October 22, 2010

Note Title 10/22/2010

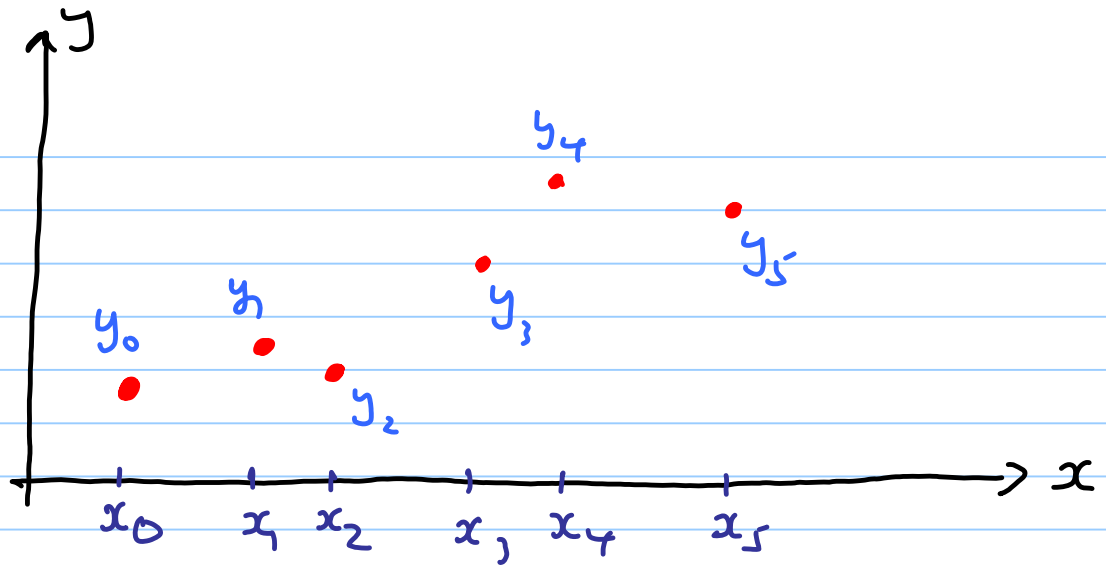
12 Smoothing of Data and the Method of Least Squares.

Suppose the following numerical table

x	x_0	x_1	x_2	....	x_m
y	y_0	y_1	y_2	...	y_m

represent data obtained from some experiment.

The $m+1$ points can be plotted on a graph. Assume the data points have the pattern given below.



It seems reasonable to conclude that the underlying function is linear. The failure of the points to fall precisely on a straight line is due to experimental error.

Proceeding with this assumption, we need to determine the correct function of the form

$$y = ax + b$$

Need to determine the coefficients a and b.

Geometrically: What line most nearly passes through the points plotted?

Suppose you comedy guess the values of a and b.

- If the k^{th} datum falls on the line, it should satisfy the equation

$$ax_k + b - y_k = 0$$

- If the k^{th} datum does not fall on the line,

$$|\text{error}| = |ax_k + b - y_k|$$

the total absolute error for all the $m+1$ points is

therefore $$\sum_{k=0}^{m} |ax_k + b - y_k|$$

which is a function of a and b.

So choose a and b so that that

$$\Phi(a,b) = \sum_{k=0}^{m} |ax_k + b - y_k|$$

assumes its minimum value.

This is an example of l_1 approximation.

Remarks

- The function $Q(a,b)$ is generally not differentiable.
- Calculus methods will not work!
- Solution methods include linear programming.

The practice is to minimize the error function

$$\Phi(a,b) = \sum_{k=0}^{m} (ax_k + b - y_k)^2$$

[for statistical reasons].