Proofs, Exercises and Literature - Linear Algebra Recap

1 Exercises

1.1 Trivia Questions from the Lecture

1. $A \in \mathbb{R}^{n \times r}$, $B \in \mathbb{R}^{m \times r}$, which product is well-defined?
   - A. BA
   - B. $A^\top B$
   - C. $AB^\top$

Solution: The correct solution is C. $AB^\top$, the connecting dimension is r.

2. $A \in \mathbb{R}^{n \times r}$, $B \in \mathbb{R}^{m \times r}$, what is equal to $(AB^\top)^\top$?
   - A. $A^\top B$
   - B. $B^\top A^\top$
   - C. $BA^\top$

Solution: The correct solution is C. $BA^\top$, we have $(AB^\top)^\top = B^{\top^\top} A^\top = BA^\top$.

3. What is the matrix product computed by $C_{ji} = \sum_{s=1}^{r} A_{is} B_{js}$?
   - A. $C = AB^\top$
   - B. $C = B^\top A$
   - C. $C = BA^\top$

Solution: The correct solution is C. $BA^\top$, we have $C_{ji} = \sum_{s=1}^{r} A_{is} B_{js} = \sum_{s=1}^{r} B_{js} A_{is} = B_{j\cdot} A_{i\cdot}^\top$, hence $C = BA^\top$.

4. $A, B \in \mathbb{R}^{n \times n}$ have an inverse A^{-1}, B^{-1}, what is generally **not** equal to $AA^{-1}B$?
   - A. $A^{-1}BA$
   - B. B
   - C. $BB^{-1}B$

Solution: the correct answer is A. $A^{-1}BA$, since the matrix product between the matrices A^{-1} and BA is generally not commutative. Answers B. and C. are correct because

$$AA^{-1}B = IB = B = BB^{-1}B.$$

5. Let $v, w \in \mathbb{R}^d$, $\alpha \in \mathbb{R}$, then $\|\alpha v + w\| \leq$
   A. $\alpha\|v + w\|$
   B. $|\alpha|\|v\| + \|w\|$
   C. $\alpha\|v\| + \|w\|$

> **Solution:** The correct answer is B. $|\alpha|\|v\| + \|w\|$, because
>
> $$\begin{aligned} \|\alpha v + w\| &\leq \|\alpha v\| + \|w\| && \text{(triangle inequality)} \\ &= |\alpha|\|v\| + \|w\| && \text{(homogenity).} \end{aligned}$$

6. Let $A, B \in \mathbb{R}^{n \times r}$, $\alpha \in \mathbb{R}$, then $\|A\| \leq$
   A. $\|A - B\| + \|B\|$
   B. $\alpha\|\frac{1}{\alpha}A\|$
   C. $\|A\|^2$

> **Solution:** The correct answer is A. $\|A - B\| + \|B\|$, because
>
> $$\|A\| \leq \|A - B + B\| \leq \|A - B\| + \|B\|,$$
>
> where the last inequality derives from the triangle inequality. Answer B. is nott correct because the inequalitty does not hold for negative α and C. is not correct for matrices A having small entries (for example take the 1×1 matrix A=(0.1), then $\|A\| = 0.1 > 0.1^2$).

7. Let $A, B \in \mathbb{R}^{n \times n}$, A is orthogonal, what is **not** equal to $\mathrm{tr}(ABA^\top)$?
   A. $\mathrm{tr}(A^\top BA)$
   B. $\mathrm{tr}(B)$
   C. $\mathrm{tr}(ABA)$

> **Solution:** The correct answer is C. $\mathrm{tr}(ABA)$. The other answers are not correct because the cycling property of the trace yields $\mathrm{tr}(ABA^\top) = \mathrm{tr}(BA^\top A)$, and
>
> $$\begin{aligned} \mathrm{tr}(ABA^\top) &= \mathrm{tr}(BA^\top A) = \mathrm{tr}(BI) = \mathrm{tr}(B) && \text{(orthogonality of } A\text{)} \\ &= \mathrm{tr}(IB) = \mathrm{tr}(AA^\top B) = \mathrm{tr}(A^\top BA). && \text{(cycling property and orthogonality)} \end{aligned}$$

1.2 Additional Exercises

1. Compute the matrix product AB inner-product-wise and outer-product-wise

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 2 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 2 \\ 3 & 1 \\ 1 & 2 \end{pmatrix}.$$

