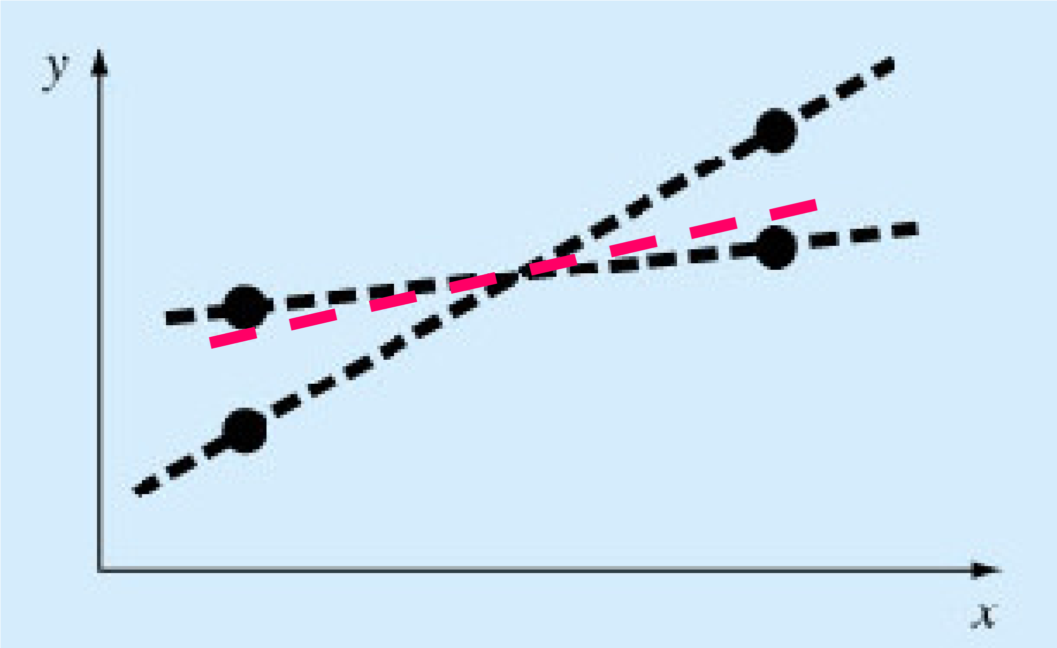
$$\min \sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)$$



$e_1 = -e_2$

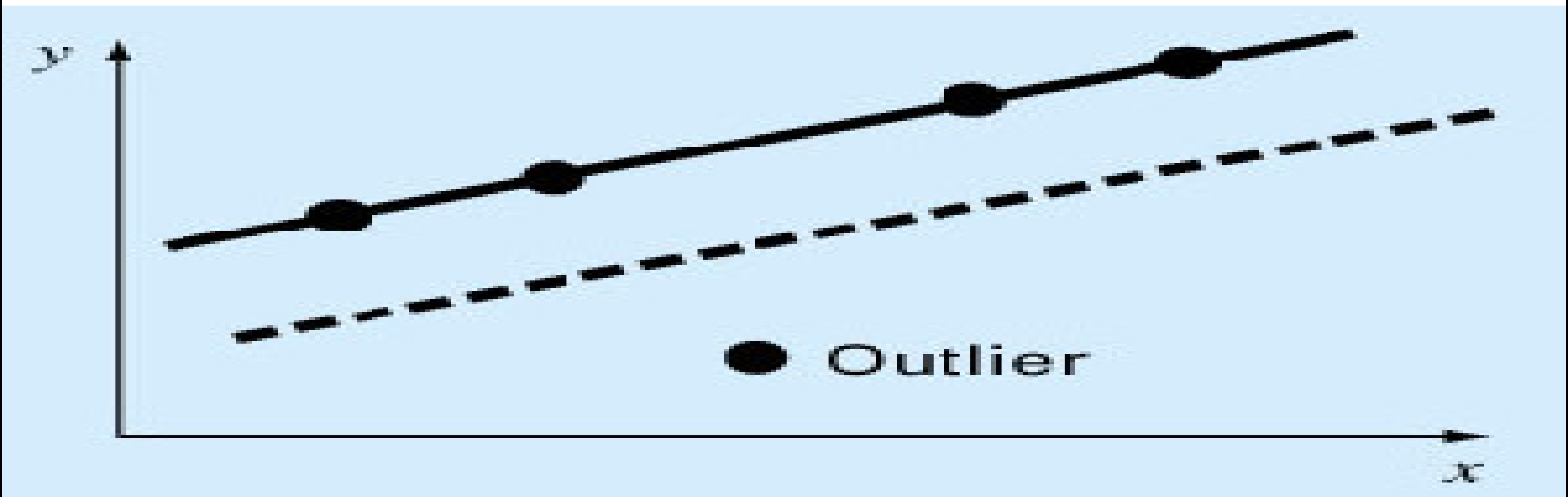
Linear Regression: Criteria for a "Best" Fit

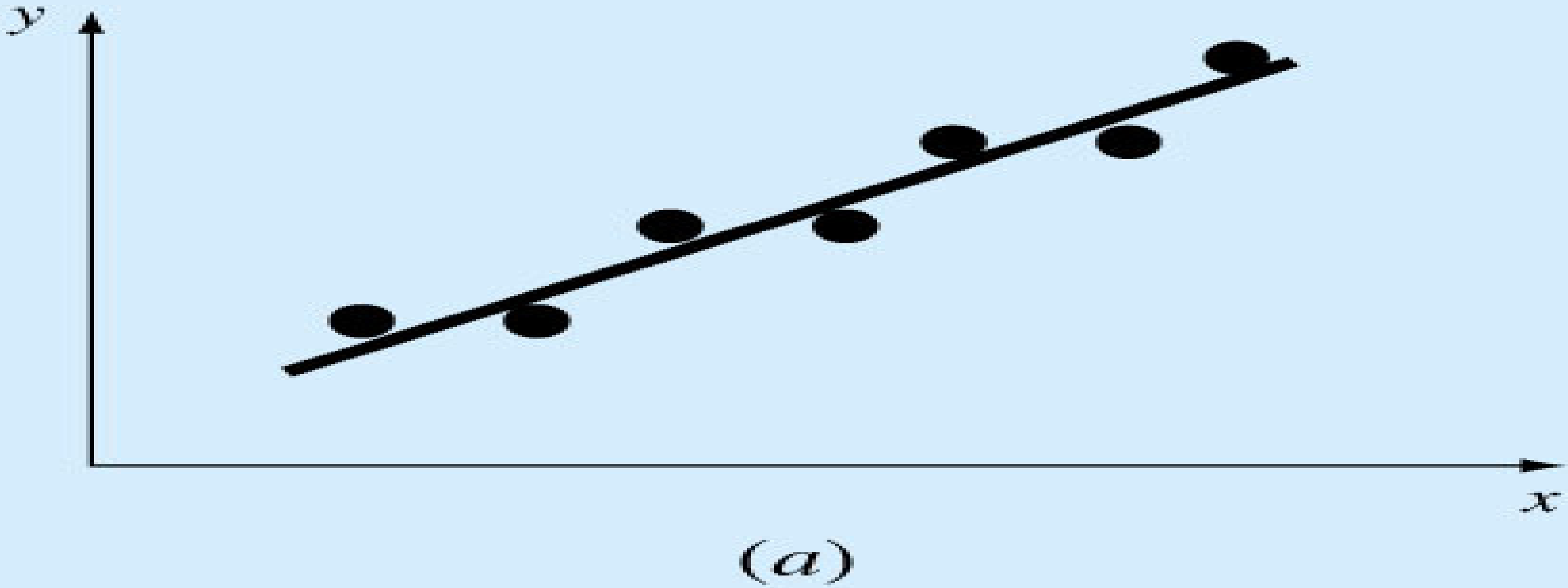
$$\min \sum_{i=1}^{n} | e_i | = \sum_{i=1}^{n} | y_i - a_0 - a_1 x_i |$$

