Math 2270 - Lecture 23 : Introduction to Least Squares Approximation

Dylan Zwick

Fall 2012

This lecture and the next will cover *section 4.3* from the textbook.

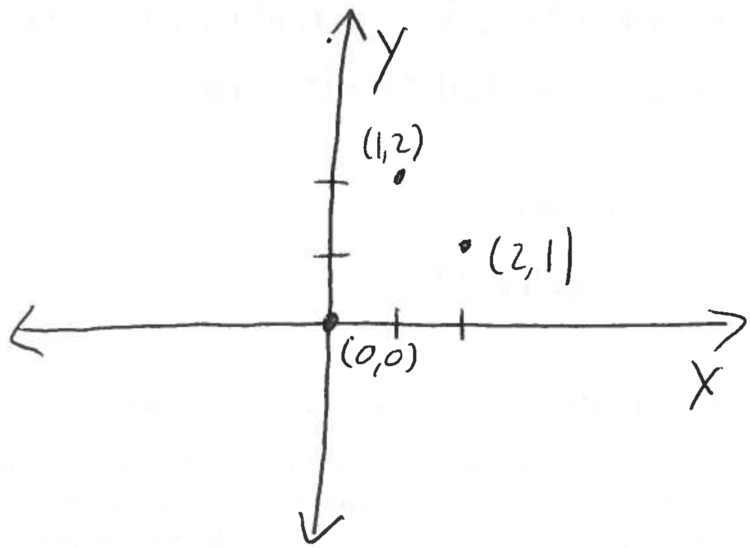
Suppose we have m points on the xy-plane, and we want to find a line that goes through them. If $m > 2$ this may not, and in fact very likely will not, be possible. Some lines will be closer to going through these m points than others, and so we can still try to find the line that gets closest to going through all the points. To do this, we'll have to first precisely define what we mean by "closest", and then with this definition in hand figure out a way to find this closest line. Those are the goals of this lecture.

The assigned problems for this section are:

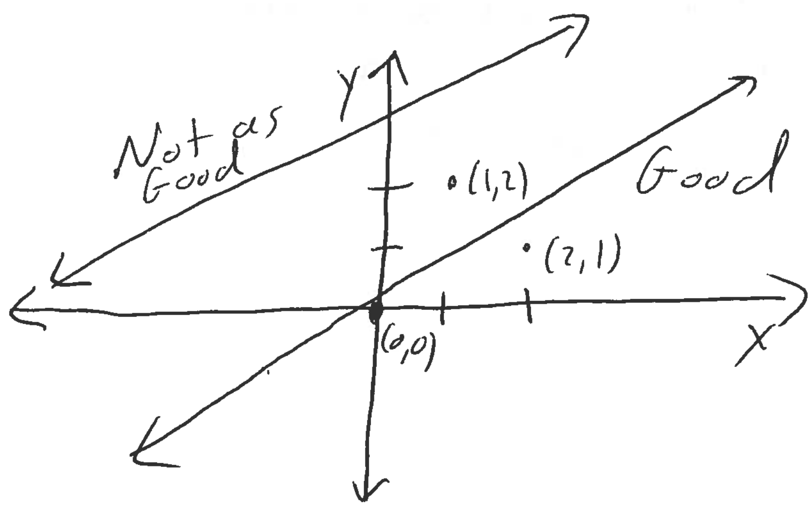
Section 4.3 - 1, 3, 9, 10, 12

1 A Very Common Problem

Suppose we take the following points on the xy-plane: $(0,0),(1,2),(2,1)$.



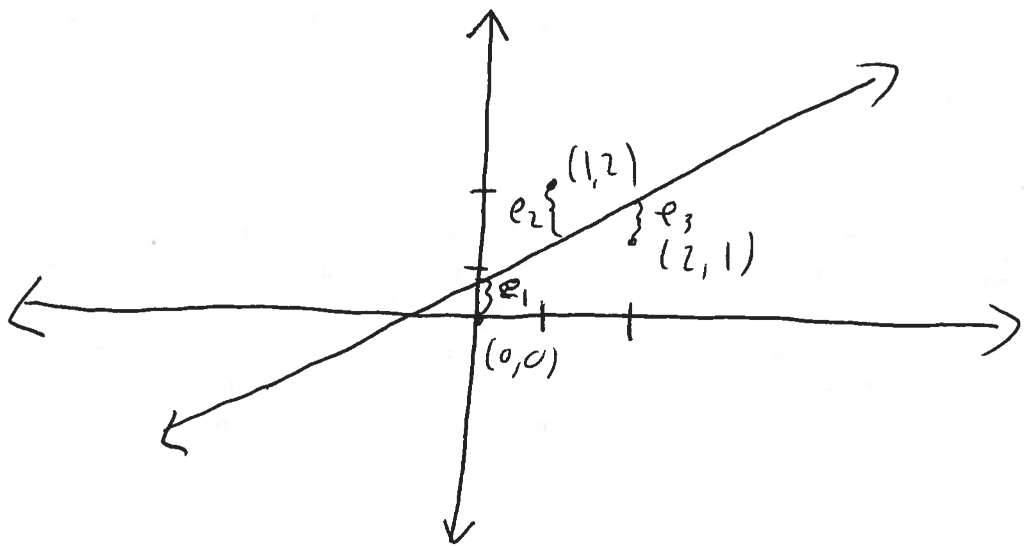
These three points are not collinear, which means there's no line that goes through all three of them. However, some lines are obviously closer to going through all three points than others:



