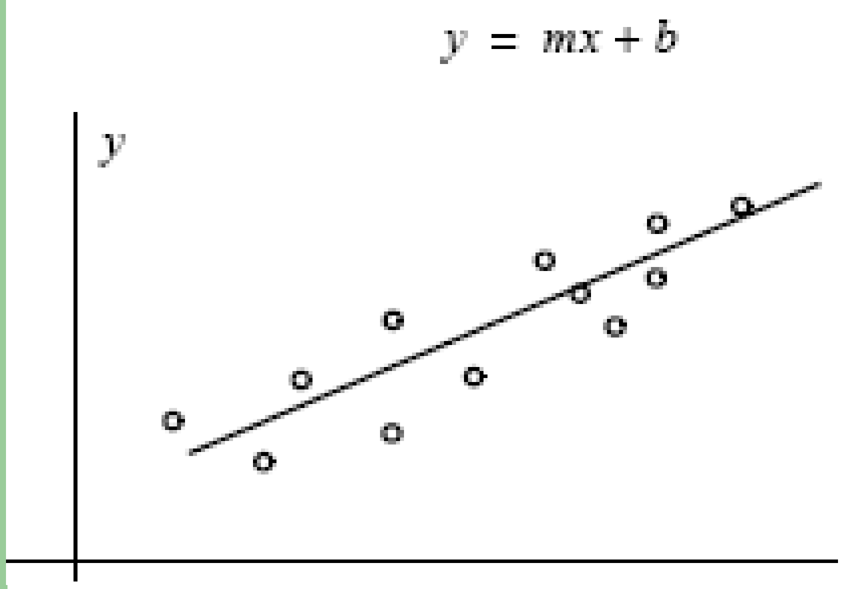
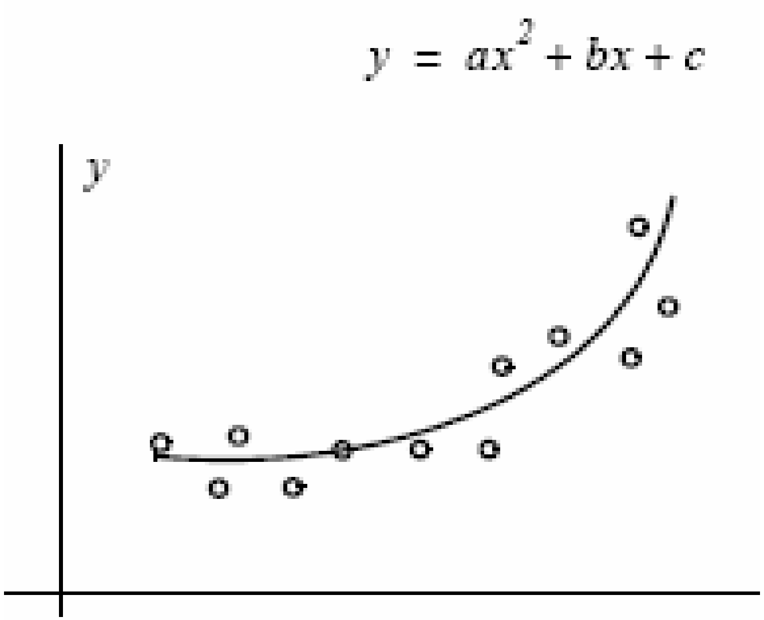
Regression

Ching-Han Chen
I-Shou University
2006-04-25

Curve Fitting

Curve fitting is the process of finding equations to approximate straight lines and curves that best fit given sets of data.





Least-squares Regression Curve

Let the distance of data point *x1* from the line be denoted as *d1,* the distance of data point *x2*, from the same line as *d2*, , and so on. The best fitting straight line or curve has the property that

$$d_1^2 + d_2^2 + \ldots + d_3^2 = minimum$$

and it is referred to as the l*east-squares curve*.

Linear Regression

In order to best fit the observed data, we compute the coefficients *m*(slope) and b(y-intercept) of the straight line equation

$$y = mx + b$$

such that the sum of the squares of the errors will be minimum. We denote the straight line equations passing through these points as

$$y_1 = mx_1 + b$$

$$y_2 = mx_2 + b$$

$$y_3 = mx_3 + b$$

$$\dots$$

$$y_n = mx_n + b$$

Minimize the squares errors

Sum of the squares errors:

$$\sum squares = [y_1 - (mx_1 + b)]^2 + [y_2 - (mx_2 + b)]^2 + \dots + [y_n - (mx_n + b)]^2$$

Minimize the squares errors:

$$\frac{\partial}{\partial m}\sum squares = -2x_1[y_1 - (mx_1 + b)] - 2x_2[y_2 - (mx_2 + b)] - \dots - 2x_n[y_n - (mx_n + b)] = 0$$

$$\frac{\partial}{\partial b}\sum squares = -2[y_1 - (mx_1 + b)] - 2[y_2 - (mx_2 + b)] - \dots - 2[y_n - (mx_n + b)] = 0$$

Regression Equation

$$(\Sigma x^2)m + (\Sigma x)b = \Sigma xy$$
$$(\Sigma x)m + nb = \Sigma y$$

Σx = *sum of the numbers x*

Σy = *sum of the numbers y*

Σxy = *sum of the numbers of the product xy*

Σx^2 = *sum of the numbers x squared*

n = *number of data x*

Solve the Regression Equation

$$y = mx + b$$

$$(\Sigma x^2)m + (\Sigma x)b = \Sigma xy$$
$$(\Sigma x)m + nb = \Sigma y$$

$$b = \frac{nS_{XY} - S_X S_Y}{nS_{XX} - S_X S_X}.$$

$$m = \frac{S_Y - bS_X}{n}.$$

Algorithm

```
void regression(float x[], float y[], int nb,
   float &m, float &b)
{
   float xy=0,sx=0,sy=0,x2=0;
   int i;
   for(i=0;i<nb;i++)
   {
       xy=xy+x[i]*y[i];
       sx=sx+x[i];
       sy=sy+y[i];
       x2=x2+x[i]*x[i];
   }
   b=(nb*xy-sx*sy)/(nb*x2-sx*sx);
   m=(sy - b*sx)/nb;
}
```

Ex1.

