

MA 402: Project 4

Matthew Murray, Johnathan Rhyne

Fall 2019

Instructions:

- Detailed instructions regarding submission are available on the class website¹.
- The zip file should contain three files project4.pdf, project4.tex, classnotes.sty.

Pen-and-paper exercises

The problems from this section total 30 points.

1) (10 points) Let the matrix $\mathbf{A} \in \mathbb{R}^{3 \times 2}$ have the SVD

$$\mathbf{A} = \begin{bmatrix} 4 & 0 \\ -5 & -3 \\ 2 & 6 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 & -2 & 2 \\ -2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix} \begin{bmatrix} 6\sqrt{2} & 0 \\ 0 & 3\sqrt{2} \\ 0 & 0 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}.$$

Let $\mathbf{b} = [1 \ 2 \ 3]^\top$. Compute the least squares solution in two different ways:

(a) Using the normal equation approach;

Solution:

$$\mathbf{A}^\top = \begin{bmatrix} 4 & -5 & 2 \\ 0 & -3 & 6 \end{bmatrix}$$

$$\mathbf{A}^\top \mathbf{A} = \begin{bmatrix} 4 & -5 & 2 \\ 0 & -3 & 6 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ -5 & -3 \\ 2 & 6 \end{bmatrix} = \begin{bmatrix} 16 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 25 & 15 \\ 15 & 9 \end{bmatrix} + \begin{bmatrix} 4 & 12 \\ 12 & 16 \end{bmatrix} = \begin{bmatrix} 45 & 27 \\ 27 & 45 \end{bmatrix}$$

$$\mathbf{A}^\top \mathbf{b} = \begin{bmatrix} 0 \\ 12 \end{bmatrix}$$

$$\mathbf{A}^\top \mathbf{A} \mathbf{x}_* = \mathbf{A}^\top \mathbf{b} \implies \begin{bmatrix} 45 & 27 \\ 27 & 45 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 12 \end{bmatrix}$$

Now we look at the 2 equations:

$$45x_1 + 27x_2 = 0 \quad \& \quad 27x_1 + 45x_2 = 12$$

$$45x_1 + 27x_2 = 0 \implies x_1 = \frac{-3x_2}{5} \implies 27\left(\frac{-3x_2}{5}\right) + 45x_2 = 12 \implies x_2 = \frac{5}{12}$$

¹https://github.ncsu.edu/asaibab/ma402_fall_2019/blob/master/projects.md

$$x_2 = \frac{5}{12} \Rightarrow x_1 = \frac{-3\frac{5}{12}}{5} = \frac{-1}{4}$$

Therefore,

$$\mathbf{x}_* = \begin{bmatrix} \frac{-1}{4} \\ \frac{5}{12} \end{bmatrix}$$

(b) Using the SVD of \mathbf{A} .

Solution:

$$\mathbf{A}^\dagger = \mathbf{V}_r \mathbf{\Sigma}_r^{-1} \mathbf{U}_r^\top$$

$$\mathbf{V}_r = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \mathbf{\Sigma}_r^{-1} = \frac{1}{3\sqrt{2}} \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix}, \mathbf{U}_r^\top = \frac{1}{3} \begin{bmatrix} 1 & -2 & 2 \\ -2 & 1 & 2 \end{bmatrix}$$

$$\begin{aligned} \mathbf{V}_r \mathbf{\Sigma}_r^{-1} \mathbf{U}_r^\top &= \frac{1}{18} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -2 & 2 \\ -2 & 1 & 2 \end{bmatrix} = \frac{1}{18} \begin{bmatrix} \frac{1}{2} & -1 \\ \frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} 1 & -2 & 2 \\ -2 & 1 & 2 \end{bmatrix} \\ &= \frac{1}{18} \begin{bmatrix} \frac{5}{2} & -2 & -1 \\ -3 & 0 & 3 \end{bmatrix} \end{aligned}$$

