University of Zurich[UZH]

Prof. Dan Olteanu,
Dr. Ahmet Kara, Dr. Nils Vortmeier,
Haozhe Zhang

DaST
Data • (Systems+Theory)

Fall 2020/2021 | Sheet 1 | 15.09.2020

The solutions will be discussed on Friday 02.10.2020, 14:00-15:45 on Zoom:

- https://uzh.zoom.us/j/96690150974?pwd=cnZmMTduWUtCeWoxYW85Z3RMYnpTZz09
- Meeting ID: 966 9015 0974, Passcode: 094609
- You need to log in to UZH zoom to access the course room

Videos with solutions will be posted in OLAT after the exercise session.

Exercise 1.1 [Vectors and Matrices]
Consider the following matrix $\mathbf{X}$ and vectors $\mathbf{y}$ and $\mathbf{z}$:

$$\mathbf{X} = \begin{pmatrix} 2 & -3 & 1 \\ 3 & 4 & 2 \\ -2 & 0 & -1 \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix} \quad \mathbf{z} = \begin{pmatrix} 3 \\ -2 \\ 4 \end{pmatrix}.$$

(a) The dot (or inner) product of $\mathbf{y}$ and $\mathbf{z}$ is denoted as $\mathbf{y} \cdot \mathbf{z}$ (or sometimes as $\langle \mathbf{y}, \mathbf{z} \rangle$, $\mathbf{y}^\top \mathbf{z}$, or $\mathbf{z}^\top \mathbf{y}$). Give the result of $\mathbf{y} \cdot \mathbf{z}$.

(b) Give the results of $\mathbf{X}\mathbf{y}$ and $\mathbf{z}^\top \mathbf{X}$.

(c) Give the determinant of $\mathbf{X}$.

(d) Does the inverse of $\mathbf{X}$ exist? What is the connection between this question and the determinant of $\mathbf{X}$? Give the inverse of $\mathbf{X}$ if it exists.

(e) What is the rank of $\mathbf{X}$? Give examples of changes to $\mathbf{X}$ that change the rank of $\mathbf{X}$.

(f) Is the matrix $\mathbf{Y} = \begin{pmatrix} 2 & 2 & 1 \\ 2 & 3 & 0 \\ 1 & 0 & 2 \end{pmatrix}$ positive (semi-)definite?

Exercise 1.2 [Eigenvalues, Eigenvectors, and Eigenbasis]
Consider the following matrix $\mathbf{X}$ and vector $\mathbf{y}$:

$$\mathbf{X} = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}.$$

(a) Give the eigenvalues and eigenvectors of $\mathbf{X}$.

(b) Give an eigenbasis for $\mathbf{X}$.

(c) We would like to compute the product of the 100th power of $\mathbf{X}$ with $\mathbf{y}$, i.e., $\mathbf{X}^{100}\mathbf{y}$. A quite tedious method to obtain the result would be to compute $\mathbf{X}(\mathbf{X}^i\mathbf{y})$ step by step for each $i \in \{1, \ldots, 99\}$. Could you find a smarter way to compute the result by expressing $\mathbf{y}$ in a coordinate system whose basis is an eigenbasis for $\mathbf{X}$? Why is this method easier than the step-by-step method described above?

Exercise 1.3 [Continuity and Differentiability]
Decide for each of the following functions whether it is continuous and/or differentiable at the given x-value. If the function is differentiable at x, give its derivative at x.

(a) $f(x) = \begin{cases} x^2 + 4 & \text{, for } x < -2 \\ 1 - 3x & \text{, for } x \geq -2 \end{cases}$ $\qquad x = -2$

(b) $f(x) = \begin{cases} x - 1 & \text{, for } x < 1 \\ (x-1)^2 & \text{, for } x \geq 1 \end{cases}$ $\qquad x = 1$

(c) The *sigmoid function* $\sigma(x) = \frac{1}{1+e^{-x}}$ $\qquad x = 0$

Exercise 1.4 [Critical Points of Univariate Functions]
Consider the univariate function $f(x) = x^3 - 4x + 5$.

(a) Give the derivative $\frac{\mathrm{d}f}{\mathrm{d}x}$.

(b) Give the maximum and minimum values of f on the interval $[0, 2]$. At which points are these values attained?

(c) Give the maximum and minimum values of f on the interval $[-3, 0]$. At which points are these values attained?

Exercise 1.5 [Critical Points of Multivariate Functions]
Consider the multivariate function $f(x, y) = 3x^3 + 2y^2 + xy$.

(a) What is the gradient of f at the input values $(2, 3)$?

(b) What is the Hessian of f at the input values $(0, 0)$?

(c) List all critical points of f. What can you say about each of them?

Exercise 1.6 [Lagrange Multipliers]
Consider the following objective function f and constraint function g:

$$f(x, y) = \ln(x^2 \cdot y^3)$$

$$g(x, y) = x^2 + y^2 - 1$$

Assume x and y to be strictly positive. Using Lagrange multipliers, find the critical points of $f(x, y)$ subject to $g(x, y) = 0$.

Exercise 1.7 [Probability Background]

(a) Assume that for a sample set of computer science students at UZH the following holds. One third of these students has Data Science and two thirds have Information Systems as major. The seminar Modern Data Analytics has been chosen by 40% of the students with major Data Science and by 20% of the students with major Information Systems. Compute the probability that a student from the sample set chooses the seminar Modern Data Analytics.

(b) Compute the probability that a student from the sample set who chooses the seminar Modern Data Analytics has Data Science as major.

(c) Have you heard about the virus DATAVID? Well, it can be caught by your smartphone while studying Data Science. It can also be transmitted from one smartphone to another. Scientists in Oxford have even developed a test for this virus with 90% sensitivity and 95% specificity. This means:

- The probability that a smartphone that has caught the virus is tested positively is 0.9.
- The probability that a smartphone that has not caught the virus is tested negatively is 0.95.

Scientists assume that already 5% of all smartphones in the world have the virus. Compute the probability that a randomly chosen smartphone has DATAVID if it is tested positively with the aforementioned test. Compute the probability that a randomly chosen smartphone has DATAVID if it is tested negatively.

Hint: You can use Bayes' rule for questions 1.8.(b) and 1.8.(c).

Exercise 1.8 [Some Statistics on Exam Scores]

The average scores of a group of 6 students in last year's Data Science exam were $(60, 75, 70, 50, 30, 75)$.

(a) Estimate the mean, median, variance, and standard deviation of the scores in last year's Data Science exam.

(b) How could you modify (standardise) the six scores so that the mean becomes 0 and the variance becomes 1?

(c) Are the estimators for the mean and the variance you used under (a) biased or unbiased?