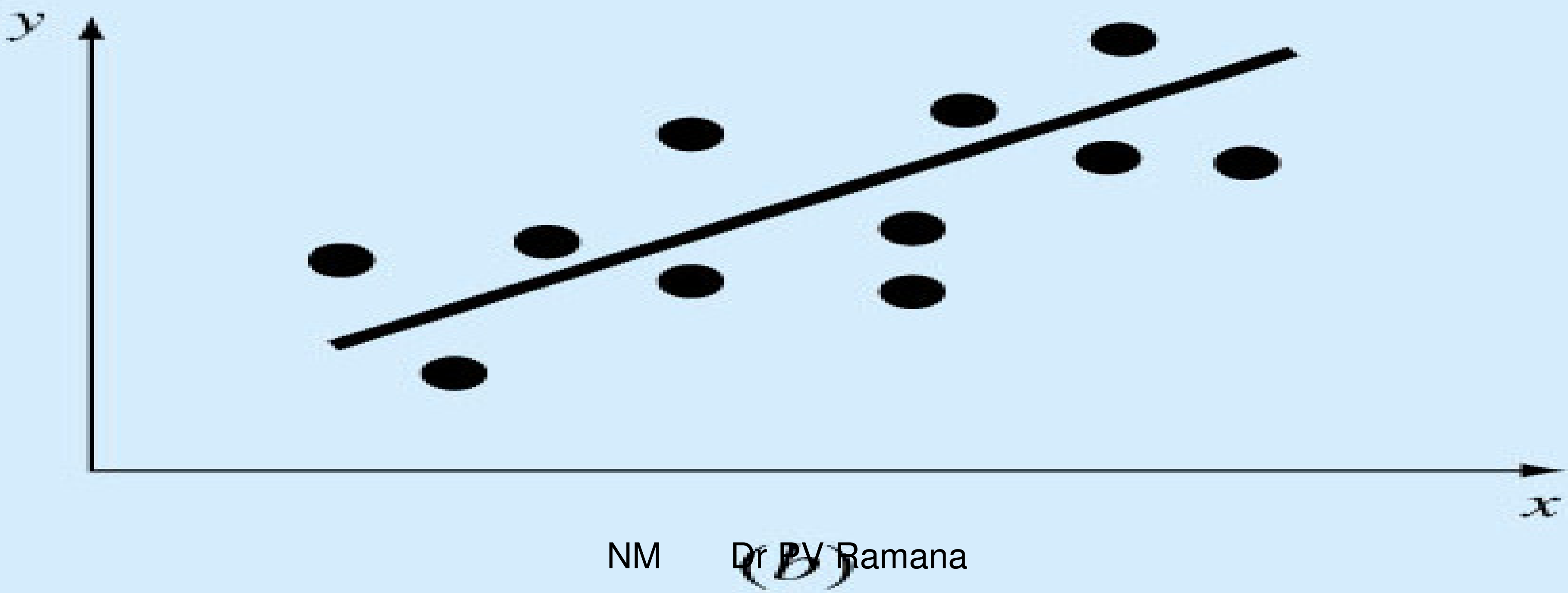


Linear Regression: Criteria for a "Best" Fit

$$\min \max_{i=1}^{n} |e_i| = |y_i - a_0 - a_1 x_i|$$




y
x
(a)


y
x
(b)

Linear Regression: Determination of a_o and a_1

$$\frac{\partial S_r}{\partial a_o} = -2\sum (y_i - a_o - a_1 x_i) = 0$$

$$\frac{\partial S_r}{\partial a_1} = -2\sum [(y_i - a_o - a_1 x_i) x_i] = 0$$

$$0 = \sum y_i - \sum a_0 - \sum a_1 x_i$$

$$0 = \sum y_i x_i - \sum a_0 x_i - \sum a_1 x_i^2$$

$$\sum a_0 = n a_0$$

$$n a_0 + \left(\sum x_i\right) a_1 = \sum y_i$$

$$\sum y_i x_i = \sum a_0 x_i + \sum a_1 x_i^2$$

2 equations with 2 unknowns, can be solved simultaneously

Linear Regression: Determination of a_o and a_1

$$\min \sum_{i=1}^{n} |e_i| = \sum_{i=1}^{n} |y_i - a_0 - a_1 x_i|$$

$$a_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2}$$

$$a_0 = \bar{y} - a_1 \bar{x}$$

Error Quantification of Linear Regression

- Total sum of the squares around the mean for the dependent variable, y, is ***S_t***

$$S_t = \sum (y_i - \bar{y})^2$$

- **Sum of the squares of residuals around the regression line is** S_r

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a_o - a_1 x_i)^2$$

Error Quantification of Linear Regression

- S_t-S_r quantifies the improvement or error reduction due to describing data in terms of a straight line rather than as an average value.

$$r^2 = \frac{S_t - S_r}{S_t}$$

***r*²:** coefficient of determination

r : correlation coefficient

Error Quantification of Linear Regression

$$r^2 = \frac{S_t - S_r}{S_t}$$

For a perfect fit:

- $S_r = 0$ and $r = r^2 = 1$, signifying that the line explains 100 percent of the variability of the data.

- For $r = r^2 = 0$, $S_r = S_t$, the fit represents no improvement.

Curve Fitting

Statistics Background

- Arithmetic mean

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

- Standard deviation

$$s_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n-1}}$$

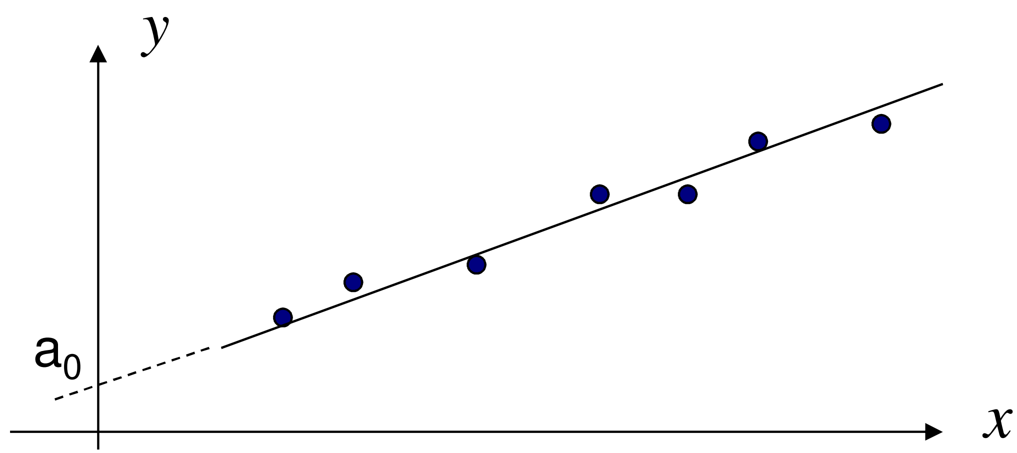
$$s_y = \sqrt{\frac{\sum y_i^2 - n\bar{y}^2}{n-1}}$$