With this solution, we know that $\mathbf{x}_* = \mathbf{A}^\dagger \mathbf{b}$, meaning that \mathbf{x}_* is equal to:

$$\frac{1}{18} \begin{bmatrix} \frac{5}{2} - 4 - 3 \\ -3 + 0 + 9 \end{bmatrix} = \frac{1}{18} \begin{bmatrix} \frac{-9}{2} \\ \frac{15}{2} \end{bmatrix} = \begin{bmatrix} \frac{-1}{4} \\ \frac{5}{12} \end{bmatrix}$$

2) (10 points) Consider the line $f(t) = \alpha + \beta t$ that we seek to fit the data

$$\mathcal{D} = \{(t_1, b_1), \dots, (t_n, b_n)\}.$$

Assume that $n(\sum_{i=1}^n t_i^2) \neq (\sum_{i=1}^n t_i)^2$. Show that the coefficients α and β , obtained by solving the least squares problem, satisfy

$$\alpha = \frac{(\sum_i t_i^2)(\sum_i b_i) - (\sum_i t_i)(\sum_i b_i t_i)}{n(\sum_i t_i^2) - (\sum_i t_i)^2} \quad \beta = \frac{n(\sum_i b_i t_i) - (\sum_i t_i)(\sum_i b_i)}{n(\sum_i t_i^2) - (\sum_i t_i)^2}.$$

Solution: note that we can write \mathbf{A} , \mathbf{b} , \mathbf{x} as:

$$\mathbf{A} = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \dots & \\ 1 & t_n \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

Now, we have $\mathbf{A}^\top \mathbf{A}$ can be written as

$$\begin{bmatrix} \sum 1 & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix}$$

which when multiplied by \mathbf{x} gives us

$$\begin{bmatrix} \alpha \sum 1 + \beta \sum t_i \\ \alpha \sum t_i + \beta \sum t_i^2 \end{bmatrix}$$

Now we look at $\mathbf{A}^\top \mathbf{b}$, which when simplified gives us:

$$\begin{bmatrix} \sum b_i \\ \sum t_i b_i \end{bmatrix}$$

Next, we solve the 2 equations $\alpha \sum 1 + \beta \sum t_i = \sum b_i$ and $\alpha \sum t_i + \beta \sum t_i^2 = \sum t_i b_i$ For α and β , which gives us the 2 following equations:

$$\frac{\sum b_i - \beta \sum t_i}{n} = \alpha = \frac{\sum t_i b_i - \beta \sum t_i^2}{\sum t_i} \quad (1)$$

and

$$\frac{\sum b_i - \alpha n}{\sum t_i} = \beta = \frac{\sum t_i b_i - \alpha \sum t_i}{\sum t_i^2} \quad (2)$$

Now we solve 2 for α

$$\begin{aligned} \sum t_i^2 (\sum b_i - \alpha n) &= \sum t_i (\sum t_i b_i - \alpha \sum t_i) \\ \sum t_i^2 \sum b_i - \alpha n \sum t_i^2 &= \sum t_i \sum t_i b_i - \alpha (\sum t_i)^2 \\ \alpha ((\sum t_i)^2 - n \sum t_i^2) &= \sum t_i \sum t_i b_i - \sum t_i^2 \sum b_i \\ \alpha &= \frac{\sum t_i \sum t_i b_i - \sum t_i^2 \sum b_i}{(\sum t_i)^2 - n \sum t_i^2} \end{aligned}$$

or

$$\alpha = \frac{\sum t_i^2 \sum b_i - \sum t_i \sum t_i b_i}{n \sum t_i^2 - (\sum t_i)^2}$$

Now we solve 1 for β

$$\begin{aligned} \sum t_i (\sum b_i - \beta \sum t_i) &= n (\sum t_i b_i - \beta \sum t_i^2) \\ \sum t_i \sum b_i - \beta \sum t_i \sum t_i &= n \sum t_i b_i - \beta n \sum t_i^2 \\ \beta n \sum t_i^2 - \beta \sum t_i \sum t_i &= n \sum t_i b_i - \sum t_i \sum b_i \\ \beta (n \sum t_i^2 - \sum t_i \sum t_i) &= n \sum t_i b_i - \sum t_i \sum b_i \\ \beta &= \frac{n \sum b_i t_i - \sum t_i \sum b_i}{n \sum t_i^2 - (\sum t_i)^2} \end{aligned}$$

3) (10 points) Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ and let $\text{rank}(\mathbf{A}) = n$. Let \mathbf{x}_* be the solution of the least squares problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{Ax} - \mathbf{b}\|_2^2.$$

Define the residual corresponding to the optimal solution $\mathbf{r}_* = \mathbf{b} - \mathbf{Ax}_*$.

