ISTA 421/521 – Homework 1

Due: Friday, September 5, 5pm (to the class D2L dropbox)
8 points total

Emanuel Carlos de Alcantara Valente

Undergraduate

All written portions of the homework must be submitted to the D2L dropbox as a .pdf.

In the following, FCMA refers to the course text: Simon Rogers and Mark Girolami (2012), *A First Course in Machine Learning*. (All questions from the book are reproduced here in full, so you do not need the book to proceed.)

For general notes on using latex to typeset math, see: `http://en.wikibooks.org/wiki/LaTeX/Mathematics`

1. **Exercise 1.1** from FCMA p.35 [0.5 pts]

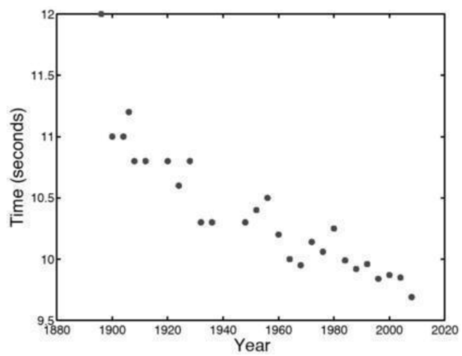

Figure 1: Reproduction of figure 1.1, Olympic men's 100m data

By examining Figure 1.1 [from p. 2 of FCMA, reproduced here], estimate (by hand / in your head) the kind of values we should expect for w_0 and w_1 (e.g., High? Low? Positive? Negative?).

Solution. We expect that w_0 will be high and positive. Through the figure, we can infer that in order to have a good model (in this case it is a linear function) it should have an intercept (where the line intercepts the t-axis in $x = 0$) high ($\sim> 11$) and positive.

The w_1 is expected to be negative and low. Through the same way, a good model will have a negative gradient because the line is descending (at least $100° < slope(line) < 170°$). Therefore, the tangent of this angle will be negative and low.

NOTE: The following three exercises (2, 3, and 4) review basic linear algebra concepts; we will review these briefly in Lecture 3.

Notation conventions:

- Script variables, such as x_{n2} and w_1 represent scalar values
- Lowercase bold-face variables, such as $\mathbf{x}_n$ and $\mathbf{w}$, represent vectors
- Uppercase bold-face variables, such as $\mathbf{X}$, represent n (rows) $\times$ m (columns) matrices
- Note that because all indexes in the following are either a value between 0, 1, ..., 9, or a scalar, n, I am representing multiple dimension indexes without a comma, as it is unambiguous; e.g., x_{32} is the element scalar value of $\mathbf{X}$ at row 3, column 2. When we have to refer to specific index values greater than 9, we'll use commas, such as $x_{32,3}$ is the scalar value in the 32nd row and 3rd column.
- '$\top$' in expressions like $\mathbf{w}^\top$ indicates the transpose operator.

2. **Exercise 1.3** from FCMA p.35 [1pt]

Show that:

$$\mathbf{w}^\top \mathbf{X}^\top \mathbf{X} \mathbf{w} = w_0^2 \left(\sum_{n=1}^{N} x_{n1}^2 \right) + 2 w_0 w_1 \left(\sum_{n=1}^{N} x_{n1} x_{n2} \right) + w_1^2 \left(\sum_{n=1}^{N} x_{n2}^2 \right),$$

where

$$\mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}, \mathbf{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \\ \vdots & \vdots \\ x_{N1} & x_{N2} \end{bmatrix}.$$

(Hint – it's probably easiest to do the $\mathbf{X}^\top \mathbf{X}$ first!)

Solution. Let's solve first $\mathbf{X}^\top \mathbf{X}$

$$\mathbf{X}^\top \mathbf{X} = \begin{bmatrix} x_{11} & x_{21} & x_{31} & \dots & x_{N1} \\ x_{12} & x_{22} & 32 & \dots & x_{N2} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \\ \vdots & \vdots \\ x_{N1} & x_{N2} \end{bmatrix} =$$

$$= \begin{bmatrix} x_{11}x_{11} + x_{21}x_{21} + \ldots + x_{N1}x_{N1} & x_{11}x_{12} + x_{21}x_{22} + \ldots + x_{N1}x_{N2} \\ x_{12}x_{11} + x_{22}x_{21} + \ldots + x_{N2}x_{N1} & x_{12}x_{12} + x_{22}x_{22} + \ldots + x_{N2}x_{N2} \end{bmatrix} = \begin{bmatrix} \sum_{n=1}^{N} x_{n1}^2 & \sum_{n=1}^{N} x_{n1}x_{n2} \\ \sum_{n=1}^{N} x_{n1}x_{n2} & \sum_{n=1}^{N} x_{n2}^2 \end{bmatrix}$$

