Math Basics for Machine Learning
Final Exam

Your Name Here

Fall 2023

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 just 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 Monday, November 27, 18:59 Moscow time. ***Late submissions will not be accepted***.

Solutions must be typed in LaTeX. Hand-written solutions, as well as late submissions, will not be accepted.

It is the idea that you 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 problem formulations below.

Always show how you obtain the solution and provide reasonably detailed explanations. Answers stated without any comments won't be accepted. For some tasks, it might be convenient to use Python rather than perform the computations by hand. If you do so, attach your code as well.

1. (4 points) Consider the following three vectors:

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

   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

2. (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

3. (3 points) Consider the following matrix:

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

   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:

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

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) Suppose you want to fit a linear regression model of the form

$$y = w_0 + w_1 \cdot x.$$

You want to do so via **regularized least squares**, i.e., by minimizing the following loss function:

$$\mathcal{L}(x, y, w) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \cdot ||w||^2.$$

Here, $\hat{y}_i$ is model's prediction for example i and $||w||^2$ is the l_2 norm of the unknown weights vector $w = (w_0, w_1)$. The effect of adding this extra term to the loss function is that it forces us to chose small values for the unknown coefficients. The larger the value of the hyperparameter λ, the larger is the effect of regularization.

(a) (4 points) You decide to set $\lambda = 0.01$. Compute the gradient of the loss $\mathcal{L}$.

Solution: Your solution here

(b) (4 points) Suppose that you have observed two examples: $x_1 = 1, \; y_1 = 2$ and $x_2 = 2, \; y_2 = -1$. Your initial guess for the unknown weights is $w^0 = (0, 1)$.

Perform one step of the gradient descent algorithm to update parameters' values. Chose learning rate $\eta = 0.1$.

Solution: Your solution here