- Variance

$$s_y^2 = \frac{\sum (y_i - \bar{y})^2}{n-1}$$

Curve Fitting

Linear Regression



The straight line equation:

$$y = f(x) = a_0 + a_1 x$$

Given: A set of data $\{ (x_i, y_i) \}$

Find: a_0, and a_1 such that the sum of square of the error, **e_i ,is minimum**

$$e_i = f(x) - y_i$$

Curve Fitting

Linear Regression

$$E = \sum_{i=1}^{n} \left[f(x_i) - y_i \right]^2$$

$$= \sum_{i=1}^{n} \left[(a_0 + a_1 x_i) - y_i \right]^2$$

E is minimum only if:

$$\frac{\partial E}{\partial a_0} = 2\sum_{i=1}^{n} (a_0 + a_1 x_i - y_i) = 0$$

$$\frac{\partial E}{\partial a_1} = 2\sum_{i=1}^{n} (a_0 + a_1 x_i - y_i)(x_i) = 0$$

Curve Fitting

Linear Regression

$$\frac{\partial E}{\partial a_0} = 2\sum_{i=1}^{n}(a_0 + a_1 x_i - y_i) = 0$$

$$\frac{\partial E}{\partial a_1} = 2\sum_{i=1}^{n}(a_0 + a_1 x_i - y_i)(x_i) = 0$$

Solving for a_1 and a_0

$$a_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

$$a_0 = \bar{y} - a_1 \bar{x}$$

Minimize the error

1- By minimizing the sum of errors at all data points. Mathematically, it means:

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)$$

2- By minimizing the sum of the absolute values of errors at all data points:

$$\sum_{i=1}^{n} |e_i| = \sum_{i=1}^{n} |(y_i - a_0 - a_1 x_i)|$$

3- By minimizing the sum of the squares of errors at all data points:

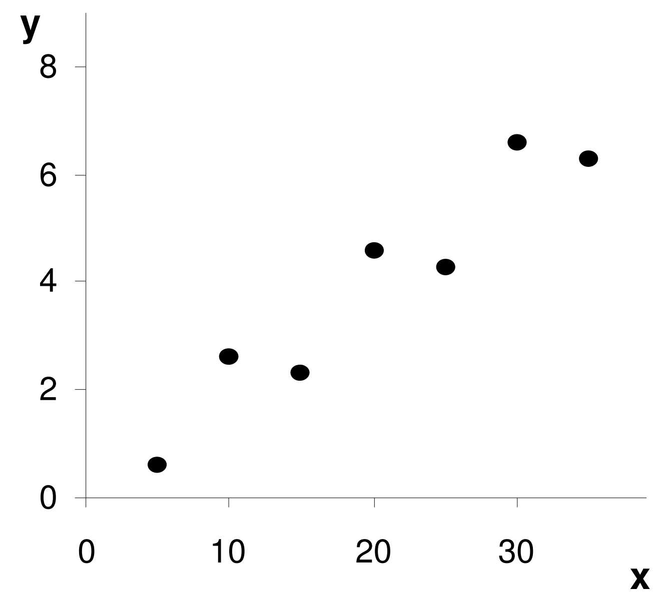
$$\sum_{i=1}^{n} e^2{}_i = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

$$a_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x^2{}_i - (\sum x_i)^2}$$

$$a_0 = \frac{\sum y_i}{n} - a_1 \frac{\sum x_i}{n} = \bar{y} - a_1 \bar{x}$$

Curve Fitting-Linear Regression

Point	x	y
1	5	0.6
2	10	2.6
3	15	2.3
4	20	4.6
5	25	4.3
6	30	6.6
7	35	6.3



$$y \approx a_0 + a_1 x$$

or

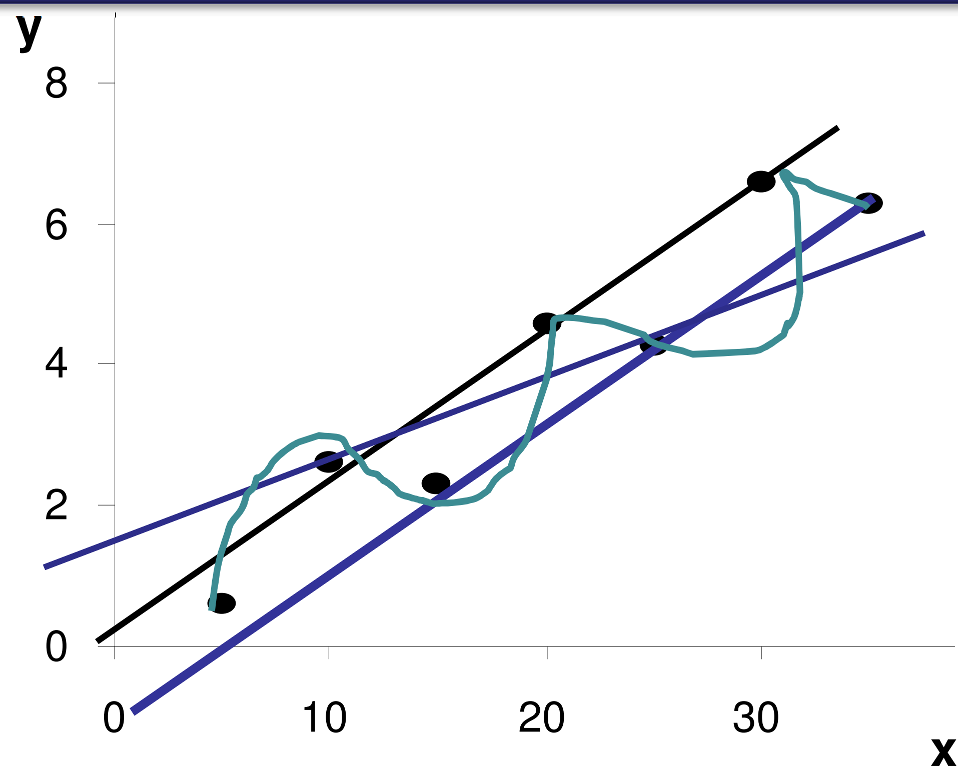
$$y = a_0 + a_1 x + e$$

$$a_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x^2{}_i - \left(\sum x_i\right)^2}$$

$$a_0 = \frac{\sum y_i}{n} - a_1 \frac{\sum x_i}{n} = \bar{y} - a_1 \bar{x}$$

Curve Fitting-Linear Regression

Point	x	y
1	5	0.6
2	10	2.6
3	15	2.3
4	20	4.6
5	25	4.3
6	30	6.6
7	35	6.3



$$a_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x^2{}_i - \left(\sum x_i\right)^2}$$

$$a_0 = \frac{\sum y_i}{n} - a_1 \frac{\sum x_i}{n} = \bar{y} - a_1 \bar{x}$$

Use least squares (linear regression) to fit a straight line to the data given above and compare the predicted values to the experimental values of y