(a) Starting with the normal equations $\mathbf{A}^\top \mathbf{Ax}_* = \mathbf{A}^\top \mathbf{b}$, show that the $\mathbf{x}_*, \mathbf{r}_*$ also satisfy

$$\begin{bmatrix} \mathbf{I} & \mathbf{A} \\ \mathbf{A}^\top & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{r}_* \\ \mathbf{x}_* \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix}.$$

Remark: This system is known as the augmented equations and maybe beneficial to use when \mathbf{A} is sparse (i.e., it has many zero entries).

Solution: From $\mathbf{r}_* = \mathbf{b} - \mathbf{Ax}_*$, we get that $\mathbf{r}_* + \mathbf{Ax}_* = \mathbf{b} = \mathbf{Ir}_* + \mathbf{Ax}_* = \mathbf{b}$. From $\mathbf{r}_* = \mathbf{b} - \mathbf{Ax}_*$,

$$\implies \mathbf{A}^\top \mathbf{r}_* = \mathbf{A}^\top \mathbf{b} - \mathbf{A}^\top \mathbf{Ax}_*$$

We know that $\mathbf{A}^\top \mathbf{Ax}_* = \mathbf{A}^\top \mathbf{b}$

$$\therefore \mathbf{A}^\top \mathbf{r}_* = \mathbf{0}$$

We can rewrite are results as,

$$\mathbf{Ir}_* + \mathbf{Ax}_* = \mathbf{b}$$

and

$$\mathbf{A}^\top \mathbf{r}_* + \mathbf{0x}_* = \mathbf{0}$$

We can combine the results into a matrix as,

$$\begin{bmatrix} \mathbf{I} & \mathbf{A} \\ \mathbf{A}^\top & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{r}_* \\ \mathbf{x}_* \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix}.$$

(b) Show: the vectors \mathbf{Ax}_* and $\mathbf{b} - \mathbf{Ax}_*$ are orthogonal and

$$\|\mathbf{b}\|_2^2 = \|\mathbf{b} - \mathbf{Ax}_*\|_2^2 + \|\mathbf{Ax}_*\|_2^2.$$

Provide an interpretation of this result in your words.

Solution: Compute the dot product between \mathbf{Ax}_* and $\mathbf{b} - \mathbf{Ax}_*$.

$$\implies (\mathbf{Ax}_*)^\top (\mathbf{b} - \mathbf{Ax}_*)$$

$$= \mathbf{x}_*^\top \mathbf{A}^\top \mathbf{b} - \mathbf{x}_*^\top \mathbf{A}^\top \mathbf{Ax}_*$$

We know that $\mathbf{A}^\top \mathbf{Ax}_* = \mathbf{A}^\top \mathbf{b}$

$$= \mathbf{x}_*^\top \mathbf{A}^\top \mathbf{b} - \mathbf{x}_*^\top \mathbf{A}^\top \mathbf{b} = 0.$$

$$\therefore (\mathbf{Ax}_*)^\top (\mathbf{b} - \mathbf{Ax}_*) = 0$$

Also the dot product is commutative,

$$\therefore (\mathbf{b} - \mathbf{Ax}_*)^\top (\mathbf{Ax}_*) = 0$$

We know that $\|\mathbf{Ax}_*\|_2^2 = (\mathbf{Ax}_*)^\top (\mathbf{Ax}_*)$.

$$\therefore \|\mathbf{Ax}_*\|_2^2 = \mathbf{x}_*^\top \mathbf{A}^\top \mathbf{Ax}_* = \mathbf{x}_*^\top \mathbf{A}^\top \mathbf{b}$$