Now, we will calculate: $\mathbf{w}^\top \mathbf{X}^\top \mathbf{X}$

$$\mathbf{w}^\top \mathbf{X}^\top \mathbf{X} = \begin{bmatrix} w_0 & w_1 \end{bmatrix} \begin{bmatrix} \sum_{n=1}^{N} x_{n1}^2 & \sum_{n=1}^{N} x_{n1}x_{n2} \\ \sum_{n=1}^{N} x_{n1}x_{n2} & \sum_{n=1}^{N} x_{n2}^2 \end{bmatrix} =$$

$$= \begin{bmatrix} w_0 \sum_{n=1}^{N} x_{n1}^2 + w_1 \sum_{n=1}^{N} x_{n1}x_{n2} & w_0 \sum_{n=1}^{N} x_{n1}x_{n2} + w_1 \sum_{n=1}^{N} x_{n2}^2 \end{bmatrix}$$

Finally:

$$\mathbf{w}^\top \mathbf{X}^\top \mathbf{X}\mathbf{w} = \begin{bmatrix} w_0 \sum_{n=1}^N x_{n1}^2 + w_1 \sum_{n=1}^N x_{n1}x_{n2} & w_0 \sum_{n=1}^N x_{n1}x_{n2} + w_1 \sum_{n=1}^N x_{n2}^2 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} =$$

$$= \begin{bmatrix} w_0^2 \sum_{n=1}^N x_{n1}^2 + w_0 w_1 \sum_{n=1}^N x_{n1}x_{n2} + w_0 w1 \sum_{n=1}^N x_{n1}x_{n2} + w_1^2 \sum_{n=1}^N x_{n2}^2 \end{bmatrix} =$$

$$= \mathrm{w}_0^2 \left(\sum_{n=1}^N x_{n1}^2 \right) + 2 w_0 w_1 \left(\sum_{n=1}^N x_{n1}x_{n2} \right) + w_1^2 \left(\sum_{n=1}^N x_{n2}^2 \right)$$

3. **Exercise 1.4** from FCMA p.35 [1pt]

   Using $\mathbf{w}$ and $\mathbf{X}$ as defined in the previous exercise, show that $(\mathbf{X}\mathbf{w})^\top = \mathbf{w}^\top \mathbf{X}^\top$ by multiplying out both sides.

   Solution.

   Multiplying the left side:

$$(\mathbf{X}\mathbf{w})^\top = \left(\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \\ \vdots & \vdots \\ x_{n1} & x_{n2} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} \right)^\top = \left(\begin{bmatrix} x_{11}w_0 + x_{12}w_1 \\ x_{21}w_0 + x_{22}w_1 \\ x_{31}w_0 + x_{32}w_1 \\ \vdots \\ x_{n1}w_0 + x_{n2}w_1 \end{bmatrix} \right)^\top = \begin{bmatrix} w_0 x_{11} + w_1 x_{12} & w_0 x_{21} + w_1 x_{22} & \dots & w_0 x_{n1} + w_1 x_{n2} \end{bmatrix}$$

   Multiplying the right side:

$$\mathbf{w}^\top \mathbf{X}^\top = \begin{bmatrix} w_0 & w_1 \end{bmatrix} \begin{bmatrix} x_{11} & x_{21} & \dots & x_{N1} \\ x_{12} & x_{22} & \dots & x_{N2} \end{bmatrix} = \begin{bmatrix} w_0 x_{11} + w_1 x_{12} & w_0 x_{21} + w_1 x_{22} & \dots & w_0 x_{n1} + w_1 x_{n2} \end{bmatrix}$$

   Therefore, $(\mathbf{X}\mathbf{w})^\top = \mathbf{w}^\top \mathbf{X}^\top$.

