Universität des Saarlandes
Prof. Dr. Dietrich Klakow
Lehrstuhl für Signalverarbeitung
NNIA Winter Term 2019/2020

Exercise Sheet 1

Linear Algebra Basics

(Solutions) **Deadline: 17.11.2020, 23:59**

Instructions

A good understanding of linear algebra is essential for understanding and working with many machine learning algorithms, especially deep learning algorithms. The purpose of the following exercises is to create the base for understanding concepts and algorithms introduced in the following lectures.

The exercises are based on the *Basic Linear Algebra for Deep Learning* post by Niklas Donges and the *Linear Algebra* chapter in the *Deep Learning* book by Ian Goodfellow, Yoshua Bengio and Aaron Courville.

Exercises

Exercise 1.1 - Mathematical objects (1 points)

Consulting Niklas's post or Deep Learning book, give short definitions and your own examples.

Note the conventional way to define different mathematical objects (you can find them in the DL book), e.g. we write scalars in italics and usually give them lowercase variable names. Throughout all the assignments (and of course outside of this course) stick to the conventional way of writing scalars, vectors etc[1].

a) scalar

b) vector

c) matrix

d) tensor

Solution 1.1

a) A **scalar** is just a single number. We write scalars in italics and usually give them lowercase variable names. $s \in R, s = 0.2$

[1]See how to make letters bold in LaTex math mode here.

b) A **vector** is an ordered array of numbers. Typically we give vectors lowercase names in bold typeface. The elements of the vector are identified by writing its name in italic typeface. $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}$

c) A **matrix** is a 2-D array of numbers, where each element is identified by indices instead of just one. We usually give matrices uppercase variable names with bold typeface, such as $\mathbf{A}$. The elements of the matrix are identified by writing its name in italic typeface.

$$\mathbf{A} = \begin{bmatrix} x_1 & x_2 & x_3 \\ x_4 & x_5 & x_6 \\ x_7 & x_8 & x_9 \end{bmatrix}$$

d) A **tensor** is an array of numbers, arranged on a regular grid, with a variable number of axes. A tensor has three indices, where the first one points to the row, the second to the column and the third one to the axis. We denote a tensor named "A" with this typeface: A.

We identify the element of A at coordinates (i, j, k) by writing $\mathbf{A}_{i,j,k}$.

Exercise 1.2 - Vectors in machine learning (2 points)

In machine learning we deal with data in multidimensional space. Each data point is characterised by a number of features: for example, a lecture can be characterised by the number of CoLi students, CS students and students from other departments (i.e. 3 features).

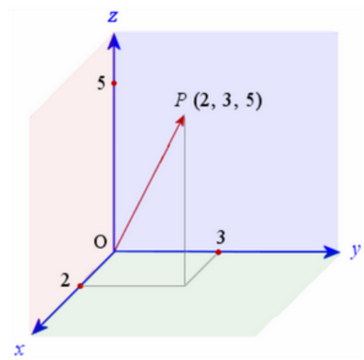
In space	Data point	In general
	$x = \begin{bmatrix} NumCoLi \\ NumCS \\ NumOther \end{bmatrix}$	$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

Figure 1: Different vector representations. Note: those are not representations of the same vector

a) Create your own data set of **three** data points where each data point is described by **two** features. Report:

  a) what are the data points in your data set (e.g. lectures);

  b) what are the features of the data points (e.g. number of CoLi students).

b) Represent the data set in physical space (as in Figure 1). Label the axes according to the features you defined.

c) Represent the data in form of vectors $\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}$.

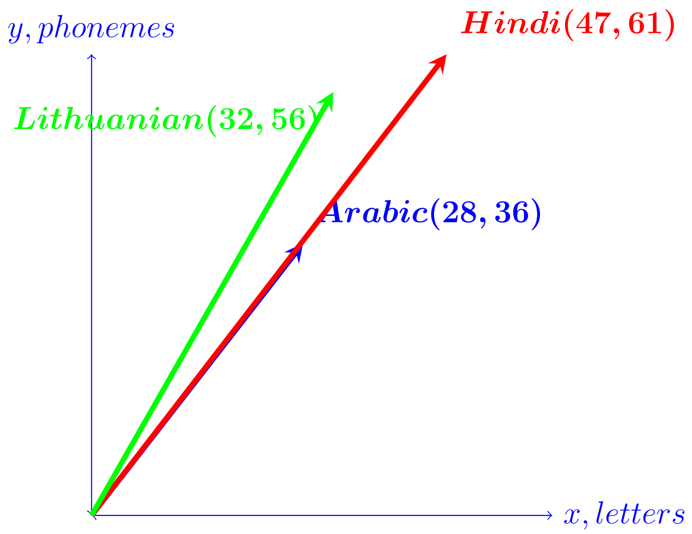
d) Represent the data in form of a matrix $\boldsymbol{X} \in \mathbb{R}^{n \times m}$, where n is the number of data points and m is the number of features.