Solution: Computing the matrix product *inner product wise*:

$$AB = \begin{pmatrix} 1\cdot 0 + 2\cdot 3 + 0\cdot 1 & 1\cdot 2 + 2\cdot 1 + 0\cdot 2 \\ 0\cdot 0 + 2\cdot 3 + 4\cdot 1 & 0\cdot 2 + 2\cdot 1 + 4\cdot 2 \end{pmatrix} = \begin{pmatrix} 6 & 4 \\ 10 & 10 \end{pmatrix}$$

Computing the matrix product *outer product wise*:

$$\begin{aligned} AB &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & 2 \end{pmatrix} + \begin{pmatrix} 2 \\ 2 \end{pmatrix} \begin{pmatrix} 3 & 1 \end{pmatrix} + \begin{pmatrix} 0 \\ 4 \end{pmatrix} \begin{pmatrix} 1 & 2 \end{pmatrix} \\ &= \begin{pmatrix} 1\cdot 0 & 1\cdot 2 \\ 0\cdot 0 & 0\cdot 2 \end{pmatrix} + \begin{pmatrix} 2\cdot 3 & 2\cdot 1 \\ 2\cdot 3 & 2\cdot 1 \end{pmatrix} + \begin{pmatrix} 0\cdot 1 & 0\cdot 2 \\ 4\cdot 1 & 4\cdot 2 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 6 & 2 \\ 6 & 2 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 4 & 8 \end{pmatrix} \\ &= \begin{pmatrix} 6 & 4 \\ 10 & 10 \end{pmatrix} \end{aligned}$$

2. You have observations of 5 symptoms of a disease for three patients represented in the binary matrix

$$A = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

Compute the matrix $AA^\top$ and $A^\top A$ and interpret the result with regard to the scenario.

Solution: The matrix represents a data table of the following format, where $\mathtt{S}_i$ denotes the feature of sympton i and `P` stands for patient ID:

P	$\mathtt{S}_1$	$\mathtt{S}_2$	$\mathtt{S}_3$	$\mathtt{S}_4$	$\mathtt{S}_5$
1	1	0	1	1	0
2	1	1	0	0	0
3	0	1	0	0	1

The matrix products

$$AA^\top = \begin{pmatrix} 3 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}, \qquad A^\top A = \begin{pmatrix} 2 & 1 & 1 & 1 & 0 \\ 1 & 2 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

contrast either the features or the patients. That is, the representation of the matrix product in the table format would look as follows:

$AA^\top$	$\mathtt{P}_1$	$\mathtt{P}_2$	$\mathtt{P}_3$
$\mathtt{P}_1$	3	1	0
$\mathtt{P}_1$	1	2	1
$\mathtt{P}_1$	0	1	2

$A^\top A$	$\mathtt{S}_1$	$\mathtt{S}_2$	$\mathtt{S}_3$	$\mathtt{S}_4$	$\mathtt{S}_5$
$\mathtt{S}_1$	2	1	1	1	0
$\mathtt{S}_2$	1	2	0	0	1
$\mathtt{S}_3$	1	0	1	1	0
$\mathtt{S}_4$	1	0	1	1	0
$\mathtt{S}_5$	0	1	0	0	1

The table of $AA^\top$ denotes in entry jl the number of symptoms patient j and patient l have in common. The table of $A^\top A$ denotes in entry ik the number of patients which exibit symptoms i and k.

3. Find a matrix/vector notation to compute the vector of average feature values for a matrix $A \in \mathbb{R}^{n\times d}$, representing n observations of d features. Make an example for your computation.

Solution: The vector representing the average for every feature value is computed by $\boldsymbol{\mu} = \frac{1}{n}A^\top \mathbf{1}$ where $\mathbf{1}$ is the n-dimensional one-vector, having all values equal to one. This is the case, because we have according to the definition of matrix multiplications by the row-times-column column scheme:

$$\boldsymbol{\mu}^\top = \frac{1}{n}\mathbf{1}^\top A = \frac{1}{n}\begin{pmatrix}\mathbf{1}^\top A_{\cdot 1} & \dots & \mathbf{1}^\top A_{\cdot d}\end{pmatrix},$$

that is, for $1 \leq j \leq n$ we have that the i-th entry of $\boldsymbol{\mu}$ is given as

$$\boldsymbol{\mu}_i = \frac{1}{n}\mathbf{1}^\top A_{\cdot i} = \frac{1}{n}\sum_{j=1}^{n} 1 \cdot A_{ji} = \frac{1}{n}\sum_{j=1}^{n} A_{ji},$$

which is equal to the average value for feature i.

