Math 2270-6

Notes of 11/12/19

• Tying up loose end from yesterday. Suppose

\[
A = \begin{bmatrix} a_1 & a_2 & \ldots & a_n \end{bmatrix}
\]

is \( m \times n \) and

\[
A = QR
\]

where

\[
Q = \begin{bmatrix} q_1 & q_2 & \ldots & q_n \end{bmatrix}
\]

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!

- Let's 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.

\[ F(m, b) = (2 - (2m + b))^2 + (5 - (5m + b))^2 + (5 - (7m + b))^2 = \min \]  

- 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.
\[ \nabla F = \begin{bmatrix} F_m(m, b) \\ F_b(m, b) \end{bmatrix} = 0. \]

- 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

\[
F_m(m, b) = -2(2 - (2m + b)) \times 2 \\
- 2(5 - (5m + b)) \times 5 \\
- 2(5 - (7m + b)) \times 7 = 0
\]

i.e.,

\[-2[2 \times 2 + 5 \times 5 + 5 \times 7 - m(2 \times 2 + 5 \times 5 + 7 \times 7) - b(2 + 5 + 7)] = 0.\]

This becomes

\[78m + 14b = 64. \quad (2)\]

- similarly,

\[F_b(m, b) = -2(2 - (2m + b)) \\
- 2(5 - (5m + b)) \\
- 2(5 - (7m + b)) = 0\]

i.e.,

\[-2[2 \times 2 + 5 \times 5 + 5 \times 7 - m(2 \times 2 + 5 \times 5 + 7 \times 7) - b(2 + 5 + 7)] = 0.\]
This gives the equation

\[ 14m + 3b = 12 \]  

(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 \((x_i, y_i), \quad i = 1, \ldots, n.\)

- We want to find \( m \) and \( b \) such that

\[ y_i \approx m x_i + b, \quad i = 1, \ldots, n \]  

(4)

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

\[ F(m, b) = \sum_{i=1}^{n} (y_i - (m x_i + b))^2 = \text{min}. \]
• As usual we compute the partial derivatives and set them equal to zero:

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

and

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

• 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

\[
m \sum_{i=1}^{n} x_i^2 + b \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i
\]

\[
m \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} 1 = \sum_{i=1}^{n} y_i
\]

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

• In our previous example we have the data

\[
\begin{array}{cc}
  x_i & y_i \\
  2 & 2 \\
  5 & 5 \\
  7 & 5 \\
\end{array}
\]
and get
\[
\begin{align*}
\sum x_i^2 &= 2^2 + 5^2 + 7^2 = 78 \\
\sum x_i &= 2 + 5 + 7 = 14 \\
\sum 1 &= 1 + 1 + 1 = 3 \\
\sum x_i y_i &= 2 \times 2 + 5 \times 5 + 7 \times 5 = 64 \\
\sum y_i &= 2 + 5 + 5 = 12
\end{align*}
\]
which leads to the same linear system
\[
\begin{align*}
78m + 14b &= 64 \\
14m + 3b &= 12
\end{align*}
\]
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
  \[
y_i = mx_i + b, \quad i = 1, \ldots, n \quad (5)
  \]

- We can write this as the linear system
  \[
  \begin{bmatrix}
  x_1 & 1 \\
  x_2 & 1 \\
  \vdots & \vdots \\
  x_n & 1 \\
  \end{bmatrix}
  \begin{bmatrix}
  m \\
  b \\
  \vdots \\
  y_n \\
  \end{bmatrix}
  =
  \begin{bmatrix}
  y_1 \\
  y_2 \\
  \vdots \\
  y_n \\
  \end{bmatrix}
  \]
• 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**

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

• Also notice that the previously obtained linear system

\[
m \sum_{i=1}^{n} x_i^2 + b \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i
\]

\[
m \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} 1 = \sum_{i=1}^{n} y_i
\]

can be rewritten in terms of our matrix notation as

\[
A^T A \mathbf{x} = A^T \mathbf{b}
\]

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

\[
A^T A \mathbf{x} = \mathbf{b}
\]

\[
A^T \mathbf{b} = \mathbf{b}
\]
Least Squares in General

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

$$Ax = b$$  \hspace{1cm} (6)

- Here $A$ is $m \times n$, $x$ is in $\mathbb{R}^n$, $b$ 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

$$\|Ax - b\|^2 = \min$$

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

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

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

- The textbook calls such an $\hat{x}$ a **Least Squares Solution** of

$$Ax = b.$$  

- I would call it a solution of

$$\|Ax - b\| = \min.$$
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 \( Ax = b \) coincides with the nonempty set of solutions of the **normal equations**

\[
A^T Ax = A^T b.
\]

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

- There

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

- We get

\[
A^T Ax = \begin{bmatrix} 78 & 14 \\ 14 & 3 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 64 \\ 12 \end{bmatrix}
\]

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

- So let’s see why Theorem 14 is true
1. Calculus exercise b

$A \in \mathbb{R}^{m \times n}$

$Ax = b$

$A = [a_1, \ldots, a_n]$

$\text{col } A = \{A \in \mathbb{R}^n\}$

$\|b - Ax\| = \text{min}$

$a_i^T(b - Ax) = 0 \quad i = 1, \ldots, n$

$A^T(b - Ax) = 0 \in \mathbb{R}^n$

$A^Tb - A^TAx = 0$

$A^TAx = A^Tb$

$\|Ax - b\|^2 = \text{min}$

$A^TAx = A^Tb$

Normal Equations
* 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 $Ax = b$ has a unique least square solution for each $b$ in $\mathbb{R}^m$.

  b. The columns of $A$ are linearly independent.

  c. The matrix $A^T A$ is invertible.
Using the $QR$ factorization

$$R = Q^T A$$
$$A = QR$$

$m 	imes n$, $m 	imes n$, $m 	imes n$

Orthogonal

$$Q^T Q = I$$

$R$, $n 	imes n$ upper triangular

$$A = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix} = \begin{bmatrix} Q_1 R_1 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} = \begin{bmatrix} q_1 & q_2 & \cdots & q_m \end{bmatrix}$$

$$\|Qx\|^2 = (Qx)^T (Qx) = x^T Q^T Q x = x^T x = \|x\|^2$$

$$\min = \|A x - b\|^2 = \|Q^T (A x - b)\|^2$$

$$= \|Q^T A x - Q^T b\|^2$$

$$= \|R x - Q^T b\|^2$$

$$= \|\begin{bmatrix} R_1 x \\ 0 \end{bmatrix} - \begin{bmatrix} Q_1^T b \\ Q_2^T b \end{bmatrix}\|^2 = \min$$
\[ = \| R_1 x - Q_1^T b \|^2 + \| Q_2^T b \|^2 = \min \]

solve \[ R_1 x = Q_1^T b \]