

## SEC. 5.4 LEAST SQUARES

We noticed that we often come across matrices with no solution. In class we have ignored them for the most part, but in real life we can't. One way to "solve" this is to throw out data points, but this results in large errors. Lets think of a way to minimize this average error instead of throwing out data points.

We think of a squared error: Suppose we have data points  $2x_1 = b_1$ ,  $3x_2 = b_2$ ,  $4x_3 = b_3$ . One way to minimize the error between  $x$  and the data points is to let

$$E^2 = (2x - b_1)^2 + (3x - b_2)^2 + (4x - b_3)^2.$$

If there was an exact solution then  $E^2 = 0$ . In the more likely case  $E^2$  will be a parabola. We can find the minimum of a parabola just by taking the derivative and finding critical points, then doing the max and min test.

$$\frac{dE^2}{dx^2} = 2[2(2x - b_1) + 3(3x - b_2) + 4(4x - b_3)] = 0.$$

Solving for  $x$  gives us

$$x = \frac{2b_1 + 3b_2 + 4b_3}{2^2 + 3^2 + 4^2} = \frac{a^T b}{a^T a},$$

which is called the least square "solution" in one variable. In order to separate the notation, we shall call the least squares solution  $\hat{x}$  instead of  $x$ .

For multiple variables, all vectors perpendicular to the column space lie on the left nullspace. Thus the error vector  $e = b - A\hat{x}$  must be in the nullspace of  $A^T$ :  $A^T(b - A\hat{x}) = 0 \Rightarrow A^T A\hat{x} = A^T b$ . Then  $\hat{x} = (A^T A)^{-1}(A^T b)$ . And the projection itself would be  $p = A\hat{x} = A(A^T A)^{-1}(A^T b)$ .

Ex: Consider

$$A = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 0 & 0 \end{pmatrix}, \quad b = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}$$

Notice that  $Ax = b$  has no solution. So we try to find a least squares solution by using the normal equation:  $A^T A\hat{x} = A^T b$ .

$$A^T A = \begin{pmatrix} 1 & 1 & 0 \\ 2 & 3 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 5 \\ 5 & 13 \end{pmatrix}$$

and

$$A^T b = \begin{pmatrix} 1 & 1 & 0 \\ 2 & 3 & 0 \end{pmatrix} \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} = \begin{pmatrix} 9 \\ 23 \end{pmatrix}$$

We can either solve this through Gaussian elimination or by inverting the matrix. For this particular problem we chose to invert.

$$\hat{x} = (A^T A)^{-1}(A^T b) = \begin{pmatrix} 13 & -5 \\ -5 & 2 \end{pmatrix} \begin{pmatrix} 9 \\ 23 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

Then the projection is

$$p = A\hat{x} = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 5 \\ 0 \end{pmatrix}$$

Ex: Consider

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \\ 1 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 2 \\ 0 \\ -3 \end{pmatrix}$$

First we find the normal equation  $A^T A x = A^T b$ ,

$$A^T A = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 6 & 5 \\ 5 & 6 \end{pmatrix}$$

And

$$A^T b = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 0 \\ -3 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

We can see the solution without even inverting or doing Gaussian elimination:  $\hat{x} = (1, -1)$ . Notice that gives us the line  $y = -x$ .

Also, the projection is

$$p = A\hat{x} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$$

Ex: Consider

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \quad b = \begin{pmatrix} 4 \\ -1 \\ 0 \\ 1 \end{pmatrix}$$

First we find the normal equation

$$A^T A = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & 2 \\ 2 & 2 & 3 \end{pmatrix}$$

and

$$A^T b = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 4 \\ -1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 0 \\ 3 \end{pmatrix}$$

Then we solve the normal equation

$$\left( \begin{array}{ccc|c} 3 & 2 & 2 & 4 \\ 2 & 3 & 2 & 0 \\ 2 & 2 & 3 & 3 \end{array} \right) = \left( \begin{array}{ccc|c} 6 & 4 & 4 & 8 \\ 6 & 9 & 6 & 0 \\ 6 & 6 & 9 & 9 \end{array} \right) = \left( \begin{array}{ccc|c} 6 & 4 & 4 & 8 \\ 0 & 5 & 2 & -8 \\ 0 & 2 & 5 & 1 \end{array} \right) = \left( \begin{array}{ccc|c} 6 & 4 & 4 & 8 \\ 0 & 10 & 25 & 5 \\ 0 & 10 & 4 & -16 \end{array} \right) = \left( \begin{array}{ccc|c} 6 & 4 & 4 & 8 \\ 0 & 10 & 25 & 5 \\ 0 & 0 & -21 & -21 \end{array} \right).$$

Then our least square solution for  $\hat{x} = (2, -2, 1)$ , and the projection is

$$p = A\hat{x} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ -1 \\ 0 \end{pmatrix}$$

Ex: What is the linear regression of the points  $(-1, 1)$ ,  $(1, 0)$ ,  $(3, -3)$ ?

Here we are given data points and we need to find the least square fit (also called regression). Lets first write this as a system of equations with  $y = mx + b$ ,

$$\begin{aligned}b + -m &= 1 \\b + m &= 0 \\b + 3m &= -3\end{aligned}$$

Then the matrix form of this is

$$\begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ -3 \end{bmatrix}$$

Then we find the normal equation

$$A^T A = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 3 & 11 \end{bmatrix}$$

and

$$A^T b = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -3 \end{bmatrix} = \begin{bmatrix} -2 \\ -10 \end{bmatrix}$$

Then we have

$$(A^T A) \begin{pmatrix} b \\ m \end{pmatrix} = A^T b \Rightarrow b = 1/3, m = -1 \Rightarrow \hat{y} = -\hat{x} + 1/3.$$

Ex: What is the linear regression for  $(-2, 1)$ ,  $(-1, 2)$ ,  $(0, 1)$ ,  $(1, 2)$ ,  $(2, 1)$ .

Again, we fill in the matrix for least squares

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

Then our normal equation becomes

$$A^T A = \begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix}, \quad \text{and} \quad A^T b = \begin{bmatrix} 7 \\ 0 \end{bmatrix}$$

So,  $b = 7/5$  and  $m = 0$ , then  $\hat{y} = 7/5$ , which is a horizontal line.

For a quadratic fit we simply fill in the quadratic equation  $y = ax^2 + bx + c$  (I'll give you this on the HW). What would we do for an exponential fit?