4. **Exercise 1.5** from FCMA p.35 [1pt]

   When multiplying a scalar by a vector (or matrix), we multiply each element of the vector (or matrix) by that scalar. For $\mathbf{x}_n = [x_{n1}, x_{n2}]^\top$, $\mathbf{t} = [t_1, ..., t_N]^\top$, $\mathbf{w} = [w_0, w_1]^\top$, and

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1{}^\top \\ \mathbf{x}_2{}^\top \\ \vdots \\ \mathbf{x}_N{}^\top \end{bmatrix}$$

   show that

$$\sum_n \mathbf{x}_n t_n = \mathbf{X}^\top \mathbf{t}$$

   and

$$\sum_n \mathbf{x}_n \mathbf{x}_n^\top \mathbf{w} = \mathbf{X}^\top \mathbf{X}\mathbf{w}$$

   Solution.

We will show that $\sum_n \mathbf{x}_n t_n = \mathbf{X}^\top \mathbf{t}$

Developing the left side:

$$\sum_n \mathbf{x}_n t_n = \mathbf{x}_1 t_1 + \mathbf{x}_2 t_2 + \ldots + \mathbf{x}_n t_n = \begin{bmatrix} x_{11}t_1 \\ x_{12}t_1 \end{bmatrix} + \begin{bmatrix} x_{21}t_2 \\ x_{22}t_2 \end{bmatrix} + \ldots + \begin{bmatrix} x_{n1}t_n \\ x_{n2}t_n \end{bmatrix} = \begin{bmatrix} x_{11}t_1 + x_{11}t_2 + \ldots + x_{n1}t_n \\ x_{12}t_1 + x_{22}t_2 + \ldots + x_{n2}t_n \end{bmatrix} =$$

$$= \begin{bmatrix} \sum_n x_{n1}t_n \\ \sum_n x_{n2}t_n \end{bmatrix}$$

Developing the right side:

$$\mathbf{X}^\top \mathbf{t} = \begin{bmatrix} x_{11} & x_{21} & \ldots & x_{n1} \\ x_{12} & x_{22} & \ldots & x_{n2} \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix} == \begin{bmatrix} x_{11}t_1 + x_{11}t_2 + \ldots + x_{n1}t_n \\ x_{12}t_1 + x_{22}t_2 + \ldots + x_{n2}t_n \end{bmatrix} = \begin{bmatrix} \sum_n x_{n1}t_n \\ \sum_n x_{n2}t_n \end{bmatrix}$$

Therefore, $\sum_n \mathbf{x}_n t_n = \mathbf{X}^\top \mathbf{t}$.

Now, we will show that $\sum_n \mathbf{x}_n \mathbf{x}_n^\top \mathbf{w} = \mathbf{X}^\top \mathbf{X} \mathbf{w}$

Starting by the left side, we will have:

$$\sum_n \mathbf{x}_n \mathbf{x}_n^\top \mathbf{w} = \sum_n \left(\begin{bmatrix} x_{n1} \\ x_{n2} \end{bmatrix} \begin{bmatrix} x_{n1} & x_{n2} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} \right) = \sum_n \begin{bmatrix} x_{n1}^2 & x_{n1}x_{n2} \\ x_{n1}x_{n2} & x_{n2}^2 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} = \sum_n \begin{bmatrix} w_0 x_{n1}^2 + w_1 x_{n1}x_{n2} \\ w_0 x_{n1}x_{n2} + w_1 x_{n2}^2 \end{bmatrix}$$

On the other hand, the right side will be:

$\mathbf{X}^\top \mathbf{X} \mathbf{w}$

We have seen in the Exercise **1.3** that: $\mathbf{X}^\top \mathbf{X} = \begin{bmatrix} \sum_{n=1}^N x_{n1}^2 & \sum_{n=1}^N x_{n1}x_{n2} \\ \sum_{n=1}^N x_{n1}x_{n2} & \sum_{n=1}^N x_{n2}^2 \end{bmatrix}$