Any non-vertical line in the xy-plane can be written as $Dx + C$, where

D and C are two numbers[1]. Suppose we're given a line $Dx + C$, and we want to know how close this line gets to our three points. We first need to precisely define how we're measuring closeness. At the three x-values of our points, $0, 1, 2$ the line will have thre y-values $C, D + C, 2D + C$. The y-values of our points for these x-values are $0, 2, 1$, respectively. At each of these x-values are line will be off by an amount:

$$\begin{aligned} e_1 &= C - 0, \\ e_2 &= (D + C) - 2, \\ e_3 &= (2D + C) - 1. \end{aligned}$$



These three values e_1, e_2, e_3 are our *error* terms, and we want to minimize them. But, what do we mean by minimize them? We do *not* mean we want to minimize the sum $e_1+e_2+e_3$. This is because each error term could be either negative or positive, and we don't want a large positive error to cancel a large negative error. Two large errors of opposite sign does not a good approximation make! So, how do we get around this problem? We minimize the *sum of squares*. That is to say, we want to minimize $e_1^2+e_2^2+e_3^2$.

OK, we've defined precisely what we mean by the "closest line" for our particular problem. But, how do we find this closest line? That's where linear algebra, and the idea of projections, comes in.

[1]The slope and y-intercept.

Finding a line that goes through our three points is the same as finding a vector $\hat{\mathbf{x}} = \begin{pmatrix} C \\ D \end{pmatrix}$ that solves:

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} C \\ D \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}$$

This, however, is impossible. There is no vector $\hat{\mathbf{x}}$ that solves the above equation. The vector $\mathbf{b} = \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}$ is not in the column space of the matrix $A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}$.

So, what we do instead is we get the next best thing. Stated more precisely, we want to find the vector in the column space of A, $A\hat{\mathbf{x}}$, that is *closest* to $\mathbf{b}$. What will this closest approximation be? It will be the projection $\mathbf{p}$ of $\mathbf{b}$ onto the column space of A. Our error terms e_1, e_2, e_3 will be the components of the error vector $\mathbf{e} = \mathbf{b} - A\hat{\mathbf{x}}$, and the projection will minimize the length of this vector. If it minimizes the length it minimizes the square of the length, and the square of the length is the sum of the squares of our error terms!

The vector $\hat{\mathbf{x}}$ that we want to find will be the vector that solves the equation $\mathbf{p} = A\hat{\mathbf{x}}$. In the last section we derived that this is the vector that solves the equation $(A^T A)\hat{\mathbf{x}} = A^T\mathbf{b}$. This vector $\hat{\mathbf{x}}$ gives us the coefficients for the line closest to our points.

Example - Calculate and draw the line that gets closest to going through our three points $(0,0), (1,2), (2,1)$.

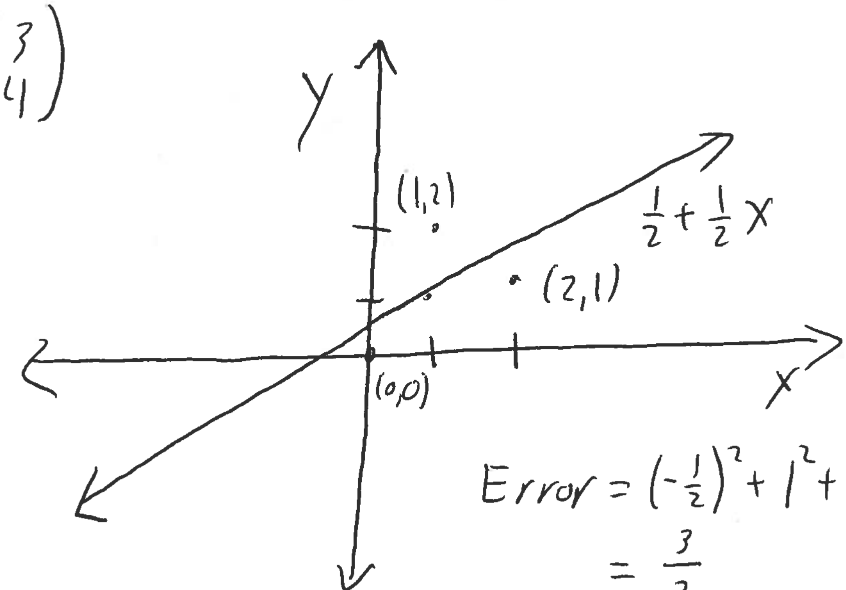
$$A^TA = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 3 \\ 3 & 5 \end{pmatrix} \qquad A^T\vec{b} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix}\begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 3 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} C \\ D \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$$

$$\Rightarrow C = D = \frac{1}{2}$$

Line

$$\boxed{\frac{1}{2} + \frac{1}{2}x}$$



$$\text{Error} = \left(-\tfrac{1}{2}\right)^2 + 1^2 + \left(-\tfrac{1}{2}\right)^2 = \frac{3}{2}$$

2 The General Problem

In general we may have many, many points, and we want to find the line that best fits them. Let's say the points are $(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)$. Then our matrix A will be:

$$A = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{pmatrix}$$

and our vector $\mathbf{b}$ will be:

$$\mathbf{b} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}.$$

We want to find the vector $\hat{\mathbf{x}}$ that solves $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$. Writing $\hat{\mathbf{x}} = \begin{pmatrix} C \\ D \end{pmatrix}$ and multiplying out A^TA and $A^T\mathbf{b}$ we get:

$$\begin{pmatrix} m & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} \begin{pmatrix} C \\ D \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}.$$

To find the best fit line we find the values of C and D that solve the above equation.