Example 1

Point	x	y
1	5	0.6
2	10	2.6
3	15	2.3
4	20	4.6
5	25	4.3
6	30	6.6
7	35	6.3

x_i	y_i	x_i^2	$x_i y_i$
5	0.6	25	3
10	2.6	100	26
15	2.3	225	34.5
20	4.6	400	92
25	4.3	625	107.5
30	6.6	900	198
35	6.3	1225	220.5
140	27.3	3500	681.5

x_i	y_i	y(estimate) $= a_0+a_1x_i$
5	0.6	0.99
10	2.6	1.96
15	2.3	2.93
20	4.6	3.9
25	4.3	4.87
30	6.6	5.84
35	6.3	6.80

$$a_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

$$a_1 = \frac{(7)(681.5)-(140)(27.3)}{(7)(3500)-(140)^2}$$

= 0.1957

$$a_0 = \bar{y} - a_1 \bar{x}$$

$$a_0 = \frac{27.3}{7} - a_1 \frac{140}{7} = 0.02857$$

$$y_{estimate} = a_0 + a_1 x_i$$
$$= 0.02857 + 0.1957 x_i$$

Linear Regression

$$a_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

$$a_0 = \bar{y} - a_1 \bar{x}$$

Example 2

Fit a straight line to the x and y values as shown:

x_i	1	2	3	4	5	6	7
y_i	**0.5**	**2.5**	**2.0**	**4.0**	**3.5**	**6.0**	**5.5**

	x_i	y_i	x_i^2	$x_i y_i$	$(y_i - \bar{y})^2$
	1	0.5	1	0.5	8.5765
	2	2.5	4	5.0	0.8622
	3	2.0	9	6.0	2.0408
	4	4.0	16	16.0	0.3265
	5	3.5	25	17.5	0.0051
	6	6.0	36	36.0	6.6122
	7	5.5	49	38.5	4.2908
Σ	28.0	24.0	140	119.5	22.7143

$$\bar{x} = \frac{28}{7} = 4$$

$$\bar{y} = \frac{24}{7} = 3.428571$$

$$a_1 = \frac{7(119.5) - 28(24)}{7(140) - (28)^2}$$

= 0.8392857

$$a_0 = 3.428571 - 0.8392857(4)$$

= 0.07142857

Least Squares Fit of a Straight Line: Ex

$$S_t = \sum (y_i - \bar{y})^2$$

$$S_r = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

Fit a straight line to the x and y values in the following Table:

x_i	y_i	$x_i y_i$	x_i^2
1	0.5	0.5	1
2	2.5	5	4
3	2	6	9
4	4	16	16
5	3.5	17.5	25
6	6	36	36
7	5.5	38.5	49
28	**24**	**119.5**	**140**

$$\sum x_i = 28 \qquad \sum y_i = 24.0$$

$$\sum x_i^2 = 140 \qquad \sum x_i y_i = 119.5$$

$$\bar{x} = \frac{28}{7} = 4$$

$$\bar{y} = \frac{24}{7} = 3.428571$$

Least Squares Fit of a Straight Line: Ex

$$a_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

$$= \frac{7 \times 119.5 - 28 \times 24}{7 \times 140 - 28^2} = 0.8392857$$

$$a_0 = \bar{y} - a_1 \bar{x}$$

$$= 3.428571 - 0.8392857 \times 4 = 0.07142857$$

Y = 0.07142857 + 0.8392857 x

Least Squares Fit of a Straight Line: Ex

$$a_0 = \bar{y} - a_1\bar{x} = 3.428571 - 0.8392857 \times 4 = 0.07142857$$

$$S_t = \sum (y_i - \bar{y})^2$$

$$S_r = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

x_i	y_i	$(y_i - \bar{y})^2$	$e_i^2 = (y_i - \hat{y})^2$
1	0.5	8.5765	0.1687
2	2.5	0.8622	0.5625
3	2.0	2.0408	0.3473
4	4.0	0.3265	0.3265
5	3.5	0.0051	0.5896
6	6.0	6.6122	0.7972
7	5.5	4.2908	0.1993
28	**24.0**	**22.7143**	**2.9911**

$$S_t = \sum (y_i - \bar{y})^2 = 22.7143$$

$$S_r = \sum e_i^{\;2} = 2.9911$$

$$r^2 = \frac{S_t - S_r}{S_t} = 0.868$$

$$r = \sqrt{r^2} = \sqrt{0.868} = 0.932$$

$$\bar{y} = \frac{24}{7} = 3.428571$$

Y = 0.07142857 + 0.8392857 x

Least Squares Fit of a Straight Line: Ex

$$S_t = \sum (y_i - \bar{y})^2 = 22.7143$$

$$S_r = \sum e_i^{\ 2} = 2.9911$$

$$r = \sqrt{r^2} = \sqrt{0.868} = 0.932$$

- **The standard deviation (quantifies the spread around the mean):**

$$s_y = \sqrt{\frac{S_t}{n-1}} = \sqrt{\frac{22.7143}{7-1}} = 1.9457$$

- The standard error of estimate (quantifies the spread around the regression line)

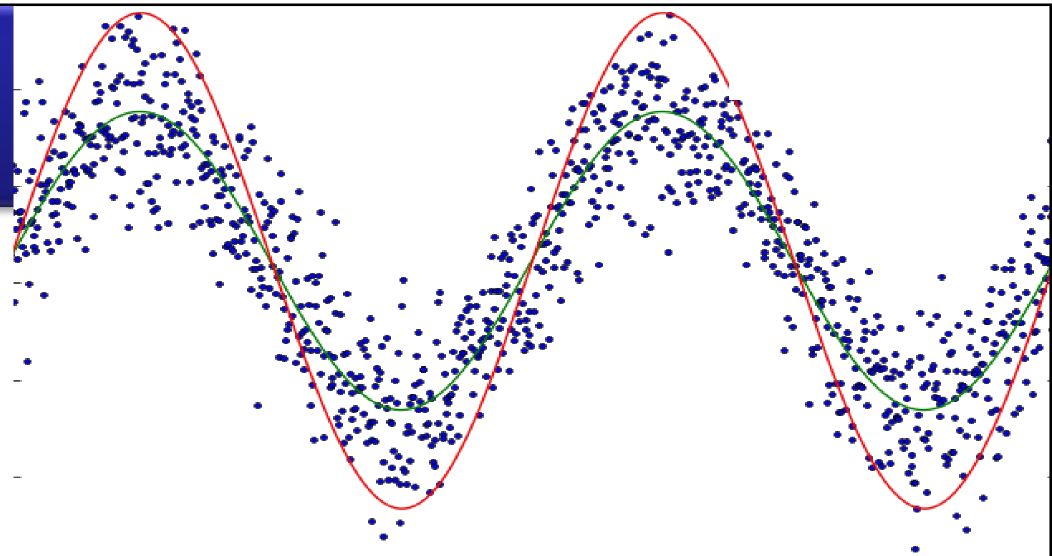
$$s_{y/x} = \sqrt{\frac{S_r}{n-2}} = \sqrt{\frac{2.9911}{7-2}} = 0.7735$$

