

# Math Basics for Machine Learning

## Final Exam

Vladimir Saraikin

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 quizzes 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:

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

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:**

To determine when these vectors form a basis, check for their linear independence by ensuring the determinant of the matrix formed by these vectors is non-zero. The matrix  $A$  is:

$$A = \begin{bmatrix} 4 & -1 & 10 \\ 5 & -4 & 7 \\ 0 & 3 & h+5 \end{bmatrix}.$$

The determinant of  $A$  is:

$$\begin{aligned} \det(A) &= 4((-4)(h+5) - 21) + (-1)(5(h+5)) + 150 \\ &= -16h - 164 + 5h + 25 + 150 \\ &= -11h + 11. \end{aligned}$$

For linear independence,  $\det(A) \neq 0$ , which gives:

$$\begin{aligned} -11h + 11 &\neq 0 \\ h &\neq 1. \end{aligned}$$

**Answer:**  $\{u, v, w\}$  form a basis for  $\mathbb{R}^3$  for all  $h$  except  $h = 1$ .

- (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:**

Choose  $h = 0$ . The vectors become:

$$u = \begin{bmatrix} 4 \\ 5 \\ 0 \end{bmatrix},$$

$$v = \begin{bmatrix} -1 \\ -4 \\ 3 \end{bmatrix},$$

$$w = \begin{bmatrix} 10 \\ 7 \\ 5 \end{bmatrix}.$$

To find the coordinates of  $x = (1, 2, 3)$  in this basis, solve the equation:

$$a \cdot u + b \cdot v + c \cdot w = x.$$

In matrix form, this is:

$$\begin{bmatrix} 4 & -1 & 10 \\ 5 & -4 & 7 \\ 0 & 3 & 5 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

$$a = \frac{128}{11},$$

$$b = \frac{81}{11},$$

$$c = -\frac{42}{11}.$$

**Answer:**, the coordinates of the vector  $x = (1, 2, 3)$  in the basis formed by  $u, v$ , and  $w$  (with  $h = 0$ ) are  $\left(\frac{128}{11}, \frac{81}{11}, -\frac{42}{11}\right)$ .

2. (3 points) Consider the following linearly independent vectors:

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

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

**Solution:**

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:**

Eigenvalues:  $\lambda_1 = 2, \lambda_2 = 2, \lambda_3 = 1$

Eigenvectors:  $v_1 = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}, v_2 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, v_3 = \begin{pmatrix} 3 \\ -1 \\ 3 \end{pmatrix}$

To check if  $A$  can be diagonalized, we verify the linear independence of its eigenvectors by calculating the determinant of the matrix  $V$  formed by these eigenvectors:

$$V = \begin{pmatrix} 2 & 2 & 3 \\ 0 & 1 & -1 \\ 1 & 0 & 3 \end{pmatrix}$$

$$\det(V) = \det \left( \begin{pmatrix} 2 & 2 & 3 \\ 0 & 1 & -1 \\ 1 & 0 & 3 \end{pmatrix} \right) = 1$$

Since  $\det(V) \neq 0$ , the eigenvectors are linearly independent, and thus  $A$  can be diagonalized.

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

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

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

**Solution:** Given the function

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

simplify it as

$$f(x) = x^{\sum_{i=1}^n a_i} e^{-nx}.$$

To find the maximum, differentiate  $f(x)$  with respect to  $x$  and set the derivative to zero:

$$f'(x) = \frac{d}{dx} \left( x^{\sum_{i=1}^n a_i} e^{-nx} \right) = -\frac{nx^{\sum_{i=1}^n a_i}}{e^{nx}} + \frac{x^{\sum_{i=1}^n a_i - 1} \sum_{i=1}^n a_i}{e^{nx}}.$$

Setting  $f'(x) = 0$  and solving for  $x$  yields

$$x = \frac{\sum_{i=1}^n a_i}{n}.$$

Thus, the value of  $x > 0$  that maximizes  $f(x)$  is  $\frac{\sum_{i=1}^n a_i}{n}$ , the average of the  $a_i$  values.

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:**

**Critical Points:** The critical points are found by solving  $\nabla f = 0$ . The partial derivatives are:

$$\begin{aligned}f_x &= 0.5 - \frac{y}{2\sqrt{x-1}}, \\f_y &= 8 - \sqrt{x-1} - 24y + 3y^2.\end{aligned}$$

Solving  $f_x = 0$  and  $f_y = 0$ , find the critical points:

$$\begin{aligned}(x, y) &\approx (1.11, 0.33), \\(x, y) &\approx (65, 8).\end{aligned}$$

**Classification:** The Hessian matrix  $H$  is given by:

$$H = \begin{pmatrix} 0 + \frac{y}{4(x-1)^{3/2}} & -\frac{1}{2\sqrt{x-1}} \\ -\frac{1}{2\sqrt{x-1}} & -24 + 6y \end{pmatrix}.$$

At  $(1.11, 0.33)$ :

$$\begin{aligned}\det(H) &= -51.75, \\f_{xx} &= 2.25.\end{aligned}$$

Since  $\det(H) < 0$ , this point is a saddle point.

At  $(65, 8)$ :

$$\begin{aligned}\det(H) &= \frac{23}{256}, \\f_{xx} &= \frac{1}{256}.\end{aligned}$$

Since  $\det(H) > 0$  and  $f_{xx} > 0$ , this point is a local minimum.

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  $\lambda$ , 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:**

Given the loss function

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

where  $\hat{y}_i = w_0 + w_1 \cdot x_i$  and  $\|w\|^2 = w_0^2 + w_1^2$ . For  $\lambda = 0.01$ , the gradient of  $\mathcal{L}$  with respect to  $w$  is:

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

The partial derivatives are computed as follows:

1. Partial derivative with respect to  $w_0$ :

$$\frac{\partial \mathcal{L}}{\partial w_0} = -2 \sum_{i=1}^n (y_i - w_0 - w_1 x_i) + 2\lambda w_0.$$

2. Partial derivative with respect to  $w_1$ :

$$\frac{\partial \mathcal{L}}{\partial w_1} = -2 \sum_{i=1}^n x_i (y_i - w_0 - w_1 x_i) + 2\lambda w_1.$$

Therefore, the gradient of the loss function  $\mathcal{L}$  is:

$$\nabla_w \mathcal{L} = \left( -2 \sum_{i=1}^n (y_i - w_0 - w_1 x_i) + 2\lambda w_0, -2 \sum_{i=1}^n x_i (y_i - w_0 - w_1 x_i) + 2\lambda w_1 \right).$$

- (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:** The gradient of the loss function  $\mathcal{L}$  at  $w^0$  is:

$$\nabla_w \mathcal{L} = \left( -2 \sum_{i=1}^n (y_i - w_0 - w_1 x_i) + 2\lambda w_0, -2 \sum_{i=1}^n x_i (y_i - w_0 - w_1 x_i) + 2\lambda w_1 \right)$$

with  $\lambda = 0.01$ ,  $n = 2$ ,  $x = \{1, 2\}$ , and  $y = \{2, -1\}$ .

Substituting the values, the gradient at  $w^0$  is computed, and the weights are updated as:

$$w^{new} = w^{old} - \eta \cdot \nabla_w \mathcal{L}$$

After calculation, the updated weights are:

$$w_0^{new} = -0.4, \quad w_1^{new} = -0.002$$