We also know that $\|\mathbf{b} - \mathbf{Ax}_*\|_2^2 = (\mathbf{b} - \mathbf{Ax}_*)^\top (\mathbf{b} - \mathbf{Ax}_*)$ and $\|\mathbf{b}\|_2^2 = \mathbf{b}^\top \mathbf{b}$

$$= \mathbf{b}^\top \mathbf{b} - \mathbf{b}^\top \mathbf{Ax}_* - \mathbf{x}_*^\top \mathbf{A}^\top \mathbf{b} + \mathbf{x}_*^\top \mathbf{A}^\top \mathbf{Ax}_*$$

$$= \|\mathbf{b}\|_2^2 - \mathbf{b}^\top \mathbf{Ax}_* - \mathbf{x}_*^\top \mathbf{A}^\top \mathbf{b} + \mathbf{x}_*^\top \mathbf{A}^\top \mathbf{b}$$

$$\therefore \|\mathbf{b} - \mathbf{Ax}_*\|_2^2 = \|\mathbf{b}\|_2^2 - \mathbf{b}^\top \mathbf{Ax}_*$$

We add the two equations to get,

$$\|\mathbf{b} - \mathbf{Ax}_*\|_2^2 + \|\mathbf{Ax}_*\|_2^2 = \|\mathbf{b}\|_2^2 - \mathbf{b}^\top \mathbf{Ax}_* + \mathbf{x}_*^\top \mathbf{A}^\top \mathbf{b}$$

The dot product is commutative, so $\mathbf{x}_*^\top \mathbf{A}^\top \mathbf{b} - \mathbf{b}^\top \mathbf{Ax}_* = 0$.

$$\therefore \|\mathbf{b} - \mathbf{Ax}_*\|_2^2 + \|\mathbf{Ax}_*\|_2^2 = \|\mathbf{b}\|_2^2$$

Interpretation: The result is the pythagorean theorem, where the hypotenuse is $\|\mathbf{b}\|_2$ and the two legs are $\|\mathbf{b} - \mathbf{Ax}_*\|_2$ and $\|\mathbf{Ax}_*\|_2$.

Numerical exercises

The problems from this section total 20 points, and were modified from Boyd and Vandenberghe's "Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares" Meyer's book "Matrix Analysis and Applied Linear Algebra." MATLAB users may use "backslash" for solving the least squares problem, Python users may use `numpy.linalg.lstsq`.

- 4) (10 points) The computer scientist and Intel corporation co-founder Gordon Moore formulated the law that bears his name in a magazine article published in 1965. *Moore's law* states that the number of transistors per integrated circuit roughly doubles every 1.5 to 2 years. The table below shows the number of transistors N in 13 microprocessors, and the year of their introduction.

Year	Transistors
1971	2,250
1972	2,500
1974	5,000
1978	29,000
1982	120,000
1985	275,000
1989	1,180,000
1993	3,100,000
1997	7,500,000
1999	24,000,000
2000	42,000,000
2002	220,000,000
2003	410,000,000

Based on this law, you hypothesize a model of the form

$$N \approx \theta_1 10^{\theta_2(t-1970)},$$

where t is the year and N is the number of transistors.

- (a) Find the coefficients θ_1 and θ_2 that "best fit" the data. *Hint:* This is a line fitting problem in disguise.

Solution: Define a line fitting problem by taking the common logarithm of both sides of the model:

$$\begin{aligned} \implies \log(N) &\approx \log(\theta_1 10^{\theta_2(t-1970)}) \\ \implies \log(N) &\approx \log(\theta_1) + \theta_2(t - 1970) \end{aligned}$$

We define the following: $y = \log(N)$, $\alpha = \log(\theta_1)$, $t' = t - 1970$, $\beta = \theta_2$, and $\mathbf{x} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$. We solve the line fitting problem by solving the least squares problem shown below.

$$\min_{\mathbf{x} \in \mathbb{R}^2} \|\mathbf{Ax} - \mathbf{b}\|_2^2$$

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 4 \\ 1 & 8 \\ 1 & 12 \\ 1 & 15 \\ 1 & 19 \\ 1 & 23 \\ 1 & 27 \\ 1 & 29 \\ 1 & 30 \\ 1 & 32 \\ 1 & 33 \end{bmatrix}$$

and

$$b = \begin{bmatrix} 2,250 \\ 2,500 \\ 5,000 \\ 29,000 \\ 120,000 \\ 275,000 \\ 1,180,000 \\ 3,100,000 \\ 7,500,000 \\ 24,000,000 \\ 42,000,000 \\ 220,000,000 \\ 410,000,000 \end{bmatrix}$$

- (b) Plot the data points (with black crosses) along with the “best fit” curve (as a solid red line). A semilogy scale is appropriate here.