Because $S_{y/x} < S_y$, the linear regression model has good fitness

Algorithm for linear regression

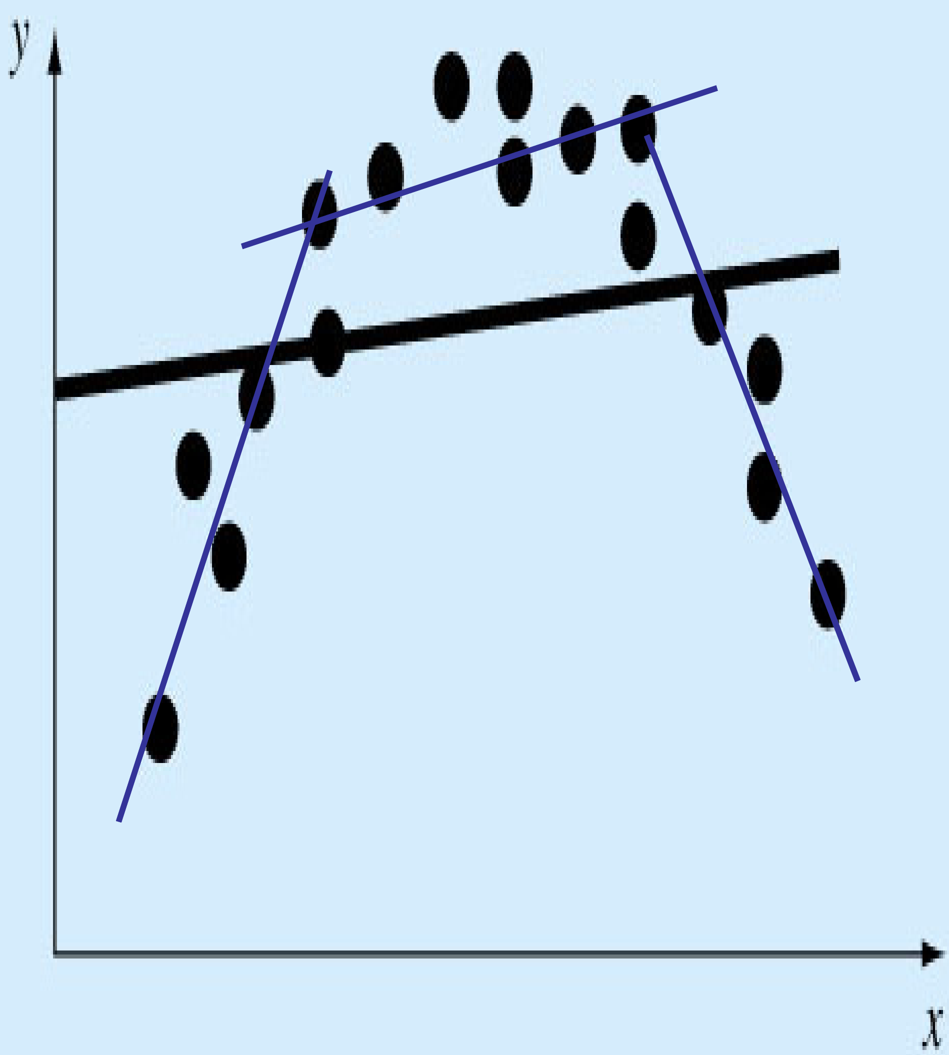
```
SUB Regress(x, y, n, a1, a0, syx, r2)

  sumx = 0: sumxy = 0: st = 0
  sumy = 0: sumx2 = 0: sr = 0
  DO i = 1, n
    sumx = sumx + x_i
    sumy = sumy + y_i
    sumxy = sumxy + x_i*y_i
    sumx2 = sumx2 + x_i*x_i
  END DO
  xm = sumx/n
  ym = sumy/n
  a1 = (n*sumxy − sumx*sumy)/(n*sumx2 − sumx*sumx)
  a0 = ym − a1*xm
  DO i = 1, n
    st = st + (y_i − ym)^2
    sr = sr + (y_i − a1*x_i − a0)^2
  END DO
  syx = (sr/(n − 2))^0.5
  r2 = (st − sr)/st

END Regress
```

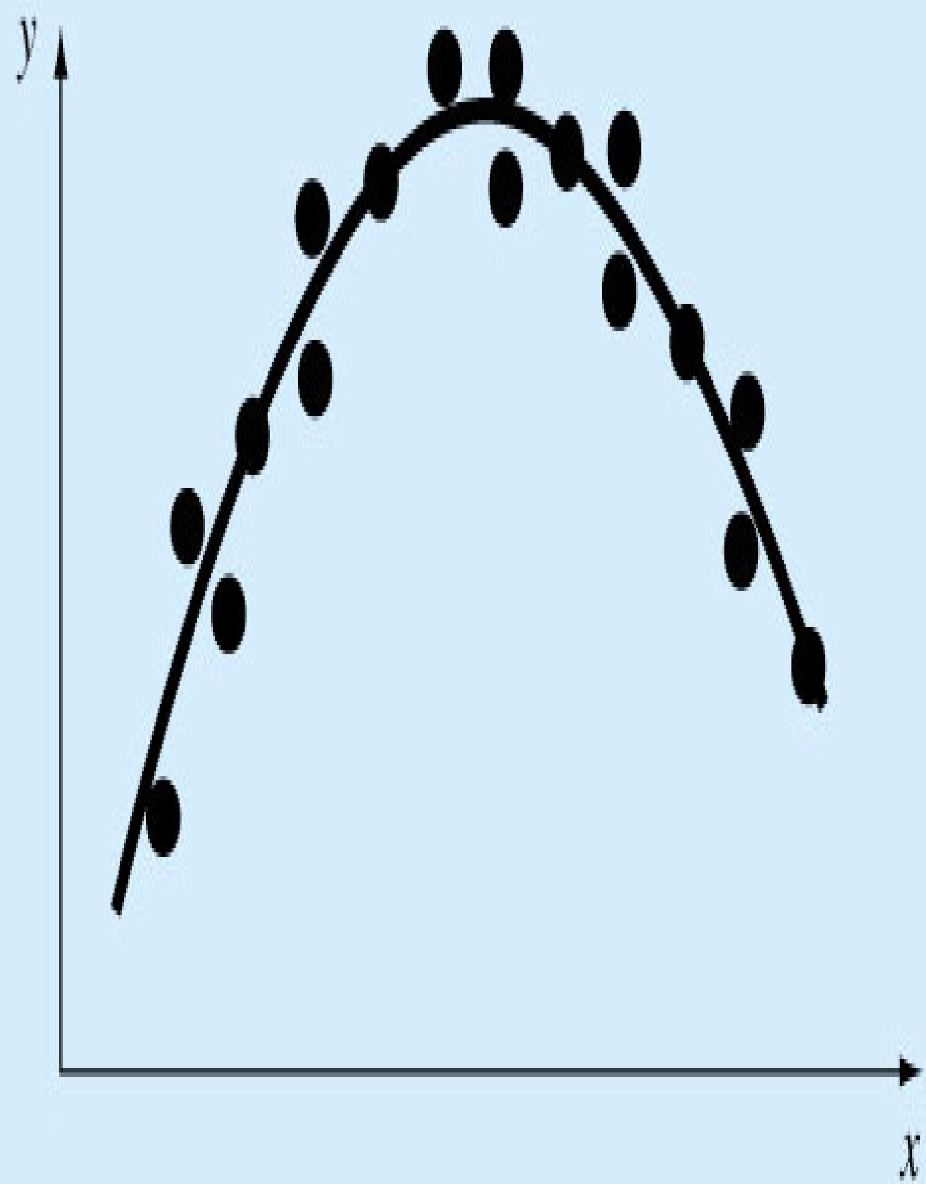
Polynomial Regression



- Some engineering data is poorly represented by a straight line.
- For these cases a curve is better suited to fit the data.
- The least squares method can readily be extended to fit the data to higher order polynomials.