So, $\mathbf{X}^\top \mathbf{X} \mathbf{w} = \begin{bmatrix} \sum_{n=1}^N x_{n1}^2 & \sum_{n=1}^N x_{n1}x_{n2} \\ \sum_{n=1}^N x_{n1}x_{n2} & \sum_{n=1}^N x_{n2}^2 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} = \begin{bmatrix} w_0 \sum_{n=1}^N x_{n1}^2 + w_1 \sum_{n=1}^N x_{n1}x_{n2} \\ w_0 \sum_{n=1}^N x_{n1}x_{n2} + w_1 \sum_{n=1}^N x_{n2}^2 \end{bmatrix} =$

$$= \begin{bmatrix} \sum_{n=1}^N w_0 x_{n1}^2 + \sum_{n=1}^N w_1 x_{n1}x_{n2} \\ \sum_{n=1}^N w_0 x_{n1}x_{n2} + \sum_{n=1}^N w_1 x_{n2}^2 \end{bmatrix} = \sum_n \begin{bmatrix} w_0 x_{n1}^2 + w_1 x_{n1}x_{n2} \\ w_0 x_{n1}x_{n2} + w_1 x_{n2}^2 \end{bmatrix}$$

Therefore, $\sum_n \mathbf{x}_n \mathbf{x}_n^\top \mathbf{w} = \mathbf{X}^\top \mathbf{X} \mathbf{w}$

5. **Plotting Exercise** – three parts

A. [0.5pt] Use the provided python script `plotlinear.py` and plot three lines (parameters of your choosing). This requires you have you python environment set up. Place the output graphic in your pdf submission and provide a descriptive caption that indicates the intercept and slope values you used to generate the lines. (LaTeX users can use the commented code in the HW latex template for inserting the figure.)

Solution.

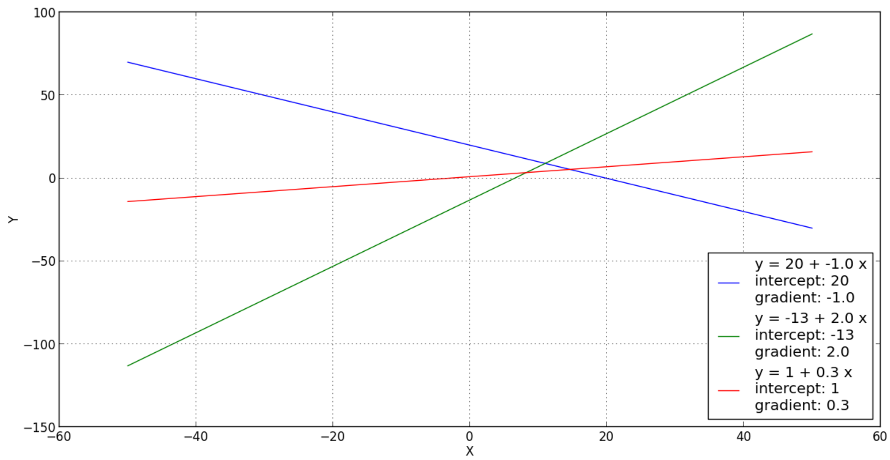

Figure 2: Solution

B. [0.5pt] Now, try another plot: generate a vector whose entries are the values of $\sin(\mathbf{x})$ for $\mathbf{x}$ in the range $[0, 10]$ in steps of 0.01, and plot it. Label the y-axis 'sin(x)', the x-axis 'x values' and provide a title for the plot, 'Sine Function for x from 0.0 to 10.0'. Include your plot in the pdf submission.

Solution.

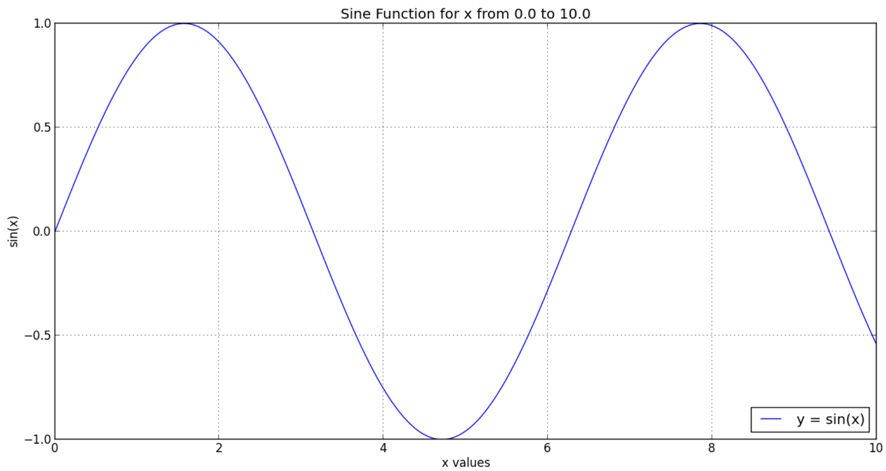

Figure 3: $\sin(\mathbf{x})$ for $\mathbf{x}$ in range $[0, 10]$ in steps of 0.01

