13 Least squares approximation

13.1 Closest point in a plane.

Problem. Given P, Q, Y, Calculate the point closest to Y in the plane containing O, P, Q. The plane is

$$\{\alpha P + \beta Q : \ \alpha, \beta \in \mathbb{R}\}$$

Let the answer be Z. If this is in 3 dimensions then we can calculate $P \times Q = R$, project Y onto the line OR getting a point W, and then set $Z = Y - W$.

For example, $P = (1, -2, 1)$ and $Q = (3, 2, -5)$. $P \times Q = (8, 8, 8)$, $W = ((1, 2, 3) \cdot (8, 8, 8)/192)(8, 8, 8) = (48/192)(8, 8, 8) = (2, 2, 2)$, and $Z = (-1, 0, 1)$.

Now, Z is the point in the plane $\alpha P + \beta Q$ which is closest to Y, which also means that the displacement vector $\vec{YZ}$ is perpendicular to $\vec{OP}$ and $\vec{OQ}$. The components of $\vec{YZ}$ are the coordinates of $Z - Y$ and so on.

$$\vec{OP} \cdot \vec{YZ} = 0$$
$$\vec{OQ} \cdot \vec{YZ} = 0$$

Viewing P, Q, Y, Z as column vectors, and replacing Z by $\alpha P + \beta Q$,

$$P^T(\alpha P + \beta Q - Y) = 0$$
$$Q^T(\alpha P + \beta Q - Y) = 0$$

If A is the matrix $[P, Q]$ (columns are P and Q), this becomes

$$A^T A \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = A^T Y$$

That is,

$$\begin{bmatrix} 0 & -6 \\ -8 & 38 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0 \\ -8 \end{bmatrix}$$

This system is easily solved: $\alpha = -1/4, \beta = -1/4$.

The closest point to Y is

$$Z = \frac{1}{4}(-(1, -2, 1) - (3, 2, -5)) = (-1, 0, 1))$$

the same as before.

13.1.1 Summarising how to calculate points closest to Y

That is, to calculate the values of α and β so that $\alpha P + \beta Q$ is closest to a given point Y, let $A_{2\times 2} = [P, Q]$ and solve the following in terms of column vectors:

$$(A^T A) \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = A^T Y$$

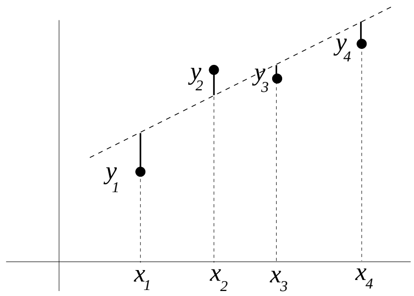
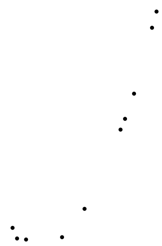


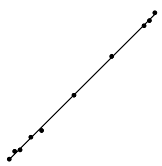
Figure 1: a hand-drawn estimate of the line $y = mx + c$.

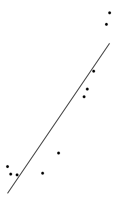
13.2 Linear and quadratic estimation

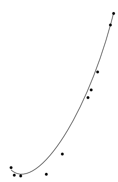
We are given a set of measurements which ought to be in a straight line or maybe on a quadratic curve, but are subject to experimental error. Two examples are shown.



Here are 'least squared error' lines fitting the data







13.3 Least squared error estimation; fitting a straight line

Often a quantity y depends linearly on x, $y = ax + b$, but where samples are taken there are measurement errors. That is, one is given a list of points $(x_1, y_1), \ldots, (x_n, y_n)$ which are supposed to be collinear, but aren't. Nevertheless, one knows that m and c exist. How can we produce a good estimate of m and c from the data?

In laboratory practice, people may simply draw a line that looks close, such as illustrated in Figure 1.

In the figure, the sample points are black dots and the 'errors', the distances between the (x_i, y_i) and $(x_i, ax_i + b)$, are indicated by short heavy lines.

We are given a set of equations for m and c. They are almost certainly inconsistent:

$$y_1 = mx_1 + c$$
$$y_2 = mx_2 + c$$
$$y_3 = mx_3 + c \ldots$$

There is a generally-accepted rule for choosing m and c: *Choose m and c so that the sum of the squares of the errors is minimal.* That is, find m and c so that

$$\sum (y_i - mx_i - c)^2$$

is minimal.

We have seen this a few times already. Let

$$P = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} Q = \begin{bmatrix} 1 \\ 1 \\ \cdots \\ 1 \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix}, \quad A = [P\ Q]$$

Then we want to find the point

$$Pm + Qc$$

closest to Y.

