GG250 Lab 9

We will try to fit a line

$$y = ax + b$$

to a data set of points in the *x*-*y* plane

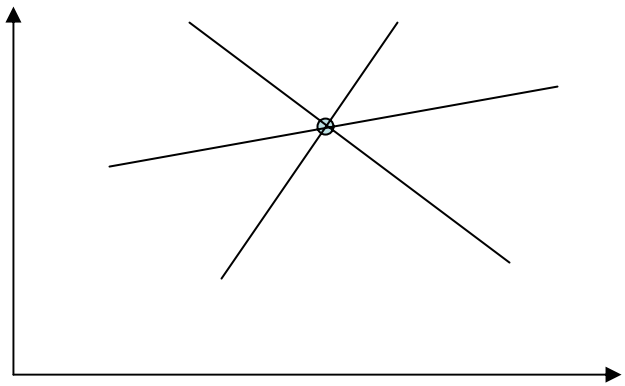
$$(x_i, y_i),\ i = 1, n$$

Unless all the points lie exactly on a line, how can we uniquely find such a line?

Case 1: $n = 1$

- We have one data point. Can we fit a unique line?

- No, any line that goes through the single point will satisfy our data constraint.



Case 2: $n = 2$

- We have two data points. Can we fit a unique line?

- Yes. The two points uniquely define the line. We set up **AX** = **B** as below and solve it for *a* and *b*

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

Case 3: $n = 3$

- We have three data points. Can we fit a unique line?

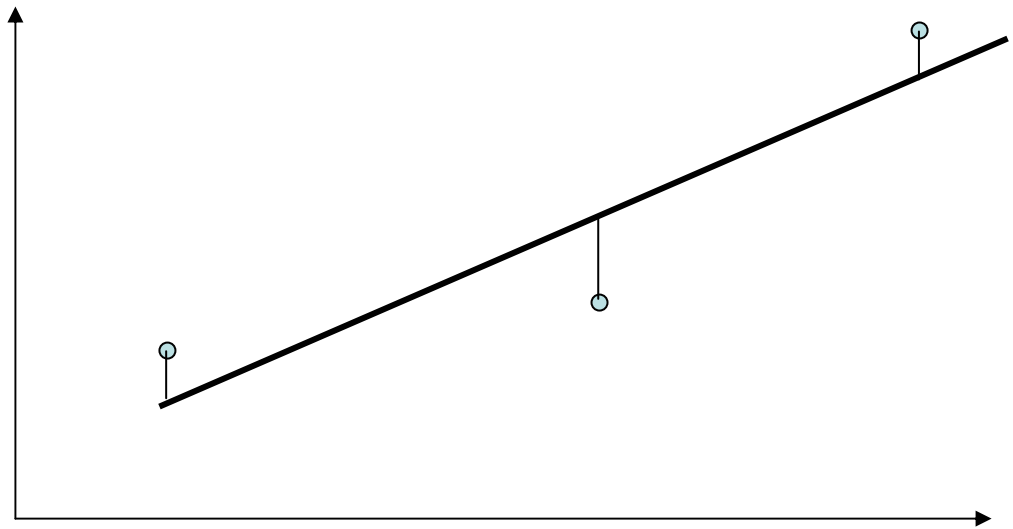
- Yes, if all three points lie of a straight line (then only two of them are required to solve the case for $n = 2$), else we have too much information.

