Math Basics for Machine Learning
Graded Assignment 4

Your Name Here

Fall 2024

Instructions

This is the fourth 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.

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 21, 23:59 Moscow Time**.

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 this assignment individually. Do not collaborate or copy answers of somebody else.

Have fun!

1. (4 points) Find and classify all the critical points of the following function

$$f(x, y) = 7x - 8y + 2xy - x^2 + y^3$$

Solution: Your solution here

2. (6 points) Fitting a machine learning model means finding the optimal values of its parameters, which comes down to optimizing some loss function $\mathcal{L}$. Suppose you want to fit a linear regression model of the form

$$y = w_0 + w_1 \cdot x_1 + ... + w_n \cdot x_n$$

You want to do so by minimizing the following loss function:

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

Here, $\hat{y}_i$ is model's prediction for example i and $||w||^2$ is the l_2 norm of the unknown weights vector $w = (w_0, w_1, ..., w_n)^T$. 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) Find the gradient of $\mathcal{L}$. *Hint: it might be useful to re-write $\mathcal{L}$ using matrix notation and use matrix calculus.*

Solution: Your solution here

(b) (2 points) Using the gradient obtained above, derive the expression for the optimal model weights w that minimize the loss $\mathcal{L}$.

Solution: Your solution here