Polynomial Regression





A parabola is preferable

Second Order Polynomial Regression

- Some engineering data is poorly represented by a straight line. A curve (polynomial) may be better suited to fit the data. The least squares method can be extended to fit the data to higher order polynomials.
- **As an example let us consider a second order polynomial to fit the data points:**

$$y = a_0 + a_1 x + a_2 x^2$$

$$\text{Minimize error:} \quad S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i - a_2 x_i^2)^2$$

$$\frac{\partial S_r}{\partial a_o} = -2\sum (y_i - a_o - a_1 x_i - a_2 x_i^2) = 0$$

$$\frac{\partial S_r}{\partial a_1} = -2\sum x_i (y_i - a_o - a_1 x_i - a_2 x_i^2) = 0$$

$$\frac{\partial S_r}{\partial a_2} = -2\sum x_i^2 (y_i - a_o - a_1 x_i - a_2 x_i^2) = 0$$

$$n a_0 + \left(\sum x_i\right) a_1 + \left(\sum x_i^2\right) a_2 = \sum y_i$$

$$\left(\sum x_i\right) a_0 + \left(\sum x_i^2\right) a_1 + \left(\sum x_i^3\right) a_2 = \sum x_i y_i$$

$$\left(\sum x_i^2\right) a_0 + \left(\sum x_i^3\right) a_1 + \left(\sum x_i^4\right) a_2 = \sum x_i^2 y_i$$

Polynomial Regression

- **A 2nd order polynomial (quadratic)** is defined by:

$$y = a_o + a_1 x + a_2 x^2 + e$$

- The residuals between the model and the data:

$$e_i = y_i - a_o - a_1 x_i - a_2 x_i^2$$

- The sum of squares of the residual:

$$S_r = \sum e_i^2 = \sum \left(y_i - a_o - a_1 x_i - a_2 x_i^2 \right)^2$$

Polynomial Regression

$$\frac{\partial S_r}{\partial a_o} = -2\sum (y_i - a_o - a_1 x_i - a_2 x_i^2) = 0$$

$$\frac{\partial S_r}{\partial a_1} = -2\sum (y_i - a_o - a_1 x_i - a_2 x_i^2) x_i = 0$$

$$\frac{\partial S_r}{\partial a_2} = -2\sum (y_i - a_o - a_1 x_i - a_2 x_i^2) x_i^2 = 0$$

$$\left.\begin{aligned}
\sum y_i &= n \cdot a_o + a_1 \sum x_i + a_2 \sum x_i^2 \\
\sum x_i y_i &= a_o \sum x_i + a_1 \sum x_i^2 + a_2 \sum x_i^3 \\
\sum x_i^2 y_i &= a_o \sum x_i^2 + a_1 \sum x_i^3 + a_2 \sum x_i^4
\end{aligned}\right\}$$

3 linear equations with 3 unknowns (a_o,a_1,a_2), can be solved

Polynomial Regression

- A system of 3x3 equations needs to be solved to determine the coefficients of the polynomial.

$$\begin{bmatrix} n & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \end{Bmatrix}$$

- The standard error & the coefficient of determination

$$s_{y/x} = \sqrt{\frac{S_r}{n-3}}$$

$$r^2 = \frac{S_t - S_r}{S_t}$$

Polynomial Regression - Ex

$$\begin{bmatrix} n & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \end{Bmatrix}$$

Example 1

Fit a second order polynomial to six data points:

x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i y_i$	$x_i^2 y_i$
0	2.1	0	0	0	0	0
1	7.7	1	1	1	7.7	7.7
2	13.6	4	8	16	27.2	54.4
3	27.2	9	27	81	81.6	244.8
4	40.9	16	64	256	163.6	654.4
5	61.1	25	125	625	305.5	1527.5
15	152.6	55	225	979	585.6	2489

$$\sum x_i = 15$$

$$\sum y_i = 152.6$$

$$\sum x_i^2 = 55$$

$$\sum x_i^3 = 225$$

$$\sum x_i^4 = 979$$

$$\sum x_i y_i = 585.6$$

$$\sum x_i^2 y_i = 2488.8$$

$$\bar{x} = \frac{15}{6} = 2.5, \quad \bar{y} = \frac{152.6}{6} = 25.433$$

Polynomial Regression - Ex

$$\begin{bmatrix} n & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \end{Bmatrix}$$

- The system of simultaneous linear equations:

$$\begin{bmatrix} 6 & 15 & 55 \\ 15 & 55 & 225 \\ 55 & 225 & 979 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} 152.6 \\ 585.6 \\ 2488.8 \end{Bmatrix}$$

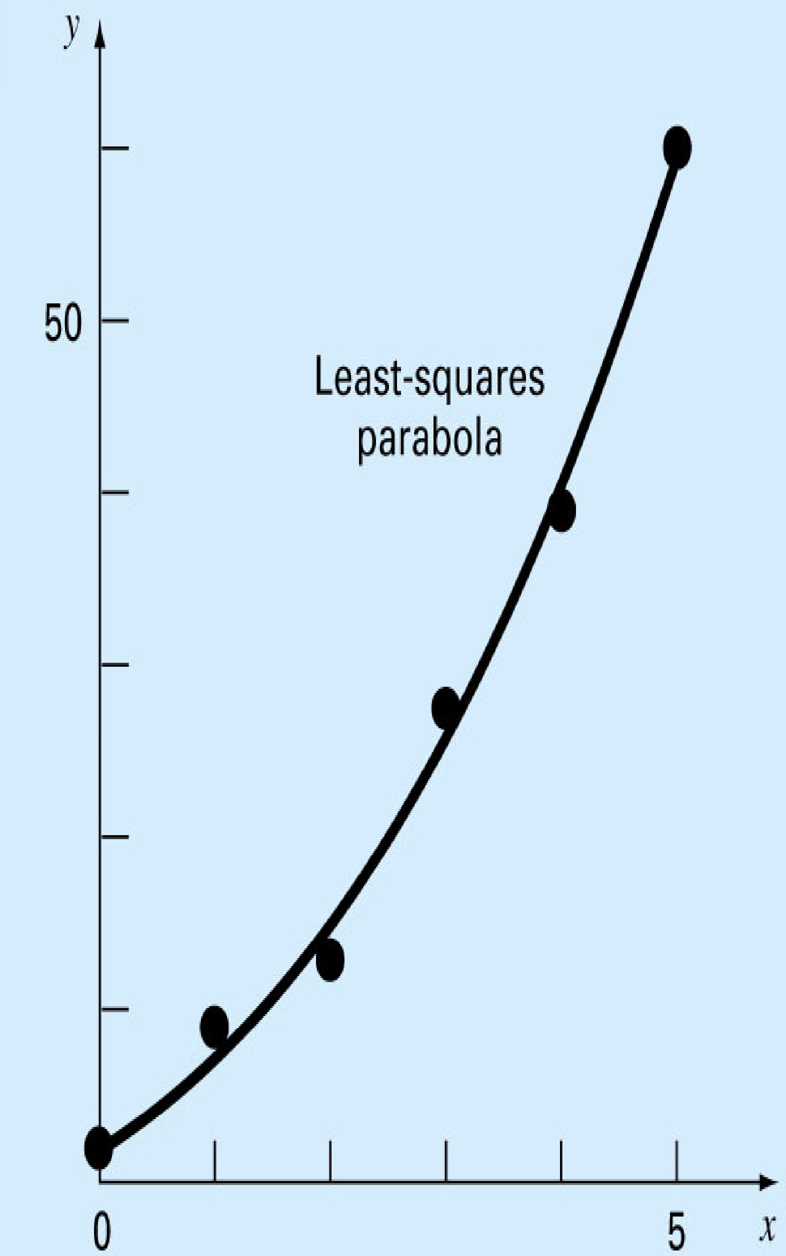
$$a_0 = 2.47857,\ a_1 = 2.35929,\ a_2 = 1.86071$$

$$y = 2.47857 + 2.35929\,x + 1.86071\,x^2$$

$$S_t = \sum (y_i - \bar{y})^2 = 2513.39 \qquad S_r = \sum e_i^{\ 2} = 3.74657$$

Polynomial Regression - Ex

x_i	y_i	y_{model}	e_i^2	$(y_i\text{-}\bar{y})^2$
0	2.1	2.4786	0.14332	544.42889
1	7.7	6.6986	1.00286	314.45929
2	13.6	14.64	1.08158	140.01989
3	27.2	26.303	0.80491	3.12229
4	40.9	41.687	0.61951	239.22809
5	61.1	60.793	0.09439	1272.13489
15	**152.6**		**3.74657**	**2513.39333**



