Math 2270-6

Notes of 11/12/19

- Tying up loose end from yesterday. Suppose

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_n \end{bmatrix}$$

is $m \times n$ and

$$A = QR$$

where

$$Q = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \ldots & \mathbf{q}_n \end{bmatrix}$$

is $m \times n$ with orthonormal columns and R is $n \times n$ upper triangular.

- Then the first k columns of A span the same space as the first k columns of Q.

- In principle, Q can be computed by applying the Gram-Schmidt process to the columns of A.

- Since

$$Q^T Q = I_n$$

with I_n being the $n \times n$ identity matrix we can compute R by

$$R = Q^T A.$$

6.5 Least Squares Problems

- We start with a somewhat elaborate example. You may have seen it before in Calculus, and we'll use Calculus notation. Then we will generalize it, turn it into a Linear Algebra Problems, and change the notation to Linear Algebra notation.

- On your calculator there may be a "linear regression" button.

- What does it do?

- It computes a line that approximates a set of specified points in an optimal sense. The points may represent, for example, inaccurately measured values.

- Suppose we are given the points $(2, 2)$, $(5, 5)$, and $(7, 5)$. There is no line that passes through all three points, but we can represent them approximately by a line, as shown in Figure 1.

- How do we find that line?

- We use three points in this example for computational simplicity, there could be many more points!

- Lets present our line as

$$y = L(x) = mx + b$$

as usual.

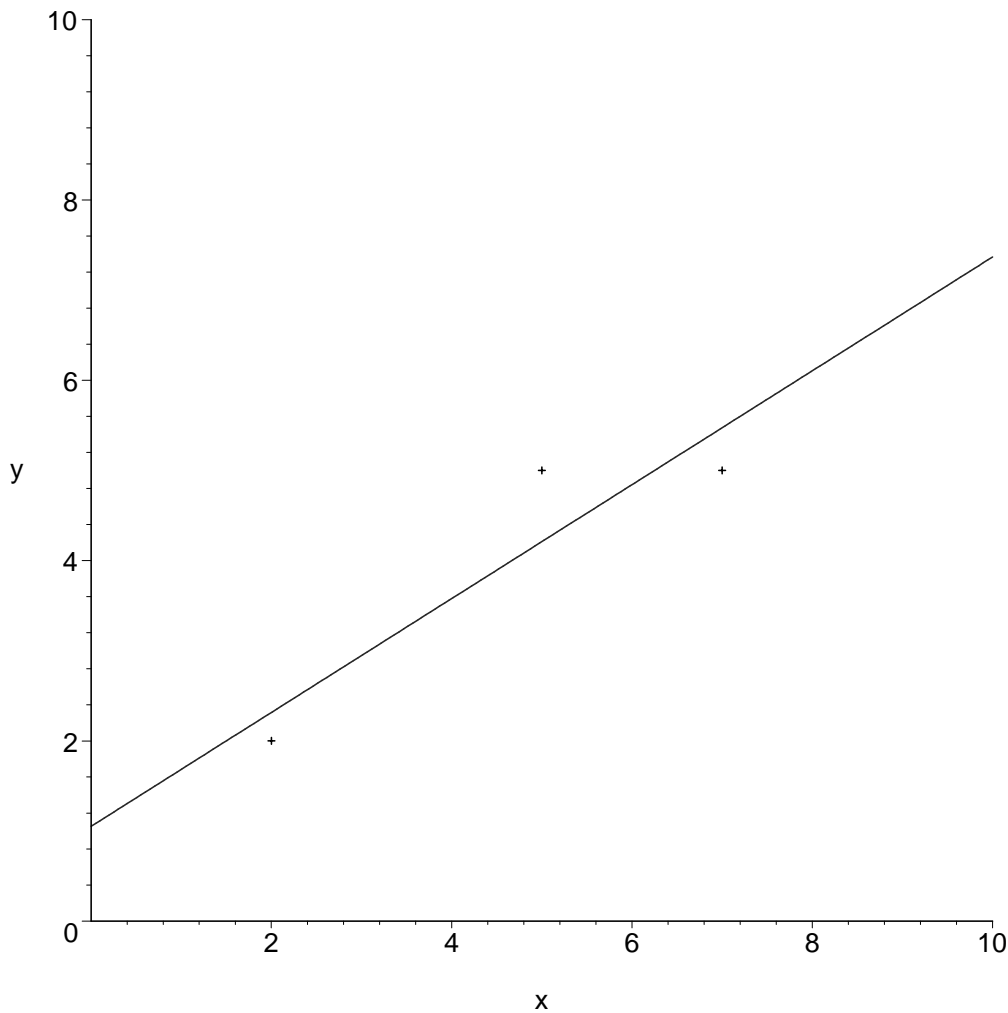- The key idea is to minimize the sum of squares:

Figure 1. 3 points and a line.

$$
\begin{aligned}
F(m, b) \quad = \quad & \left(2 - (2m + b)\right)^2 \\
+ \quad & \left(5 - (5m + b)\right)^2 \qquad\qquad (1) \\
+ \quad & \left(5 - (7m + b)\right)^2 \quad = \quad \min
\end{aligned}
$$

- Thus we want to minimize a function of two variables. In Calculus we learned that the way to do is to compute the gradient, set it to zero, and then solve the resulting system.

$$\nabla F = \begin{bmatrix} F_m(m,b) \\ F_b(m,b) \end{bmatrix} = \mathbf{0}.$$

- In our case, since we are minimizing a sum of squares, we will get a **linear system**.

- We could expand the squares in (1) and then differentiate, or differentiate first. The second approach is better. We get

$$
\begin{aligned}
F_m(m,b) &= -2\big(2 - (2m+b)\big) \times 2 \\
&\quad - 2\big(5 - (5m+b)\big) \times 5 \\
&\quad - 2\big(5 - (7m+b)\big) \times 7 \qquad = \quad 0
\end{aligned}
$$

i.e.,

$$-2\big[2\times2+5\times5+5\times7-m(2\times2+5\times5+7\times7)-b(2+5+7)\big] = 0.$$

This becomes

$$78m + 14b = 64. \qquad\qquad (2)$$

- similarly,

$$
\begin{aligned}
F_b(m,b) &= -2\big(2 - (2m+b)\big) \\
&\quad - 2\big(5 - (5m+b)\big) \\
&\quad - 2\big(5 - (7m+b)\big) \qquad = \quad 0
\end{aligned}
$$

i.e.,

$$-2\big[2\times2+5\times5+5\times7-m(2\times2+5\times5+7\times7)-b(2+5+7)\big] = 0.$$

This gives the equation

$$14m + 3b = 12 \qquad (3)$$

- The equations (2) and (3) are two linear equations in the 2 unknowns m and b. Solving them gives

$$m = \frac{12}{19} \quad \text{and} \quad b = \frac{20}{19}.$$

- The points and the line

$$y = \frac{12}{19}x + \frac{20}{19}$$

are shown in Figure 1.

- We might be given hundreds of points. The calculation just indicated would become quite tedious. So let's do the problem in general.

- Suppose we are given n points (x_i, y_i), $i = 1, \ldots, n$.

- We want to find m and b such that

$$y_i \approx mx_i + b, \qquad i = 1, \ldots, n \qquad (4)$$

- To that end we pick m and b so as to minimize

$$F(m, b) = \sum_{i=1}^{n} \left(y_i - (mx_i + b)\right)^2 = \min.$$

- As usual we compute the partial derivatives and set them equal to zero:

$$\frac{\partial}{\partial m} F(m, b) = -2 \sum_{i=1}^{n} \left(y_i - (mx_i + b) \right) x_i = 0$$

and

$$\frac{\partial}{\partial b} F(m, b) = -2 \sum_{i=1}^{n} \left(y_i - (mx_i + b) \right) = 0$$

- Dividing by -2, distributing the sums, and collecting the m and b terms on one side and the constant terms on the other side gives the linear system

$$m \sum_{i=1}^{n} x_i^2 \quad + \quad b \sum_{i=1}^{n} x_i \quad = \quad \sum_{i=1}^{n} x_i y_i$$

$$m \sum_{i=1}^{n} x_i \quad + \quad b \sum_{i=1}^{n} 1 \quad = \quad \sum_{i=1}^{n} y_i$$

- This is the linear system your calculator solves when you press the linear regression button.

