# Tutorial 02: Mathematical Background (20P)

Please submit your answers by 4th Nov 23:59 to manuel.milling@informatik.uni-augsburg.de.

## 1 Linear Algebra

### 1.1 Multiple Transpositions (2P)

Simplify $(\boldsymbol{A}^{\mathrm{T}})^{\mathrm{T}}$. Explain your solution!

### 1.2 Transposing a Matrix Product 2 (3P)

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

### 1.3 Brackets in Matrix Multiplications (4P)

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

How many operations does $\boldsymbol{ABC}$ with $\boldsymbol{A} \in \mathbb{R}^{16\times 2}$, $\boldsymbol{B} \in \mathbb{R}^{2\times 4}$ and $\boldsymbol{C} \in \mathbb{R}^{4\times 8}$ need? Use the associative property of matrix multiplications to find the fastest solution.

## 2 Differential Calculus

### 2.1 Quotient Rule (3P)

Derive the Quotient rule

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

using the chain rule and the product rule of differentiation.

### 2.2 Derivative of the Sigmoid Function (4P)

Derive the so-called sigmoid function

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

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

## 2.3 Applying Gradients (4P)

Calculate the gradient of the function

$$f(\boldsymbol{x}) = x_1 + x_2.$$

For a given point $\boldsymbol{a}$, by how much does the value of $f$ change, if we change $\boldsymbol{a}$ by a magnitude of $\epsilon$ ($\epsilon \ll 1$) in the direction of the gradient, i.e., calculate

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

with

$$\boldsymbol{\epsilon} = \epsilon \frac{\nabla f(\boldsymbol{x})|_{\boldsymbol{x}=\boldsymbol{a}}}{\|\nabla f(\boldsymbol{x})|_{\boldsymbol{x}=\boldsymbol{a}}\|_2}.$$

Compare this to a change of $\boldsymbol{a}$ with magnitude of $\epsilon$ in the direction of $x_1$ or $x_2$. Does the result satisfy your expectations?