- The standard error of estimate:

$$s_{y/x} = \sqrt{\frac{3.74657}{6-3}} = 1.12$$

- **The coefficient of determination:**

$$r^2 = \frac{2513.39 - 3.74657}{2513.39} = 0.99851, \quad r = \sqrt{r^2} = 0.99925$$

1st Order Polynomial Regression

$y = a_0 + a_1 x$

2 X 2

$$\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum y_i x_i \end{bmatrix}$$

$$a_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2}$$

using $\bar{y} = a_0 + a_1\bar{x}$, a_0 can be expressed as $a_0 = \bar{y} - a_1\bar{x}$

2nd Order Polynomial Regression

3 X 3

$y = a_0 + a_1 x + a_2 x^2$

$$\begin{bmatrix} n & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \end{Bmatrix}$$

3^{rd} Order Polynomial Regression

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3$$

4 X 4

Minimize error : $$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i - a_2 x_i^2 - a_3 x_i^3)^2$$

$$\begin{bmatrix} n & \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 & \sum x_i^5 \\ \sum x_i^3 & \sum x_i^4 & \sum x_i^5 & \sum x_i^6 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{Bmatrix} = \begin{Bmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \\ \sum x_i^3 y_i \end{Bmatrix}$$

m^{th} order Polynomial Regression

- To fit the data to an m^{th} order polynomial, need to solve the following system of linear equations ((m+1) equations with (m+1) unknowns)

$$\begin{bmatrix} n & \sum x_i & \cdots & \sum x_i^m \\ \sum x_i & \sum x_i^2 & \cdots & \sum x_i^{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum x_i^m & \sum x_i^{m+1} & \cdots & \sum x_i^{m+m} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \\ \vdots \\ \sum x_i^m y_i \end{bmatrix}$$

Matrix Form

Polynomial Regression

General:

The mth-order polynomial:

$$y = a_o + a_1 x + a_2 x^2 + \ldots\ldots + a_m x^m + e$$

- A system of (m+1)x(m+1) linear equations must be solved for determining the coefficients of the mth-order polynomial.
- The standard error:

$$s_{y/x} = \sqrt{\frac{S_r}{n-(m+1)}}$$

- The coefficient of determination:

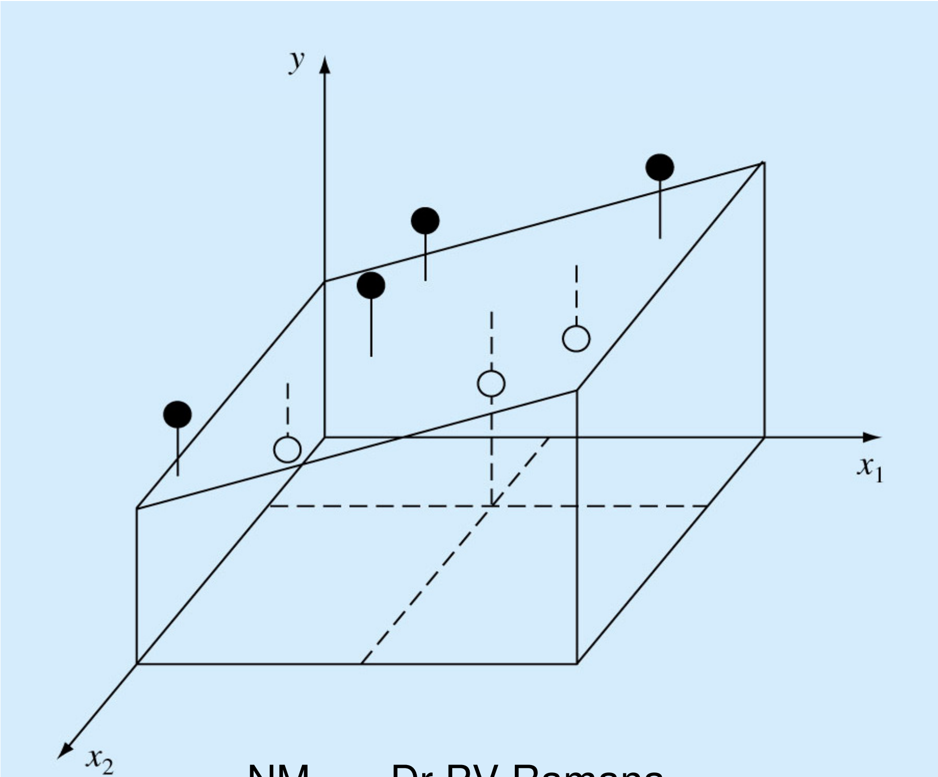
$$r^2 = \frac{S_t - S_r}{S_t}$$

Multiple Linear Regression

- A useful extension of linear regression is the case where *y* is a linear function of two or more independent variables. For example:

$$y = a_o + a_1x_1 + a_2x_2 + e$$

- For this 2-dimensional case, the regression line becomes a plane as shown in the figure below.



Multiple Linear Regression

$$\begin{bmatrix} n & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \end{Bmatrix}$$

Example (2 - vars) : Minimize error : $S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_{1i} - a_2 x_{2i})^2$

$$\frac{\partial S_r}{\partial a_o} = -2\sum (y_i - a_o - a_1 x_{1i} - a_2 x_{2i}) = 0$$

$$\frac{\partial S_r}{\partial a_1} = -2\sum x_{1i}(y_i - a_o - a_1 x_{1i} - a_2 x_{2i}) = 0$$

$$\frac{\partial S_r}{\partial a_2} = -2\sum x_{2i}(y_i - a_o - a_1 x_{1i} - a_2 x_{2i}) = 0$$

$$n a_0 + \left(\sum x_{1i}\right) a_1 + \left(\sum x_{2i}\right) a_2 = \sum y_i$$

$$\left(\sum x_{1i}\right) a_0 + \left(\sum x_{1i}^2\right) a_1 + \left(\sum x_{1i} x_{2i}\right) a_2 = \sum x_{1i} y_i$$

$$\left(\sum x_{2i}\right) a_0 + \left(\sum x_{1i} x_{2i}\right) a_1 + \left(\sum x_{2i}^2\right) a_2 = \sum x_{2i} y_i$$

$$\begin{bmatrix} n & \sum x_{1i} & \sum x_{2i} \\ \sum x_{1i} & \sum x_{1i}^2 & \sum x_{1i} x_{2i} \\ \sum x_{2i} & \sum x_{1i} x_{2i} & \sum x_{2i}^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_{1i} y_i \\ \sum x_{2i} y_i \end{bmatrix}$$

Which method would one can use to solve this Linear System of Equations?