Solution 1.2

a) This dataset comprises number of letters in the alphabet vs. number of phonemes in three languages: Arabic, Hindi and Lithuanian.

   a) Data points - Arabic, Hindi, Lithuanian.

   b) Features of the data points - number of letters in alphabet, number of phonemes.

b) 

c) Arabic:

$$\boldsymbol{x1} = \begin{bmatrix} 28 \\ 36 \end{bmatrix}$$

Hindi:

$$\boldsymbol{x2} = \begin{bmatrix} 47 \\ 61 \end{bmatrix}$$

Lithuanian:

$$\boldsymbol{x3} = \begin{bmatrix} 32 \\ 56 \end{bmatrix}$$

d) $\boldsymbol{X} \in \mathbb{R}^{n \times m}$

$\boldsymbol{X} = \begin{bmatrix} 28 & 47 & 32 \\ 36 & 61 & 56 \end{bmatrix}$

Exercise 1.3 - Operations on vectors and matrices (2.5 points)

We can perform mathematical operations such as addition, subtraction, multiplication on vectors, matrices and scalars. Carefully read the *Computational rules* part of the post[2] and perform the following exercises. If the computation is impossible, write *impossible* and argue why.

a) $\begin{bmatrix} 3 & 4 & 1 \\ 0 & 2 & 3 \end{bmatrix} \div 2 =$

b) $\begin{bmatrix} 3 & 4 & 1 \\ 0 & 2 & 3 \end{bmatrix} * \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix} =$

c) $\begin{bmatrix} 3 & 4 & 1 \\ 0 & 2 & 3 \end{bmatrix} - \begin{bmatrix} -1 & 2 & 10 \\ 3 & -6 & 3 \end{bmatrix} =$

d) $\begin{bmatrix} 3 & 4 & 1 \\ 0 & 2 & 3 \end{bmatrix} * \begin{bmatrix} 0 & 2 \\ 1 & -2 \\ 4 & 0 \end{bmatrix} =$

e) $\begin{bmatrix} 3 & 4 & 1 \\ 0 & 2 & 3 \end{bmatrix} * \begin{bmatrix} 10 & 5 \\ 0 & 2 \end{bmatrix} =$

If you are interested in the physical meaning of vectors, matrices and operations on them, watch videos from the Linear Algebra series by 3Blue1Brown.

Solution 1.3

a) $\begin{bmatrix} 1.5 & 2 & 0.5 \\ 0 & 1 & 1.5 \end{bmatrix}$

b) $\begin{bmatrix} 15 \\ 17 \end{bmatrix}$

c) $\begin{bmatrix} 4 & 2 & -9 \\ -3 & 8 & 0 \end{bmatrix}$

d) $\begin{bmatrix} 8 & -2 \\ 14 & -4 \end{bmatrix}$

e) Impossible. As number of columns in first matrix doesn't match number of rows in the second one.

Exercise 1.4 - Multiplication properties and types of matrices (3.5 points)

Read chapters *2.2 Multiplying Matrices and Vectors* and *2.3 Identity and Inverse Matrices* of the DL book or corresponding parts of the Niklas's post

On the internet, find the following definitions and write them down:

a) symmetric matrix

b) orthogonal matrix

[2]Pay attention to the dimensions and use the cheat sheets provided by Niklas.

c) unit vector

d) orthogonal vectors

Let $\boldsymbol{A}^{n\times n}$ be an orthogonal matrix, $\boldsymbol{B}^{m\times m}$ – a symmetric matrix and $\boldsymbol{C}^{m\times n}$ – a regular matrix, $\boldsymbol{I}$ – identity matrix, and λ - a scalar.
For each expression choose its equivalent, show the intermediate steps, give the dimensions of the result.