C. [0.5pt] Finally, put your script code in the pdf file.

For LaTeX users: you can use the following python code listing environment. The code currently listed here is from the `plotlinear.py` script; replace the code there with the code you wrote for generating the sin function.

Solution.

Code Listing 1: `plot-sin.py` script

```
## plot-sin.py
import numpy as np
import matplotlib.pyplot as plt

plt.ion()

#x-axis: from 0.0 to 10.0 (steps: 0.01)
x = np.arange(0, 10, 0.01);

#plt.figure(0)

#Define x and y labels as well as title
plt.xlabel("x values")
plt.ylabel("sin(x)")
plt.title("Sine Function for x from 0.0 to 10.0 ")
#our grid (just for a better reading)
plt.grid()

#Ploting our function and its label
plt.plot(x, np.sin(x), label="y = sin(x)")
#loc = 4 bottom right
plt.legend(loc=4)

#holds our graph on screen
raw_input("\nPress <ENTER> to exit...");
```

6. **More Programming in Python Practice**

A. [1pt] Write a short script that initializes the random number generator

python: `numpy.random.seed(seed=1)`

Followed by creating two three-dimensional column vectors using

python: `random.rand` (used in the context of code to generate the vectors)

Represent the random variables as **a** and **b** (be sure you issue the call to set the random seed immediately before creating these variables). Print them at the terminal and copy-and-paste the result here. (If using LaTeX, use the `verbatim` environment to display).

Solution.

```
[emanuel@localhost code]$ python hw1-6a.py
a = [  4.17022005e-01   7.20324493e-01   1.14374817e-04]
b = [ 0.30233257  0.14675589  0.09233859]
```

B. [2pts] Using the values of **a** and **b**, compute the following and display the result two ways: (1) copy-and-paste the output (from the python interpreter/terminal; again, in LaTeX use the `verbatim` environment), (2) typeset the output (e.g., using the LaTeX math environment).

1. $\mathbf{a} + \mathbf{b} =\ ?$

Solution:

```
[emanuel@localhost code]$ python hw1-6b1.py
a = [  4.17022005e-01   7.20324493e-01   1.14374817e-04]
b = [ 0.30233257  0.14675589  0.09233859]
a + b = [ 0.71935458  0.86708038  0.09245297]
```

Soluton in LaTeX:

$$\mathbf{a} + \mathbf{b} = \begin{bmatrix} 4.17022005e * 10^{-1} \\ 7.20324493 * 10^{-1} \\ 1.14374817 * 10^{-4} \end{bmatrix} + \begin{bmatrix} 0.30233257 \\ 0.14675589 \\ 0.09233859 \end{bmatrix} = \begin{bmatrix} 0.71935458 \\ 0.86708038 \\ 0.09245297 \end{bmatrix}$$

2. $\mathbf{a} \circ \mathbf{b} =\ ?$ (element-wise multiply; Note: the notation $\mathbf{a} \circ \mathbf{b}$ is also known as the Hadamard product, the entrywise product, or the Schur product.)

Solution:

```
[emanuel@localhost code]$ python hw1-6b2.py
a = [  4.17022005e-01   7.20324493e-01   1.14374817e-04]
b = [ 0.30233257  0.14675589  0.09233859]
a * b = [  1.26079336e-01   1.05711863e-01   1.05612099e-05]
```

Solution in LaTeX:

$$\mathbf{a} \circ \mathbf{b} = \begin{bmatrix} 4.17022005e * 10^{-1} \\ 7.20324493 * 10^{-1} \\ 1.14374817 * 10^{-4} \end{bmatrix} \circ \begin{bmatrix} 0.30233257 \\ 0.14675589 \\ 0.09233859 \end{bmatrix} = \begin{bmatrix} 1.26079336 * 10^{-1} \\ 1.05711863 * 10^{-1} \\ 1.05612099 * 10^{-5} \end{bmatrix}$$