While temperature increments, the observed resistance values are shown in

$T\ (^{\circ}C)$	x	0	10	20	30	40	50	60	70	80	90	100
$R\ (\Omega)$	y	27.6	31.0	34.0	37	40	42.6	45.5	48.3	51.1	54	56.7

Compute the straight line equation that best fits the observed data.

Ans: $y = mx + b = 0.288x + 28.123$

Using Cramer's Rule

$$(\Sigma x^2)m + (\Sigma x)b = \Sigma xy$$
$$(\Sigma x)m + nb = \Sigma y$$

With Cramer's rule, *m* and *b* are computed from

$$m = \frac{D_1}{\Delta} \qquad b = \frac{D_2}{\Delta}$$

$$\Delta = \begin{vmatrix} \Sigma x^2 & \Sigma x \\ \Sigma x & n \end{vmatrix} \qquad D_1 = \begin{vmatrix} \Sigma xy & \Sigma x \\ \Sigma y & n \end{vmatrix} \qquad D_2 = \begin{vmatrix} \Sigma x^2 & \Sigma xy \\ \Sigma x & \Sigma y \end{vmatrix}$$

Polynomial Regression

$$y = ax^2 + b + c$$

Find the *least-squares polynomial with* coefficients *a*, *b* and *c*

$$(\Sigma x^2)a + (\Sigma x)b + nc = \Sigma y$$

$$(\Sigma x^3)a + (\Sigma x^2)b + (\Sigma x)c = \Sigma xy$$

$$(\Sigma x^4)a + (\Sigma x^3)b + (\Sigma x^2)c = \Sigma x^2 y$$

Algorithm

$$(\Sigma x^2)a + (\Sigma x)b + nc = \Sigma y$$

$$(\Sigma x^3)a + (\Sigma x^2)b + (\Sigma x)c = \Sigma xy$$

$$(\Sigma x^4)a + (\Sigma x^3)b + (\Sigma x^2)c = \Sigma x^2 y$$

```
for(i=0;i<nb;i++)
{
        u[0][0] += x[i]*x[i];
        u[0][1] += x[i];
        u[1][0] += pow(x[i],3);
        u[2][0] += pow(x[i],4);
        v[0] += y[i];
        v[1] += x[i]*y[i];
        v[2] += x[i]*x[i]*y[i];
}
u[0][2]=nb;
u[1][1]=u[0][0];
u[1][2]=u[0][1];
u[2][1]=u[1][0];
u[2][2]=u[1][1];
```

Ex.2

Find the *least-squares polynomial with* observed data :

x	1.2	1.5	1.8	2.6	3.1	4.3	4.9	5.3
y	4.5	5.1	5.8	6.7	7.0	7.3	7.6	7.4
x	5.7	6.4	7.1	7.6	8.6	9.2	9.8	
y	7.2	6.9	6.6	5.1	4.5	3.4	2.7	

$$530.15a + 79.1b + 15c = 87.8$$

$$4004.50a + 530.15b + 79.1c = 437.72$$

$$32331.49a + 4004.50b + 530.15c = 2698.37$$

Solution for Ex.2

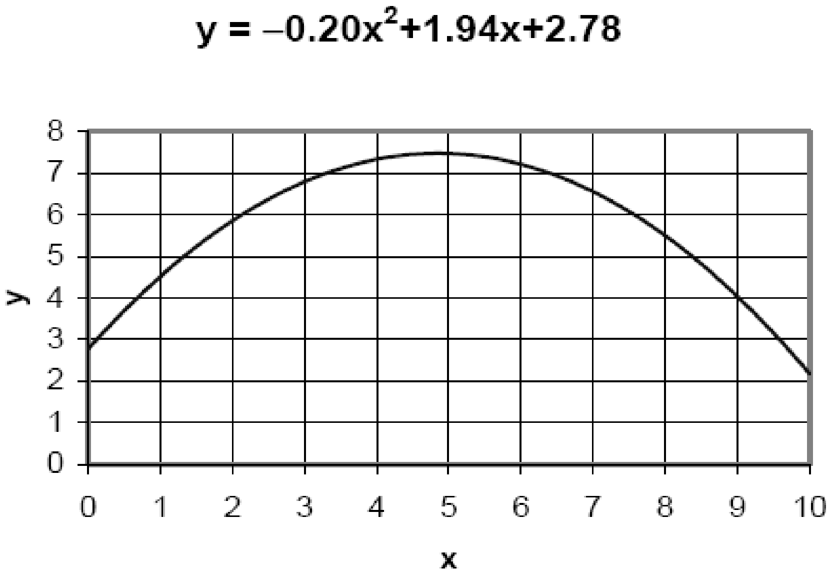
$$530.15a + 79.1b + 15c = 87.8$$

$$4004.50a + 530.15b + 79.1c = 437.72$$

$$32331.49a + 4004.50b + 530.15c = 2698.37$$

The least-squares polynomial is obtained :

$$y = -0.20x^2 + 1.94x + 2.78$$



Ex.3

In a non-linear resistance device, experiments yielded the observed data:

millivolts	*100*	*120*	*140*	*160*	*180*	*200*
milliamps	*0.45*	*0.55*	*0.60*	*0.70*	*0.80*	*0.85*

Use respectively the linear and polynomial regression to compute the straight line and polynomial equation that best fits the given data.