Solution:

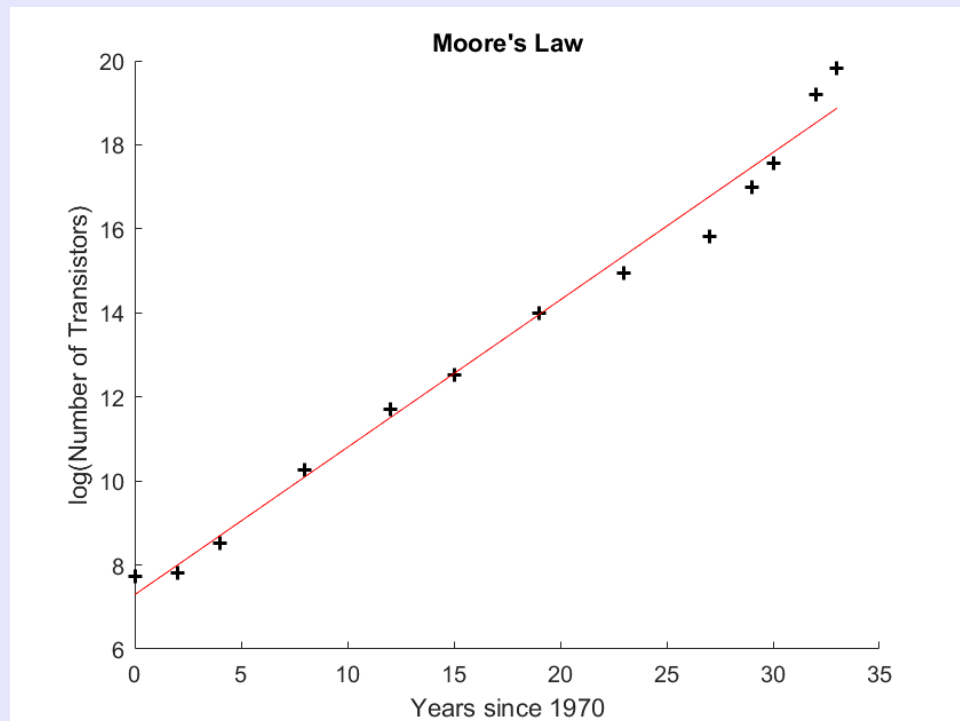


Figure 1: Number of Transistors over Time

- (c) Use your model to predict the number of transistors in a microprocessor introduced in 2015. Compare the prediction to the IBM Z13 microprocessor, released in 2015, which has around 4×10^9 transistors.

Solution: Our model predicts about 1.2018×10^{23} transistors by 2015. This estimate is dramatically higher than number of transistors in the IBM Z13 processor. The relative error between the two numbers is nearly zero.

- (d) Comment on whether the data supports Moore’s law.

Solution: The data here doesn't appear to support Moore's Law. This doesn't mean Moore's Law is not valid in the 'bigger' picture.

- 5) (10 points) Consider the time (T) it takes for a runner to complete a marathon (26 miles and 385 yards). Many factors such as height, weight, age, previous training, etc. can influence an athlete's performance, but experience has shown that the following three factors are particularly important:

$$\begin{aligned} x_1 &= \text{Ponderal index} = \frac{\text{height (in.)}}{[\text{weight (lbs.)}]^{1/3}}, \\ x_2 &= \text{Miles run during the previous 8 weeks,} \\ x_3 &= \text{Age (years).} \end{aligned}$$

A linear model hypothesizes that the time T (in minutes) is given by $T = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + r$, where r is a residual term that accounts for other factors.

T	x_1	x_2	x_3
181	13.1	619	23
193	13.5	803	42
212	13.8	207	31
221	13.1	409	38
248	12.5	482	45

- (a) Determine the least squares estimates for the coefficients $\alpha_0, \dots, \alpha_3$ from the available data.

