Math Basics for Machine Learning
Final Exam

Your Name Here

Fall 2024

Instructions

This is the final exam for the Math Basics for Data Science course.

This exam is in two parts:

- PART 1 consists of a quiz similar to the short quizes from our classes. You will need to answer a number of questions, but you do **not** need to provide detailed solutions.
- PART 2 consists of six free-response questions for which you will need to attach detailed solutions, similar to the graded assignments.

You can get 30 points for each of the parts, resulting in 60 points in total for the exam.

You can submit the exam by filling in the corresponding Google form. **You can submit your answers only once.** After you have submitted the form, you should receive a confirmation email. If you have submitted your solutions but did not receive any confirmation, contact me.

You must submit your answers by Friday January 24, 18:59 Moscow time. Solutions must be typed in LaTeX. ***Hand-written solutions, as well as late submissions, will not be accepted.***

The idea is to complete the exam individually. Do not collaborate, share your solutions with anyone, or copy answers of somebody else.

Good luck!

Part 1 (30 points)

The first part of the exam is a quiz. You will need to answer a number of questions, but you do **not** need to provide detailed solutions.

The questions for this part can be found in the submission Google form.

Part 2 (30 points)

The second part of the exam contains six free response questions for which you will need to attach detailed solutions. You can find the problem formulations below.

Always show how you obtained the solution and provide reasonably detailed explanations. The answers stated without any comments will not be accepted. For some tasks, it might be convenient to use Python rather than performing the computations by hand. If you do, then you can also attach your code.

1. (3 points) Consider the following linearly independent vectors:

$$u = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}, \ v = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \ w = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}$$

   Apply the Gram-Schmidt process to obtain an *orthonormal* basis for $\mathbb{R}^3$.

   Solution: Your solution here

2. (4 points) Consider the following three vectors:

$$u = \begin{bmatrix} -1 \\ -4 \\ 3 \end{bmatrix}, \ v = \begin{bmatrix} 10 \\ 7 \\ h+5 \end{bmatrix}, \ w = \begin{bmatrix} 4 \\ 5 \\ 0 \end{bmatrix}$$

   where h is some real number.

   (a) (2 points) Find all possible values of h for which $\{u, v, w\}$ form a basis for $\mathbb{R}^3$.

   Solution: Your solution here

   (b) (2 points) Pick **one** value of h for which $\{u, v, w\}$ form a basis for $\mathbb{R}^3$ and find the coordinates of the vector $x = (1, 2, 3)$ in that basis.

   Solution: Your solution here

3. (3 points) Consider the following matrix:

$$A = \begin{bmatrix} -6 & 5 & -6 \\ 4 & -1 & 2 \\ -6 & 3 & -4 \end{bmatrix}$$

   Can it be diagonalized? If so, determine its diagonal form and a basis in which A is diagonal. If not, explain why.

Solution: Your solution here

4. (8 points) Let $a_1, ..., a_n$ be some real numbers. Consider the following function:

$$f(x) = \prod_{i=1}^{n} [x^{a_i} e^{-x}], \ x > 0$$

Find the value of $x > 0$ that maximizes $f(x)$.

Solution: Your solution here

5. (4 points) Find and classify all the critical points of the following function:

$$f(x, y) = x^3 - 12x^2 + 8x - x\sqrt{y-1} + 0.5y$$

Note: if the second derivative test is indecisive, you don't need to investigate the correspondent point any further.

Solution: Your solution here

6. (8 points) Fitting a machine learning model means finding the optimal values of its parameters, which comes down to optimizing some loss function $\mathcal{L}$. Consider the following loss function:

$$\mathcal{L} = \sum_{i=1}^{n} \left[y_i \log \sigma_i + (1 - y_i) \log (1 - \sigma_i) \right],$$

where $\sigma_i = \sigma_i(w_0, w_1) = \frac{1}{1+e^{-(w_0+w_1 x_i)}}$.

Here, $(x_i, y_i), \ i = 1, ..., n$ are the observed data points, and w_0 and w_1 are the parameters of the model.

Find the gradient of the loss function above.

Solution: Your solution here