- In our previous example we have the data

x_i	y_i
2	2
5	5
7	5

and get

$$
\begin{array}{rcccc}
\sum x_i^2 & = & 2^2 + 5^2 + 7^2 & = & 78 \\
\sum x_i & = & 2 + 5 + 7 & = & 14 \\
\sum 1 & = & 1 + 1 + 1 & = & 3 \\
\sum x_i y_i & = & 2 \times 2 + 5 \times 5 + 7 \times 5 & = & 64 \\
\sum y_i & = & 2 + 5 + 5 & = & 12
\end{array}
$$

which leads to the same linear system

$$
\begin{array}{rcrcl}
78m & + & 14b & = & 64 \\
14m & + & 3b & = & 12
\end{array}
$$

as before.

- Of course, instead of a linear function you could use a quadratic function. You'll get 3 equations in 3 unknowns.

- The concept we discussed is much more general. You could consider polynomials of degree greater than 2 or even non-polynomial functions.

- Let's take a different tack. Suppose all the points actually are on the lines. Then we would have the equations

$$
y_i = mx_i + b, \qquad i = 1, \ldots, n \qquad (5)
$$

- We can write this as the linear system

$$
\begin{bmatrix}
x_1 & 1 \\
x_2 & 1 \\
\vdots & \vdots \\
x_n & 1
\end{bmatrix}
\begin{bmatrix}
m \\
b
\end{bmatrix}
=
\begin{bmatrix}
y_1 \\
y_2 \\
\vdots \\
y_n
\end{bmatrix}
$$

- However, minimizing the sum of squares of the difference between the left and right sides of the equations (5) in this case gives the **Least Squares Problem**

$$\left\| \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} - \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \right\|^2 = \min$$

- Also notice that the previously obtained linear system

$$m \sum_{i=1}^n x_i^2 \quad + \quad b \sum_{i=1}^n x_i \quad = \quad \sum_{i=1}^n x_i y_i$$

$$m \sum_{i=1}^n x_i \quad + \quad b \sum_{i=1}^n 1 \quad = \quad \sum_{i=1}^n y_i$$

can be rewritten in terms of our matrix notation as

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}^T \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}^T \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Least Squares in General

- Let's take a fresh start (and also make a change in notation). Suppose we have an overdetermined linear system

$$A\mathbf{x} = \mathbf{b} \tag{6}$$

- Here A is $m \times n$, $\mathbf{x}$ is in $\mathbb{R}^n$, $\mathbf{b}$ is in $\mathbb{R}^m$, and $m \geq n$ (and typically, $m > n$).

- Usually, the system (6) will not have a solution. In that case, the next best thing is to solve the alternative problem

$$\|A\mathbf{x} - \mathbf{b}\| = \min$$

- In other words (the words of our textbook), we want to find a vector $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$$

for all $\mathbf{x}$ in $\mathbb{R}^n$.

- The textbook calls such an $\hat{\mathbf{x}}$ a **Least Squares Solution** of
$$A\mathbf{x} = \mathbf{b}.$$

- I would call it a solution of

$$\|A\mathbf{x} - \mathbf{b}\| = \min.$$

We will soon see that $\hat{x}$ is unique if the columns of A are linearly independent.

- First: **Theorem 13** (p. 363) The set of least square solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the **normal equations**

$$A^T A\mathbf{x} = A^T \mathbf{b}.$$

- Before seeing why this is true, let's go back to our introductory example.

- There

$$A = \begin{bmatrix} 2 & 1 \\ 5 & 1 \\ 7 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 2 \\ 5 \\ 5 \end{bmatrix}.$$

- We get

$$A^T A\mathbf{x} = \begin{bmatrix} 78 & 14 \\ 14 & 3 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 64 \\ 12 \end{bmatrix}$$

- This is the same set of equations as before, with the same solution.

- So let's see why Theorem 14 is true

- We can tell more than Theorem 13!

- **Theorem 14** (p. 365) Let A be an $m \times n$ matrix. The following statements are logically equivalent. (This means they are either all true or all false):

a. The equation $A\mathbf{x} = \mathbf{b}$ has a unique least square solution for each $\mathbf{b}$ in $\mathbb{R}^m$.

b. The columns of A are linearly independent.

c. The matrix $A^T A$ is invertible.

Using the QR factorization