As previously, this becomes

$$A^T A \begin{bmatrix} m \\ c \end{bmatrix} = A^T Y.$$

Also,

$$A^T A = \begin{bmatrix} \sum_i x_i^2 & \sum_i x_i \\ \sum_i x_i & n \end{bmatrix}, \quad A^T Y = \begin{bmatrix} \sum_i x_i y_i \\ \sum_i y_i \end{bmatrix}$$

Example. Find the least-squares approximating line for the data $(0,1),\ (2,2),\ (4,0),\ (6,3)$.

$$\begin{bmatrix} m \\ c \end{bmatrix} = (A^T A)^{-1} A^T Y,$$

where

$$A = \begin{bmatrix} 0 & 2 & 4 & 6 \\ 1 & 1 & 1 & 1 \end{bmatrix}^T, \quad \text{and} \quad Y = [1\ 2\ 0\ 3]^T.$$

$$\begin{bmatrix} m \\ c \end{bmatrix} = \begin{bmatrix} 56 & 12 \\ 12 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 22 \\ 6 \end{bmatrix}.$$

Solution: $m = 1/5, c = 9/10$.

13.4 Fitting a quadratic curve to data points

Sometimes a quadratic curve could fit the data much better, sometimes cubic, and so on. Strange to relate, exactly the same formula holds in each case, provided we interpret it correctly.

Quadratic least-squared-error curves are about the limit of what can be hand-calculated. This involves finding a quadratic function

$$ax^2 + bx + c$$

rather than a linear function to fit the data points. X and Y are as before, but this time the matrix A is

$$\begin{bmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ \cdots & & \\ x_n^2 & x_n & 1 \end{bmatrix}.$$

However the formula is the same.

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = (A^T A)^{-1} A^T Y.$$

Note

$$A^T A = \begin{bmatrix} \sum_i x_i^4 & \sum_i x_i^3 & \sum_i x_i^2 \\ \sum_i x_i^3 & \sum_i x_i^2 & \sum_i x_i \\ \sum_i x_i^2 & \sum_i x_i & n \end{bmatrix}, \quad A^T Y = \begin{bmatrix} \sum_i x_i^2 y_i \\ \sum_i x_i y_i \\ \sum_i y_i \end{bmatrix}$$

Example. Find the linear and the quadratic least-squared-error curves to fit the data

i	1	2	3	4	5
x_i	-2	-1	0	1	2
y_i	10	3	3	1	2

$$A^T = \begin{bmatrix} 4 & 1 & 0 & 1 & 4 \\ -2 & -1 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$A^T A = \begin{bmatrix} 34 & 0 & 10 \\ 0 & 10 & 0 \\ 10 & 0 & 5 \end{bmatrix}$$

$$A^T Y = \begin{bmatrix} 52 \\ -18 \\ 19 \end{bmatrix}.$$

Applying Gauss-Jordan elimination, we get

```
1     0     0     1
0     1     0  -9/5
0     0     1   9/5  in rref
```

so the least-squared-error quadratic curve is $y = x^2 - 1.8x + 1.8$.

All the data for the linear least-squares approximation is there; rows and columns $2, 3$ of $A^T A$ are as required for linear approximation, and we drop the top row of $A^T Y$. In other words,

$$\begin{bmatrix} 10 & 0 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} m \\ c \end{bmatrix} = \begin{bmatrix} -18 \\ 19 \end{bmatrix},$$

and the least-squared-error line is $y = -1.8x + 3.8$.

13.5 Another example.

Calculate the least-squared-error quadratic approximation $y = ax^2 + bx + c$ to the data points

$$(-2, 6), \quad (-1, -21), \quad (0, 10), \quad (2, 1)$$

Answer.

$$A^T A = \begin{bmatrix} \sum x_i^4 & \sum x_i^3 & \sum x_i^2 \\ \sum x_i^3 & \sum x_i^2 & \sum x_i \\ \sum x_i^2 & \sum x_i & n \end{bmatrix} =$$

$$\begin{bmatrix} 33 & -1 & 9 \\ -1 & 9 & -1 \\ 9 & -1 & 4 \end{bmatrix}$$

$$A^T Y = \begin{bmatrix} \sum x_i^2 y_i \\ \sum x_i y_i \\ \sum y_i \end{bmatrix} =$$

$$A^T Y = \begin{bmatrix} 7 \\ 11 \\ -4 \end{bmatrix}$$

```
33      -1        9        7  *(1/33)  =R1
-1       9       -1       11   +1*R1
 9      -1        4       -4   -9*R1

 1  -1/33   3/11   7/33    +1/33*R2
 0 296/33  -8/11 370/33  *(33/296)  =R2
 0  -8/11  17/11 -65/11    +8/11*R2

 1       0  10/37     1/4   -10/37*R3
 0       1  -3/37     5/4   +3/37*R3
 0       0  55/37      -5  *(37/55)  =R3
```

```
1        0        0  51/44
0        1        0  43/44
0        0        1 -37/11  in rref
```

$$\boxed{y = \frac{51}{44}x^2 + \frac{43}{44}x - \frac{37}{11}}$$