# Übung zu Deep Learning, SoSe 21

**Tutorial 02: Mathematical Background**

Pavlo Mospan and Anastasia Karsten

24.04.2021

## 1 Linear Algebra

### 1.1 Multiple Transpositions

**Task :** *Simplify $(A^T)^T$. Justify your solution for instance using the index notation of matrices!*

**Solution :** The transpose of a matrix is an operator which flips a matrix over its diagonal; that is, it switches the row (i) and column (j) indices of the matrix: $A_{ij} = (A^T)_{ji}$

The $(j,i)$-entry of $A^T$ is the $(i,j)$-entry of A, so the $(i,j)$-entry of $(A^T)^T$ is the $(j,i)$-entry of $A^T$, which is the $(i,j)$-entry of A. Thus all entries of $(A^T)^T$ coincide with the corresponding entries of A.

**So, the transpose of the transpose of A is A, i.e.** $(A^T)^T = A$

### 1.2 Transposing a Matrix Product 2

**Task :** *Show that $(AB)^T = B^T A^T$ for a matrices $A \in R^{m \times n}$ and $B \in R^{n \times p}$. What can be said about the dimensions of $(AB)^T$?*

**Solution :**

*Option 1*: If $A = \begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix}$ and $B = \begin{vmatrix} 1 & 0 \\ 2 & -1 \end{vmatrix}$ , then the product $(AB) = \begin{vmatrix} 5 & -2 \\ 11 & -4 \end{vmatrix}$.

And the transpose of (AB) is: $(AB)^T = \begin{vmatrix} 5 & 11 \\ -2 & -4 \end{vmatrix}$.

If we take the transpose of A and B separately and multiply A with B, then we have:
$B^T A^T = \begin{vmatrix} 1 & 2 \\ 0 & -1 \end{vmatrix} \begin{vmatrix} 1 & 3 \\ 2 & 4 \end{vmatrix} = \begin{vmatrix} 5 & 11 \\ -2 & -4 \end{vmatrix}$. So, $(AB)^T = B^T A^T$

*Option 2* : We can show, that $((AB)_{ij})^T = (\sum_{l=1}^{n} a_{il} b_{lj})^T = \sum_{l=1}^{n} b_{jl} a_{li}$

**Answer to the second question:** Dimensions of $(AB)^T \in R^{p \times m}$

### 1.3 Brackets in Matrix Multiplications

**Task 1:** *How many operations (additions or multiplications of scalars) does a trivial implementation of a matrix multiplication $AB$ with $A \in R^{m \times n}$ and $B \in R^{n \times p}$ need?*

**Solution:** Multiplications: $n * m * p$. Additions: $(n-1) * (m * p)$

**Task 2:** *How many operations does $(AB)C$ with $A \in R^{16 \times 2}$, $B \in R^{2 \times 4}$ and $C \in R^{4 \times 8}$ need? Use the associative property of the matrix multiplication to find the fastest solution.*

**Solution:** Associative property of the matrix multiplication : $(AB)C = A(BC)$

**Let's calculate for (AB)C**.

MULT: $AB = 2 * 16 * 4 = 128$. $(AB)C = 4 * 16 * 8 = 512$. MULT Total: **640**

ADD: $AB = 16 * 4 = 64$. $(AB)C = 3 * 16 * 8 = 384$. ADD Total: **448**

**Let's calculate for A(BC)**.

MULT: $BC = 4 * 2 * 8 = 64$. $A(BC) = 2 * 16 * 8 = 256$. MULT Total: **320**

ADD: $BC = 3 * 2 * 8 = 48$. $A(BC) = 16 * 8 = 128$. ADD Total: **176**

**The fastest solution will take 320 multiplications, 176 additions**

## 2 Differential Calculus

### 2.1 Quotient Rule

**Task :** *Derive the Quotient rule*

$$f(x) = \frac{g(x)}{h(x)} \rightarrow \frac{d}{dx} f(x) = \frac{h(x) \frac{d}{dx} g(x) - g(x) \frac{d}{dx} h(x)}{h(x)^2}$$

*using the chain rule and the product rule of differentiation.*

**Solution :**

$$\frac{d}{dx}\frac{g(x)}{h(x)} = \frac{d}{dx}(g(x)h(x)^{-1})$$

$$= \frac{d}{dx}g(x)h(x)^{-1} + g(x)(-1)h(x)^{-2}\frac{d}{dx}h(x)$$

$$= \frac{\frac{d}{dx}g(x)}{h(x)} - \frac{g(x)\frac{d}{dx}h(x)}{h(x)^2}$$

$$= \frac{\frac{d}{dx}g(x)h(x)}{h(x)^2} - \frac{g(x)\frac{d}{dx}h(x)}{h(x)^2}$$

$$= \frac{\frac{d}{dx}g(x)h(x) - g(x)\frac{d}{dx}h(x)}{h(x)^2}$$

## 2.2 Derivative of the Sigmoid Function

**Task :** *Derive the so-called sigmoid function*

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

*and express the derivative in terms of* $\sigma(x)$*.*

$$\frac{d}{dx}\sigma(x) = \frac{d}{dx}(\frac{1}{1+e^{-x}})$$

$$= \frac{(1+e^{-x})\frac{d}{dx}(1) - (1)\frac{d}{dx}(1+e^{-x})}{(1+e^{-x})^2}$$

$$= \frac{(1+e^{-x})(0) - (1)(-e^{-x})}{(1+e^{-x})^2}$$

$$= \frac{e^{-x}}{(1+e^{-x})^2}$$

$$= \frac{-1+1+e^{-x}}{(1+e^{-x})^2}$$

$$= \frac{1+e^{-x}}{(1+e^{-x})^2} - \frac{1}{(1+e^{-x})^2}$$

$$= \frac{1}{(1+e^{-x})} - \frac{1}{(1+e^{-x})^2}$$

$$= \frac{1}{(1+e^{-x})}(1 - \frac{1}{(1+e^{-x})})$$

$$= \sigma(x)(1-\sigma(x))$$

## 2.3 Applying Gradients

**Task :** Calculate the gradient of the function

$$f(x) = x_1 + x_2$$

For a given point **a**, by how much does the value of **f** change, if we change **a** by a magnitude of $\varepsilon (0 < \varepsilon \ll 1)$ in the direction of the gradient, i.e., calculate

$$\Delta = f(\mathbf{a} + \epsilon) - f(\mathbf{a})$$

with

$$\epsilon = \varepsilon \frac{\bigtriangledown f(x)|_{x=a}}{\| \bigtriangledown f(x)|_{x=a} \|_2}$$

Compare this to a change of **a** with magnitude of $\varepsilon$ in the direction of $x_1$ or $x_2$. How can the difference be explained qualitatively?

**Solution :** Let's find out $\epsilon$ :

$\bigtriangledown f(x)|_{x=a}$ is a vector of first order partial derivatives : $\bigtriangledown f(x) = \left[\frac{df}{dx_1}; \frac{df}{dx_2}\right] = \left[1; 1\right]$

$\| \bigtriangledown f(x)|_{x=a} \|_2$ is $l^2$ norm of a vector of 1st order partial derivatives $\Rightarrow \| \bigtriangledown f(x) \|_2 = \sqrt{2}$

So, $\epsilon = \varepsilon \frac{(1;1)}{\sqrt{2}} = \varepsilon \frac{1}{\sqrt{2}}$

Thereof

$$\Delta = f(\mathbf{a} + \varepsilon \frac{1}{\sqrt{2}}) - f(\mathbf{a})$$

$\Delta$ is changes by the magnitude of $\varepsilon$