EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

MATH. - NATURWISS. FAKULTÄT
FACHBEREICH INFORMATIK
KOGNITIVE SYSTEME · PROF. A. ZELL

Deep Neural Networks

Assignment 1

Assignment due by: 08.05.2018, Discussions on: 15.05.2018

Important: If you have not already done so, please register via ILIAS until 01.05.2018.

Notes: Like the lecture slides, the assignments will be kept in English, however answers can be submitted in either German or English. Please always justify your answers, even for yes/no questions, unless the description explicitly states otherwise. Full grades can only be awarded when answers are justified.

Question 1 Advances in neural network technology (6 Points)

Which do you think are the three most important factors that gave rise to modern deep neural networks? Explain why you think this is the case? (Pick one of the point from Chapter 1, slide 20)

Please don't write more than one page handwritten (or about half a page typed) in total.

Question 2 Eigenvectors and eigenvalues (8 Points)

The eigenvalues of an $n \times n$ matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ are the solutions for λ of the equation $\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$ (the corresponding values for $\boldsymbol{x}$ are the eigenvectors).

(a) Suppose that the matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is diagonalizable, that is, $\boldsymbol{A} = \boldsymbol{T}\boldsymbol{\Lambda}\boldsymbol{T}^{-1}$ for an invertible matrix $\boldsymbol{T} \in \mathbb{R}^{n \times n}$, where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, ..., \lambda_n)$ is the diagonal matrix of eigenvalues. Use the notation $\boldsymbol{t}^{(i)}$ for the columns of $\boldsymbol{T}$, so that $\boldsymbol{T} = [\boldsymbol{t}^{(1)} \cdots \boldsymbol{t}^{(n)}]$. Show that $\boldsymbol{A}\boldsymbol{t}^{(i)} = \lambda_i \boldsymbol{t}^{(i)}$, so that the eigenvector/eigenvalue pairs of $\boldsymbol{A}$ are $(\boldsymbol{t}^{(i)}, \lambda_i)$.

(b) A matrix $\boldsymbol{U} \in \mathbb{R}^{n \times n}$ is orthogonal if $\boldsymbol{U}^T = \boldsymbol{U}^{-1}$. Suppose that the matrix $\boldsymbol{B} \in \mathbb{R}^{n \times n}$ is diagonalized by an orthogonal matrix $\boldsymbol{U} \in \mathbb{R}^{n \times n}$, such that $\boldsymbol{B} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$. Show that $\boldsymbol{B}$ has to be a symmetric matrix, i.e. $\boldsymbol{B}^T = \boldsymbol{B}$.

(c) Show that if $\boldsymbol{A}$ is positive semi-definite (i.e. $\boldsymbol{A}$ is symmetric and $\forall \boldsymbol{z} \,.\, \boldsymbol{z}^T \boldsymbol{A} \boldsymbol{z} \geq 0$), then all of its eigenvalues are non-negative.

Question 3 Trace operator, matrix derivatives (6 Points)

Let A,B and C be three matrices such that products are well defined. Show the following properties (the matrix derivative means $(\nabla_{\boldsymbol{A}})_{ij} = \frac{\partial}{\partial A_{ij}}$):

Hint: Write out the expressions explicitly in terms of sums over indices.

(a) $\nabla_{\boldsymbol{A}} \mathrm{tr}(\boldsymbol{A}\boldsymbol{B}) = \boldsymbol{B}^T$

(b) $\nabla_{\boldsymbol{A}^T} f(\boldsymbol{A}) = (\nabla_{\boldsymbol{A}} f(\boldsymbol{A}))^T$

(c) $\nabla_{\boldsymbol{A}} \mathrm{tr}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^T\boldsymbol{C}) = \boldsymbol{C}\boldsymbol{A}\boldsymbol{B} + \boldsymbol{C}^T\boldsymbol{A}\boldsymbol{B}^T$

(d) If $\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}$, then $\nabla_{\boldsymbol{A}} f(\boldsymbol{z}) = (\nabla_{\boldsymbol{z}} f(\boldsymbol{z})) \otimes \boldsymbol{x}$

Note: $\boldsymbol{u} \otimes \boldsymbol{v}$ denotes the outer product $\boldsymbol{u}\boldsymbol{v}^T$