

# Uebung04-Hofmann-Jeackel

June 4, 2019

## Übungszettel 4

Inhaltsbasierte Bild- und Videoanalyse - SoSe 19

Hinweis: Alle Lösungen sollen in einem IPython Notebook realisiert werden, wobei Teilaufgaben und Zwischenergebnisse ausgegeben bzw. visualisiert werden sollen. Die Übungszettel sollen in Gruppen von 3 Personen abgegeben werden. Die Datei soll nach folgendem Muster benannt werden: uebung01 {nachname1} {nachname2} {nachname3}.ipynb

Für eine Einleitung in IPython siehe z.B.: <http://cs231n.github.io/ipython-tutorial/>

## 0.1 Aufgabe 1 - Künstliche Neuronale Netze

### 0.1.1 Aufgabe 1.1: Grundlagen (0,5+0,5+0,5+0,5+0,5 Punkte)

**a) Erklären Sie den Unterschied zwischen Trainings-, Validierungs- und Testdatensatz.** Den gesamten Datensatz der zum Training eines Neuronalen Netzes genutzt wird teilt man im allgemeinen in drei Datensätze auf. Der **Trainingsdatensatz** (meist zusätzlich in sg. Batches unterteilt) wird zum eigentlichen Training des Netzes benutzt.

Auf dem **Validierungsdatensatz** wird überprüft ob das Netzwerk auch das gelernt hat was wir von ihm wollen und nicht etwa nur die Trainingsdatensätze auswendig kann.

Der Test wie gut das trainierte Netz abschneidet, wird mit Hilfe des **Testdatensatzes** durchgeführt. Mit diesen frischen, dem Netz unbekannten Daten kann dann etwa die *accuracy* bestimmt werden.

**b) Erklären Sie den Begriff Overfitting.** Overfitting beschreibt den Prozess, wenn ein Netzwerk die Trainingsdaten auswendig lernt. Anstatt allgemeiner Features herauszuarbeiten, werden Features erlernt, die zu sehr den Trainingsdaten entsprechen. Somit schneidet das Netz zwar auf den Trainingsdaten großartig ab, kann aber mit neuen Daten nicht umgehen.

**c) Was bewirken L1 und L2 Regularisierung?** Regularisierung addiert eine "Strafe" auf die Loss Funktion, je nachdem wie komplex das Netz wird. Hierdurch soll dem Overfitting entgegen gewirkt werden.

**L1**

$$R(W) = \sum_k \sum_l |W_{k,l}|$$

**L2**

$$R(W) = \sum_k \sum_l W_{k,l}^2$$