4. Show that $\|A - B\|^2 = -2\operatorname{tr}(AB^\top) + 2n$ for orthogonal matrices $A, B \in \mathbb{R}^{n\times n}$.

Solution: *Proof:* Let $A, B \in \mathbb{R}^{n\times n}$ be orthogonal matrices. Orthogonal matrices satisfy the property $AA^\top = A^\top A = I$ and $BB^\top = B^\top B = I$. Thus, we have

$$\begin{aligned}
\|A - B\|^2 &= \|A\|^2 - 2\operatorname{tr}(AB^\top) + \|B\|^2 && \text{(binomial formula for matrix norms)}\\
&= \operatorname{tr}(A^\top A) - 2\operatorname{tr}(AB^\top) + \operatorname{tr}(B^\top B) && \text{(definition of elementwise matrix } L_2\text{-norm)}\\
&= \operatorname{tr}(I) - 2\operatorname{tr}(AB^\top) + \operatorname{tr}(I) && \text{(orthogonality of } A \text{ and } B)\\
&= -2\operatorname{tr}(AB^\top) + 2n,
\end{aligned}$$

because $\operatorname{tr}(I) = \underbrace{1 + \ldots + 1}_{n \text{ times}} = n$. This concludes the proof.

5. Show that the following norms are orthogonal invariant
   - the vector L_2-norm
   - the Frobenius norm (matrix L_2-norm)
   - the operator norm

Solution: A norm is orthogonal invariant if multiplying the argument with an orthogonal matrix from the left does not change the value of the norm. Let A be a $(n \times n)$ orthogonal matrix. For the

L_2-norm of a vector $\mathbf{v} \in \mathbb{R}^n$ we have:

$$\begin{aligned}
\|A\mathbf{v}\|^2 &= (A\mathbf{v})^\top A\mathbf{v} && \text{(Definition)}\\
&= \mathbf{v}^\top A^\top A\mathbf{v}\\
&= \mathbf{v}^\top \mathbf{v} && (A^\top A = I)\\
&= \|\mathbf{v}\|^2
\end{aligned}$$

For the L_2-norm of a $(n \times d)$ matrix AD we have:

$$\begin{aligned}
\|AD\|^2 &= \operatorname{tr}((AD)^\top AD) && \text{(Definition)}\\
&= \operatorname{tr}(D^\top A^\top AD)\\
&= \operatorname{tr}(D^\top D) && (A^\top A = I)\\
&= \|D\|^2
\end{aligned}$$

For the operator norm of the matrix AD, we have:

$$\begin{aligned}
\|AD\|_{op} &= \max_{\mathbf{x}:\|\mathbf{x}\|=1} \|AD\mathbf{x}\| && \text{(Definition)}\\
&= \max_{\mathbf{x}:\|\mathbf{x}\|=1} \|D\mathbf{x}\| && \text{(orthogonal invariance of } L_2 \text{ vector norm)}\\
&= \|D\|_{op}
\end{aligned}$$

2 Recommended Literature

As always, the best exercise is to go through the lecture and see if you can follow the steps with pen and paper and to make the exercises. The linear algebra lecture is tailored to the needs for this lecture. If you want a more general and extensive overview, the following material is recommended.

Linear Algebra and Optimization for Machine Learning by Charu C. Aggarwal

Chapter one, in particular Sections 1.1-1.3 give a good introduction to vector spaces and matrices, norms and matrix multiplication. If you want to go deeper into the subject of linear algebra, then I would recommend to have a look at the Sections 2.1-2.4 and 7.1 and 7.2 as well.

The Course Linear Algebra and Applications (2DBI00) from Michiel Hochstenbach

Michiel is giving a very good course at TU/e about linear algebra and applications, where the applications are often data mining/machine learning problems. You can see the video lectures of 2018/2019 at the videocollege. Search for 2DBI00 in channels (not videos!). Select 2018-2019 to watch the latest recorded videos.