Math Basics for Machine Learning
Graded Assignment 3

Your Name Here

Fall 2023

Instructions

This is the second graded assignment for the Math Basics for Machine Learning course. It contains two tasks. The instructions, as well as links to supplementary material, are given in the task descriptions.

Provide **detailed solutions** to the tasks in this assignment. Then, save your solution document as a .pdf file and submit it by filling in the corresponding Google form.

For some tasks, you might use Python rather than perform computations by hand. If you do so, attach your code as well (e.g., link to the Colab notebook with code).

In total, you can earn 10 points for this assignment. This score will contribute to you final score for this course.

You must submit your answers by **Monday, October 23, 18:59 Moscow Time**. *Late submissions will* ***not*** *be accepted.*

It is the idea that you complete this assignment individually. Do not collaborate or copy answers of somebody else.

Have fun!

1. (5 points) ML researchers work either at universities ("academia") or in private companies ("industry"). Official statistics shows that every year, 20% of academic workers transition to industry, while only 1% moves from industry to academia. How will this distribution evolve over years? In this task, you will answer this question using a simple model.

   We can collect the information we have into a single matrix T, where the first column and the first row correspond to industry, while the second column and the second row are related to academia:

   $$T = \begin{bmatrix} 0.99 & 0.2 \\ 0.01 & 0.8 \end{bmatrix}$$

   You can read this matrix as follows:

   - *first column:* 0.99 of ML researchers in industry stay there, while 0.01 transition to academia;
   - *second column:* 0.2 of the academic researchers transition to industry, while the remaining 0.8 stays in academia.

   Suppose that in some year 0, there were $x_0^{industry}$ researchers working in industry and $x_0^{academia}$ academic researchers. Given the statistics above, we can then estimate industry vs. academia research distribution next year as follows:

   $$x_1 = Tx_0, \;\; x_0 = \left[x_0^{industry}, x_0^{academia}\right]^T$$

   (a) (2 points) Imagine that in 2023, 30% of the ML researchers work in industry, while 70% work in academia. How will this distribution be changing over the next 100 years (assuming that the transition matrix stays the same)?

   Make a plot to visualize industry vs. academia researchers distribution for each year from 2023 up to 2123 according to our model. Explain what you see.

   Solution: Your solution here

   (b) (1 point) In some alternative universe, in 2023, the statistics look different: 90% of the researchers work in academia and only 10% are in industry. What will the distribution between industry and academia research be by 2123 in that case? How does it compare to the previous result?

   Solution: Your solution here

   (c) (2 points) Compute eigenvalues and eigenvectors of the transition matrix T and note that one of the eigenvalues is equal to 1. How is this (and the corresponding eigenvector) related to the evolution of the industry vs. academia research distribution that you have observed in the previous task? Explain.

   Solution: Your solution here

2. (5 points) This is a programming task dedicated to SVD. You can find the assignment in the Google Colab notebook.

First, make your own copy of the notebook (*File* → *Save a copy in Drive*) or download the notebook to your machine if you prefer to work locally (*File* → *Download*).

Then, implement your solutions to the tasks formulated in the notebook. You can add **code cells** to write some code and **text cells** in case you want to include additional explanations to your answers in plain English.

Finally, attach the link to you notebook to the submission form. **Make sure that all the cells are executed and all relevant outputs are being printed out**.