$(\boldsymbol{B}\lambda\boldsymbol{I})^T\boldsymbol{C} =$
a) $\lambda\boldsymbol{CB}$ b) $\boldsymbol{CB}^T\lambda$ c) $\lambda\boldsymbol{BCI}$

$\boldsymbol{A}^{-1}(\boldsymbol{CA}^{-1})^T\lambda =$
a) $\lambda\boldsymbol{C}^T$ b) $\lambda\boldsymbol{C}$ c) $\boldsymbol{A}^T\boldsymbol{C}^T\lambda$

$\boldsymbol{AA}^T\boldsymbol{B}^T\boldsymbol{C} =$
a) $\boldsymbol{CB}$ b) $\boldsymbol{B}^{-1}\boldsymbol{C}$ c) $\boldsymbol{BC}$

Solution 1.4

a) A **symmetric matrix** is a square matrix that satisfies $A^T = A$. The entries of a symmetric matrix are symmetric with respect to the main diagonal. If A is a symmetric matrix, then for every i,j

$$a_{ij} = a_{ji}$$

b) An **orthogonal matrix** is a real square matrix whose columns and rows are orthogonal unit vectors. When matrix Q is orthogonal, the given is true:

$$Q{\cdot}Q^T = Q^T{\cdot}Q = I$$

c) A **unit vector** is a vector of length 1 , and is denoted by circumflex or “hat”: $\hat{u}$

d) **Orthogonal vectors** are two vectors u and v whose dot product-

$$u{\cdot}v = 0$$

a) The answer is **c**.
$(B_{m\times m}\lambda I_{m\times m})^T C_{m\times n}$
$= \lambda B^T_{m\times m}C_{m\times n}$ (Property used- $B^T = B$)
$= \lambda B_{m\times m}C_{m\times n}I_{n\times n}$
Matrix dimension - $m \times n$

b) The answer is **a**.
$A^{-1}(CA^{-1})^T\lambda$
$= A^{-1}{\cdot}(A^{-1})^T{\cdot}C^T\lambda$ (Property used- $AA^T = I$)
$= IC^T\lambda$
$= \lambda C^T$
Matrix dimension - $n \times m$

c) $(A_{n\times n}A^T_{n\times n})B^T_{m\times m}C_{m\times n}$
$= I_{n\times n}B_{m\times m}C_{m\times n}$
There is an ambiguity in the expression above as AA^T results an identity matrix $\boldsymbol{I}$ of size $n \times n$. However, proceeding further we get $\boldsymbol{IBC}$ which is an undefined expression as size of $\boldsymbol{B}$ is $m \times m$ while the size of identity matrix $\boldsymbol{I}$ is $n \times n$.
Matrix dimension - undefined.

Exercise 1.5 - Vector norms (1 points)

Read the chapter *2.5 Norms* of the DL book.

Calculate L^2 and L^1 norms of the following vector: $\boldsymbol{a} = \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}$.

Bonus exercise (2 points): Draw all the vectors $\boldsymbol{x} \in \mathbb{R}^2$ for which

a) $||\boldsymbol{x}||_1 = 1$;

b) $||\boldsymbol{x}||_2 = 1$.

Solution 1.5

$$\begin{aligned} L^2 = \|x\|_2 &= (\sum_i |x_i|^2)^{\frac{1}{2}} \\ &= \sqrt{4+9+1} = \sqrt{14} = 3.741 \end{aligned}$$

$$\begin{aligned} L^1 = \|x\|_1 &= \sum_i |x_i| \\ &= 2+3+1 = 6 \end{aligned}$$

Bonus exercise

Let $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ represent a 2-dimensional vector. Given that,

$$||\boldsymbol{x}||_1 = x_1 + x_2 = 1 \quad (1)$$

$$||\boldsymbol{x}||_2 = \sqrt{x_1 + x_2} = 1 \quad (2)$$

Squaring equation 1 both sides-

$$\begin{aligned} &\implies (x_1 + x_2)^2 = 1 \\ &\implies x_1^2 + x_2^2 + 2x_1x_2 = 1 \end{aligned}$$

Squaring equation 2 both sides and replacing it in the above equation-

$$\begin{aligned} &\implies x_1^2 + x_2^2 = 1 \\ &\implies 1 + 2x_1x_2 = 1 \\ &\implies x_1x_2 = 0 \end{aligned}$$

Therefore, all possible vectors are-

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ 0 \end{bmatrix}, \boldsymbol{x} = \begin{bmatrix} 0 \\ x_2 \end{bmatrix}$$

$$\boldsymbol{x} = \begin{bmatrix} x_1 & 0 \end{bmatrix}, \boldsymbol{x} = \begin{bmatrix} 0 & x_2 \end{bmatrix}$$

Submission instructions

The following instructions are mandatory. If you are not following them, tutors can decide to not correct your exercise.

- You have to submit the solutions of this assignment sheet as a team of 2-3 students.
- Hand in a **single** PDF file with your solutions.
- Therefore Make sure to write the student ID and the name of each member of your team on your submission.
- Your assignment solution must be uploaded by only **one** of your team members to the course website.
- If you have any trouble with the submission, contact your tutor **before** the deadline.