Solution:

$$\alpha_0 = 492.0442, \alpha_1 = -23.4355, \alpha_2 = -.0761, \alpha_3 = 1.862$$

- (b) Estimate the expected marathon time for a 43-year-old runner of height 74 in., weight 180 lbs., who has run 450 miles during the previous eight weeks.

Solution: The expected marathon time is: 230.7209 mins, which is found by solving: $\alpha_0 + \frac{74}{180^{1/3}} * \alpha_1 + 450 * \alpha_2 + 43 * \alpha_3$

- (c) Your instructor (height 68 in., weight 137 lbs, and 33 years old) wants to qualify for the Boston Marathon this year. How many miles should he run in an eight week period before the marathon to qualify for the Boston marathon? (In his age group, the cutoff for qualification is 3 hours 5 minutes).

Solution: We solve: $t = \alpha_0 + x_1 * \alpha_1 + x_2 * \alpha_2 + x_3 * \alpha_3$

for x_2 where we are given the rest of the information, which gives us: $x_2 = 779.8626$ miles

For part (c), feel free to use a different person (such as yourself, someone you admire, or a famous mathematician) than the instructor. The qualification times are available at this link: <https://www.baa.org/2019-boston-marathon-qualifier-acceptances>.

Solution: For number 4

For Figure 1

```
t = [1970 1972 1974 1978 1982 1985 1989 1993 1997 1999 2000 2002 2003]; %
      years
T = transpose(t-1970); %time since 1970
n = size(T);
```

```

one = ones(n(1),n(2));
A = cat(2,one,T);
b = transpose([2250 2500 5000 29000 120000 275000 1180000 3100000 7500000
24000000 42000000 220000000 410000000]);
y= log(b);
x = (transpose(A)*A)\(transpose(A)*y);
yhat = x(1) + x(2)*T;
scatter(T,y, 'k+', 'LineWidth',1.5)
hold on
plot(T,yhat, 'r')
hold off
title("Moore's Law")
xlabel('Years_since_1970')
ylabel('log(Number_of_Transistors)')
y2015 = 10^((x(1))+x(2)*(2015-1970)); %prediction at t= 2015

```

For number 5

```

%%
syms A; %Design Matrix
syms T; %Our 'b'
syms a; %matrix of variables
syms AtA;
syms At;
syms RHS; %AtT
syms finalMatrix;
%Setting up our A

A = [1 13.1 619 23;
1 13.5 803 42;
1 13.8 207 31;
1 13.1 409 38;
1 12.5 482 45];

disp("Rank of A is: " + rank(A)) % gives if 4 which means that A has
%full column rank meaning AtA is invertible

At = transpose(A);
AtA = At * A;

T = [181;
193;
212;
221;
248];

RHS = At * T;
a = AtA\RHS; %gives us (AtA)^(-1) * AtT
%a_0 = 492.0442, a_1 = -23.4355, a_2 = -.0761, a_3 = 1.862
disp("a_0 is: " + a(1));
disp("a_1 is: " + a(2));
disp("a_2 is: " + a(3));

```



```

disp("a_3 is: " + a(4));

%%
%For part b. Height = 74 in. Weight = 180 lbs. MilesRan = 450. Age = 43
x1 = 74/(180)^(1/3);
x2 = 450;
x3 = 43;
t = a(1) + x1 * a(2) + x2 * a(3) + x3 * a(4);
% gives us 230.7209 mins
disp("Estimated time is " + t + " minutes for a person with the given
information")

%%
%For part c. Height = 68 in. Weight = 137 lbs. MilesRan = solve for. Age =
%33. Goal is 3 hrs and 5 mins or 185 mins

x1 = 68 / (137)^(1/3);
x3 = 33;
t = 185;

%Now we solve t = a(1) + x1 * a(2) + x2 * a(3) + x3 * a(4) for x2

x2 = (t - a(1) - x1 * a(2) - x3 * a(4)) / (a(3));
% gives us 779.8626 miles
disp("This person will need to run approximately " + x2 + " miles within 8
weeks of the marathon");

```