Math Basics for Machine Learning
Graded Assignment 4

Vladimir Saraikin

Fall 2023

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, November 6, 18:59 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:

1. Critical Points:

$$\frac{\partial f}{\partial x} = 7 - 2x + 2y = 0,$$
$$\frac{\partial f}{\partial y} = -8 + 2x + 3y^2 = 0.$$

$$(x, y) = \left(\frac{5}{2}, -1\right),$$
$$(x, y) = \left(\frac{23}{6}, \frac{1}{3}\right).$$

2. To classify these critical points, use the second derivative test.

The determinant D is given by:

$$D = (-2)(6y) - (2)(2) = -12y - 4,$$

Evaluating D at each critical point:

- For $\left(\frac{5}{2}, -1\right)$, $D = 8$. Since $D > 0$ and $\frac{\partial^2 f}{\partial x^2} < 0 \Rightarrow$ this point is a local maximum.
- For $\left(\frac{23}{6}, \frac{1}{3}\right)$, $D = -8$. Since $D < 0 \Rightarrow$ this point is a saddle point.

Answer: $f(x, y)$ has one local maximum at $\left(\frac{5}{2}, -1\right)$ and one saddle point at $\left(\frac{23}{6}, \frac{1}{3}\right)$.

2. (6 points) Fitting a machine learning model means finding the optimal values of its parameters, which comes down to optimizing some loss function $\mathcal{L}$. In class, we saw the least-squares example. Now, let's consider a so-called *logistic loss*:

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

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

Here, $\{x_i, y_i\}_{i=1,\ldots,n}$ are the observed data points, and w_0 and w_1 are the parameters of the model.

(a) (4 points) Find the gradient of the loss function above.

Solution:

$$\nabla \mathcal{L} = \left(\frac{\partial \mathcal{L}}{\partial w_0}, \frac{\partial \mathcal{L}}{\partial w_1}\right).$$

1. $\frac{\partial \mathcal{L}}{\partial w_0}$:

$$\frac{\partial \mathcal{L}}{\partial w_0} = \sum_{i=1}^{n} \left[\frac{y_i}{\sigma_i} \frac{\partial \sigma_i}{\partial w_0} - \frac{1-y_i}{1-\sigma_i} \frac{\partial \sigma_i}{\partial w_0} \right],$$

2. $\frac{\partial \mathcal{L}}{\partial w_1}$:

$$\frac{\partial \mathcal{L}}{\partial w_1} = \sum_{i=1}^{n} \left[\frac{y_i}{\sigma_i} \frac{\partial \sigma_i}{\partial w_1} - \frac{1-y_i}{1-\sigma_i} \frac{\partial \sigma_i}{\partial w_1} \right].$$

The derivatives of σ_i with respect to w_0 and w_1 are:

$$\frac{\partial \sigma_i}{\partial w_0} = \sigma_i(1-\sigma_i),$$

$$\frac{\partial \sigma_i}{\partial w_1} = x_i \sigma_i(1-\sigma_i).$$

Substituting these into the partial derivatives of $\mathcal{L}$:

$$\frac{\partial \mathcal{L}}{\partial w_0} = \sum_{i=1}^{n} (y_i - \sigma_i),$$

$$\frac{\partial \mathcal{L}}{\partial w_1} = \sum_{i=1}^{n} x_i(y_i - \sigma_i).$$

Answer: $\nabla \mathcal{L} = \left(\sum_{i=1}^{n} (y_i - \sigma_i), \sum_{i=1}^{n} x_i(y_i - \sigma_i) \right).$

(b) (2 points) Suppose that you have a single observation:

$$x_1 = 2, \; y_1 = 1.$$

Let's assume the initial weights w_0 and w_1 are both set to 0. Perform one step of a gradient descent update of the parameter values. Use learning rate $\eta = 0.1$

Solution:

1. σ_1 with the initial weights:

$$\sigma_1 = \frac{1}{1 + \exp - (w_0 + w_1 x_1)} = \frac{1}{1 + \exp - (0 + 0 \cdot 2)} = \frac{1}{1+1} = \frac{1}{2}.$$

2. The gradient of the loss function with respect to w_0 and w_1:

$$\frac{\partial \mathcal{L}}{\partial w_0} = y_1 - \sigma_1 = 1 - \frac{1}{2} = \frac{1}{2},$$

$$\frac{\partial \mathcal{L}}{\partial w_1} = x_1(y_1 - \sigma_1) = 2\left(1 - \frac{1}{2}\right) = 2 \cdot \frac{1}{2} = 1.$$

3. The updates are:

$$w_0^{new} = w_0 - \eta \frac{\partial \mathcal{L}}{\partial w_0} = 0 - 0.1 \cdot \frac{1}{2} = 0 - 0.05 = -0.05,$$

$$w_1^{new} = w_1 - \eta \frac{\partial \mathcal{L}}{\partial w_1} = 0 - 0.1 \cdot 1 = 0 - 0.1 = -0.1.$$

Answer: $w_0 = -0.05, \quad w_1 = -0.1.$