3. $\mathbf{a}^\top \mathbf{b} =\ ?$ (also called the dot-product)

Solution.

```
[emanuel@localhost code]$ python hw1-6b3.py
a = [  4.17022005e-01   7.20324493e-01   1.14374817e-04]
b = [ 0.30233257  0.14675589  0.09233859]
aT.b = 0.231801759448
```

Solution in LaTeX:

$$\mathbf{a}^\top \mathbf{b} = \begin{bmatrix} 4.17022005e * 10^{-1} \\ 7.20324493 * 10^{-1} \\ 1.14374817 * 10^{-4} \end{bmatrix}^\top \begin{bmatrix} 0.30233257 \\ 0.14675589 \\ 0.09233859 \end{bmatrix} =$$

$$= \begin{bmatrix} 4.17022005e * 10^{-1} & 7.20324493 * 10^{-1} & 1.14374817 * 10^{-4} \end{bmatrix} \begin{bmatrix} 0.30233257 \\ 0.14675589 \\ 0.09233859 \end{bmatrix} = 0.231801759448$$

Now, set the random seed to 2 and immediately generate a random 3×3 matrix $\mathbf{X}$. In your solution, display the value of $\mathbf{X}$. Using $\mathbf{X}$ and the earlier values of a and b, compute the following in python and typeset the results in two ways, as before.

```
[emanuel@localhost code]$ python hw1-6b4.py
X =
[[ 0.4359949   0.02592623  0.54966248]
 [ 0.43532239  0.4203678   0.33033482]
 [ 0.20464863  0.61927097  0.29965467]]
```

Solution in LaTeX:

$$\mathbf{X} = \begin{bmatrix} 0.4359949 & 0.02592623 & 0.54966248 \\ 0.43532239 & 0.4203678 & 0.33033482 \\ 0.20464863 & 0.61927097 & 0.29965467 \end{bmatrix}$$

4. $\mathbf{a}^\top \mathbf{X} =\ ?$

   Solution.

```
[emanuel@localhost code]$ python hw1-6b4.py
aT.X = [ 0.49541626  0.31368386  0.46720388]
```

Solution in LaTeX:

$$\mathbf{a}^\top \mathbf{X} = \begin{bmatrix} 4.17022005e * 10^{-1} & 7.20324493 * 10^{-1} & 1.14374817 * 10^{-4} \end{bmatrix} \begin{bmatrix} 0.4359949 & 0.02592623 & 0.54966248 \\ 0.43532239 & 0.4203678 & 0.33033482 \\ 0.20464863 & 0.61927097 & 0.29965467 \end{bmatrix} =$$

$$= \begin{bmatrix} 0.49541626 & 0.31368386 & 0.46720388 \end{bmatrix}$$

5. $\mathbf{a}^\top \mathbf{X} \mathbf{b} =\ ?$

   Solution.

```
[emanuel@localhost code]$ python hw1-6b5.py
aT.X.b =
0.238956376181
```

Solution in LaTeX:

$$\mathbf{a}^\top\mathbf{X}\mathbf{b} = (\mathbf{a}^\top\mathbf{X})\mathbf{b} = \begin{bmatrix}0.49541626 & 0.31368386 & 0.46720388\end{bmatrix}\begin{bmatrix}0.30233257\\0.14675589\\0.09233859\end{bmatrix} = 0.238956376181$$

6. $\mathbf{X}^{-1} =$?

   Solution.

```
[emanuel@localhost code]$ python hw1-6b6.py
X^-1 =
[[-1.20936675  5.11771977 -3.42333228]
 [-0.96691719  0.279414    1.46561347]
 [ 2.82418088 -4.07257903  2.64627411]]
```

   Solution in LaTeX:

$$\mathbf{X}^{-1} = \begin{bmatrix}0.4359949 & 0.02592623 & 0.54966248\\0.43532239 & 0.4203678 & 0.33033482\\0.20464863 & 0.61927097 & 0.29965467\end{bmatrix}^{-1} = \begin{bmatrix}-1.20936675 & 5.11771977 & -3.42333228\\-0.96691719 & 0.279414 & 1.46561347\\2.82418088 & -4.07257903 & 2.64627411\end{bmatrix}$$