How to get a unique solution

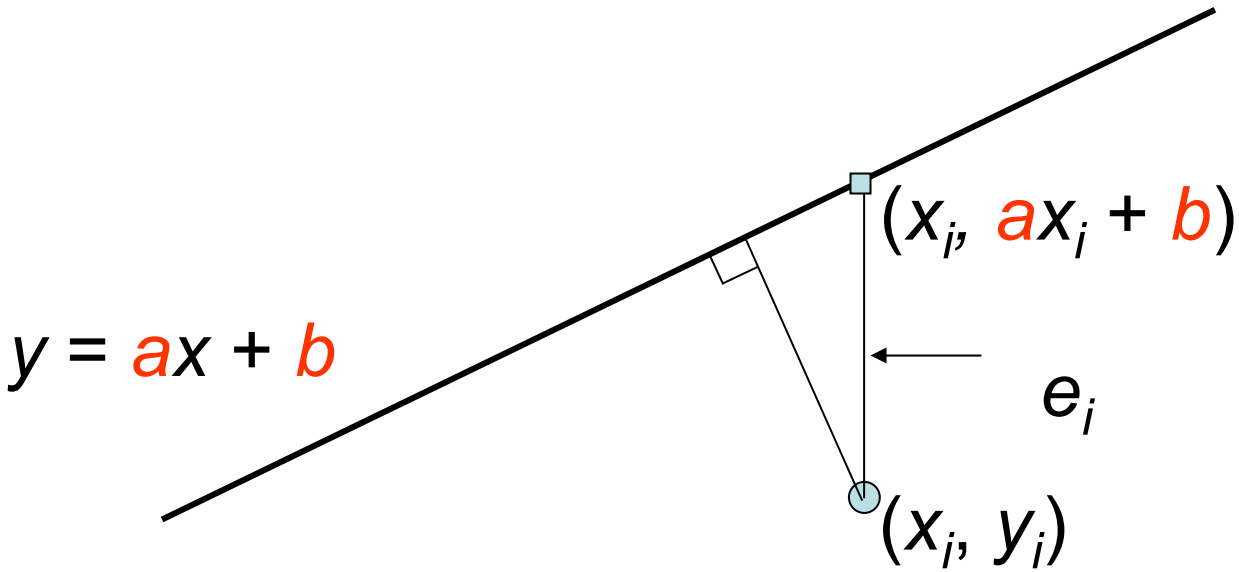
- Pick two points and eyeball a trend
- Pick all possible pairs of points, solve for *a* and *b* for each pair, and average the results.
- Look at the difference between predictions of a given line $y = ax + b$ and compare to observations, and use these differences to determine a solution

Quantifying the misfit

- Let us measure the distance between a line and a data point.



The vertical misfit



Deciding how to use misfits

- So the misfit at point # *i* is

$$e_i = y(x_i) - y_i = ax_i + b - y_i$$

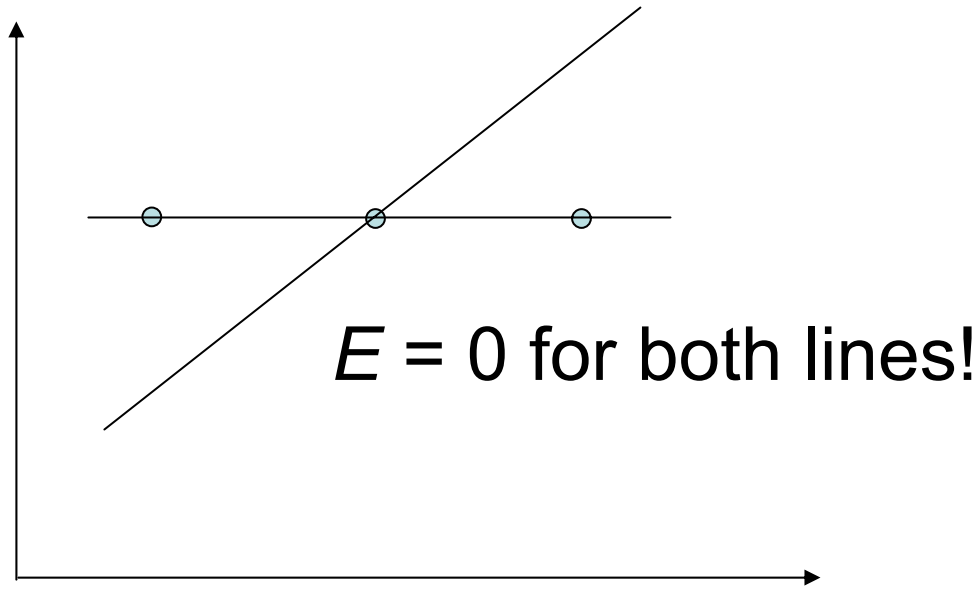
- If we want to use all the points we could
  - Sum up all of them, i.e. E = sum of all e_i
  - Give some more weight that others in the sum
  - Pick a representative misfit, i.e., $E = e_k$ for some k

and then find *a* and *b* so that E is as small as possible.

Which do we use?

- Thought experiment: Why won't this work?

$$E = \sum_{i=1}^{n} e_i$$



E = 0 for both lines!