Multiple Linear Regression

$$\begin{bmatrix} n & \sum x_{1i} & \sum x_{2i} \\ \sum x_{1i} & \sum x_{1i}^2 & \sum x_{1i}x_{2i} \\ \sum x_{2i} & \sum x_{1i}x_{2i} & \sum x_{2i}^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_{1i} y_i \\ \sum x_{2i} y_i \end{bmatrix}$$

Example 1

The following data is calculated from the equation

x_1	x_2	y	x^2_1	x^2_2	x_1x_2
0	0	5	0	0	0
2	1	10	4	1	2
2.5	2	9	6.25	4	5
1	3	0	1	9	3
4	6	3	16	36	24
7	2	27	49	4	14
16.5	**14**	**54**	**76.25**	**54**	**48**

Use multiple linear regression to fit this data.

Solution:

$y = 5 + 4x_1 - 3x_2$

$$\begin{bmatrix} 6 & 16.5 & 14 \\ 16.5 & 76.25 & 48 \\ 14 & 48 & 54 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 54 \\ 243.5 \\ 100 \end{bmatrix}$$

this system can be solved using Gauss Elimination

The result is: $a_0=5$ $a_1=4$ and $a_2= -3$

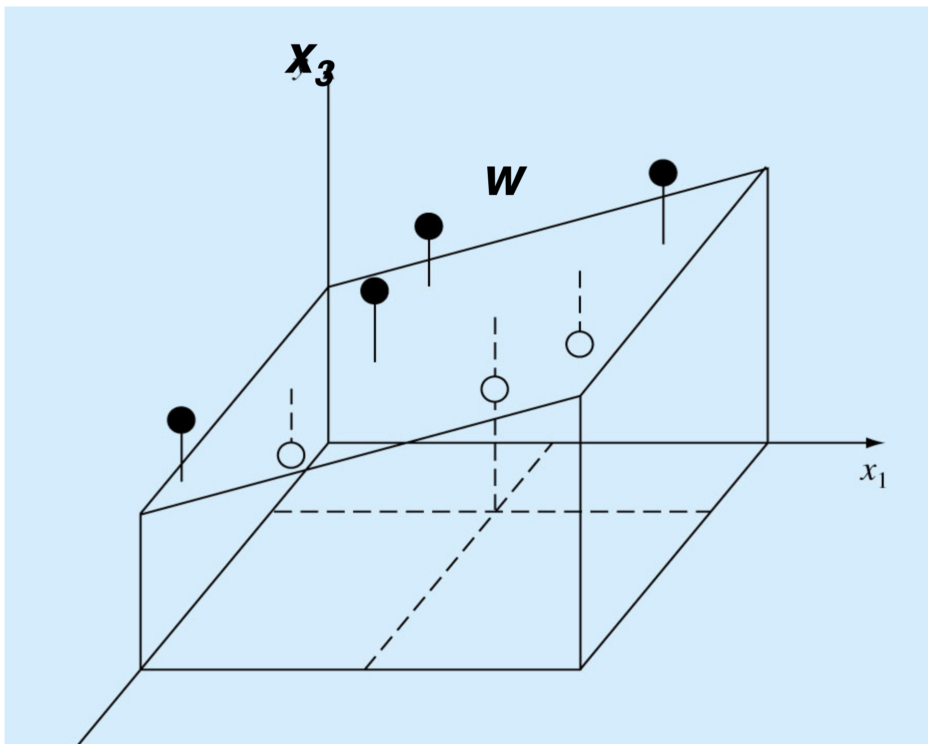
$y = 5 + 4x_1 -3x_2$

3D Multiple Linear Regression

- A useful extension of linear regression is the case where w is a linear function of three or more independent variables. For example:

$$w = a_o + a_1x_1 + a_2x_2 + a_3x_3 + e$$

- For this three - dimensional case, the regression line becomes a plane as shown in the figure below.



Multiple Linear Regression

$$\begin{bmatrix} n & \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 & \sum x_i^5 \\ \sum x_i^3 & \sum x_i^4 & \sum x_i^5 & \sum x_i^6 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{Bmatrix} = \begin{Bmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \\ \sum x_i^3 y_i \end{Bmatrix}$$

Example (3 - variables): Minimize error: $S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_{1i} - a_2 x_{2i} - a_3 x_{3i})^2$

$$\frac{\partial S_r}{\partial a_o} = -2\sum (y_i - a_o - a_1 x_{1i} - a_2 x_{2i} - a_3 x_{3i}) = 0$$

$$n a_0 + \left(\sum x_{1i}\right) a_1 + \left(\sum x_{2i}\right) a_2 + \left(\sum x_{3i}\right) a_3 = \sum y_i$$

$$\frac{\partial S_r}{\partial a_1} = -2\sum x_{1i} (y_i - a_o - a_1 x_{1i} - a_2 x_{2i} - a_3 x_{3i}) = 0$$

$$\left(\sum x_{1i}\right) a_0 + \left(\sum x_{1i}^2\right) a_1 + \left(\sum x_{1i} x_{2i}\right) a_2 + \left(\sum x_{1i} x_{3i}\right) a_3 = \sum x_{1i} y_i$$

$$\frac{\partial S_r}{\partial a_2} = -2\sum x_{2i} (y_i - a_o - a_1 x_{1i} - a_2 x_{2i} - a_3 x_{3i}) = 0$$

$$\left(\sum x_{2i}\right) a_0 + \left(\sum x_{1i} x_{2i}\right) a_1 + \left(\sum x_{2i}^2\right) a_2 + \left(\sum x_{2i} x_{3i}\right) a_3 = \sum x_{2i} y_i$$

$$\frac{\partial S_r}{\partial a_3} = -2\sum x_{3i} (y_i - a_o - a_1 x_{1i} - a_2 x_{2i} - a_3 x_{3i}) = 0$$

$$\left(\sum x_{3i}\right) a_0 + \left(\sum x_{1i} x_{3i}\right) a_1 + \left(\sum x_{2i} x_{3i}\right) a_2 + \left(\sum x_{3i}^2\right) a_3 = \sum x_{3i} y_i$$

$$\begin{bmatrix} n & \sum x_{1i} & \sum x_{2i} & \sum x_{3i} \\ \sum x_{1i} & \sum x_{1i}^2 & \sum x_{1i} x_{2i} & \sum x_{1i} x_{3i} \\ \sum x_{2i} & \sum x_{1i} x_{2i} & \sum x_{2i}^2 & \sum x_{2i} x_{3i} \\ \sum x_{3i} & \sum x_{1i} x_{3i} & \sum x_{2i} x_{3i} & \sum x_{3i}^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_{1i} y_i \\ \sum x_{2i} y_i \\ \sum x_{3i} y_i \end{bmatrix}$$

Interpolation

- General formula for an *n-th* order polynomial
  - $y = a_0 + a_1x + a_2x^2 + \ldots a_mx^m$
- For *m+1* data points, there is one, and only one polynomial of order *m* or less that passes through all points

Example: $y = a_0 + a_1x$
fits between 2 points
1^{st} order

Example: $y = a_0 + a_1x + a_2x^2$
fits between 3 points
2^{nd} order

Interpolation

- One can explore two mathematical methods well suited for computer implementation

- **Lagrange Interpolating Polynomial**

- **Newton's Divided Difference Interpolating Polynomials**