Classical Solution

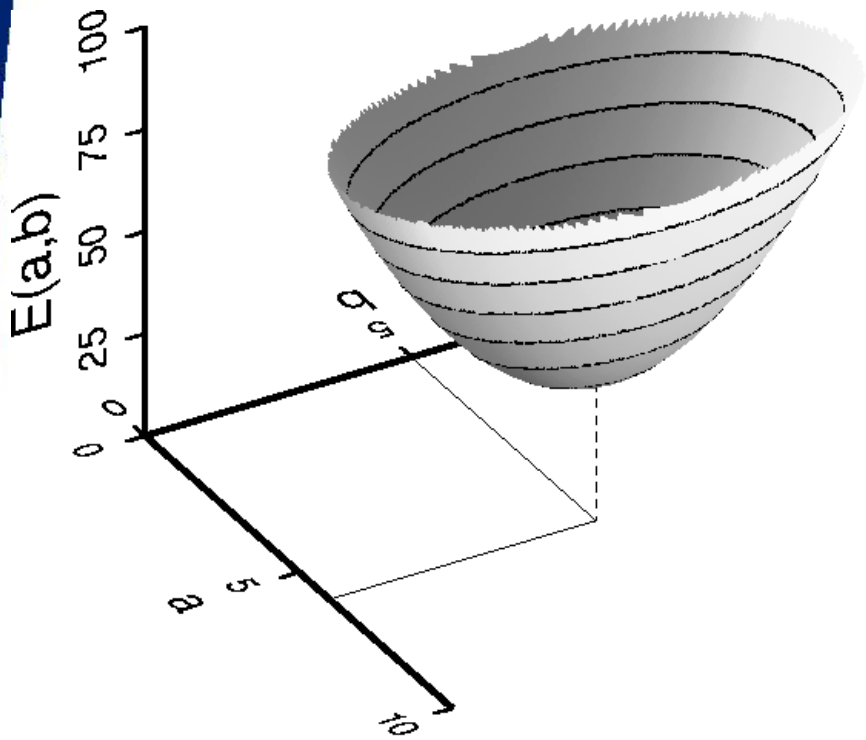
- The classical approach is to penalize solutions that give large misfits by taking the square of the lengths:

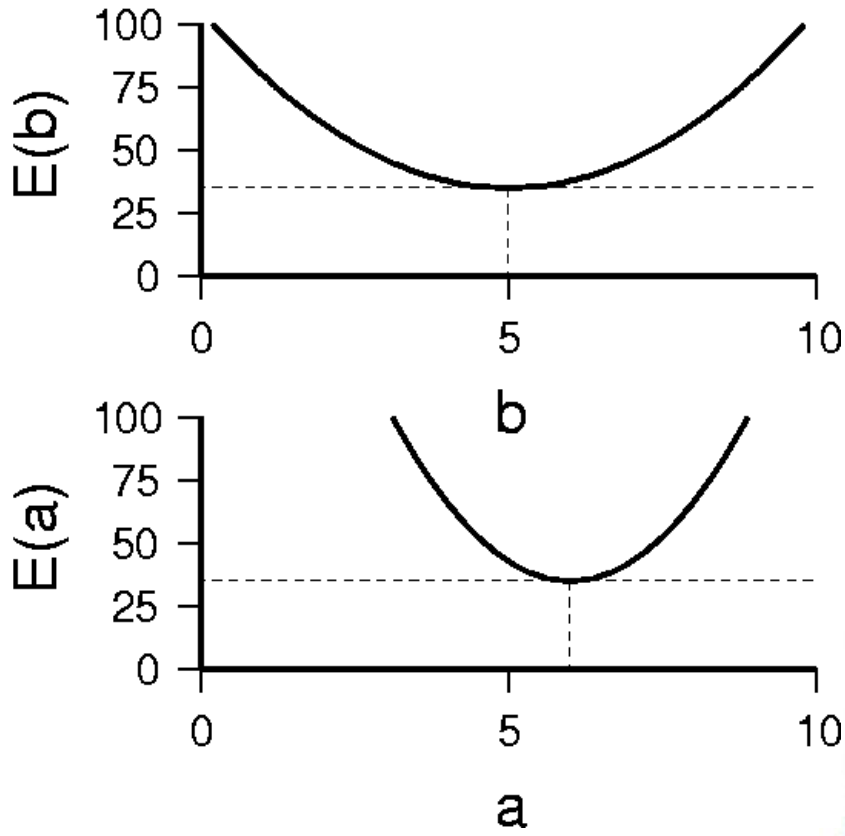
$$E = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(ax_i + b - y_i \right)^2$$

How many variables is *E* a function of?

$$E(a,b) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(ax_i + b - y_i \right)^2$$

Envisioning the misfit function, $E(a,b)$





Classical solution

- Setting the slope of *E* in the two directions to zero yields two equations in two unknowns:

$$\frac{\partial E}{\partial a} = 0$$
$$\frac{\partial E}{\partial b} = 0$$

- Thus, we are back to **AX** = **B** for a 2x2 linear system for *a* & *b* which we know how to solve.
- However, others misfit functions *E* may require a brute force solution. This is our choice.