

Computerlinguistische Anwendungen

Benjamin Roth

CIS LMU

Organisatorisches

- 3 Stunden Vorlesung Mi 14:00-17:00 (c.t.), L 155
- 2 Stunden Übung Do 14:00-16:00 (c.t), Sibirien/Gobi
- ECTS: 9
- Prüfung: Klausur (wird noch festgelegt, voraussichtlich letzte Vorlesungswoche)

Übungsaufgaben

- Bearbeitung in Teams von 2-3 Personen. (Bitte überlegen Sie sich bis zur Übung morgen die Teams → Webformular)
- Ausgabe der Übungen: Donnerstags 14:00
- Abgabe-Deadline jeweils: Freitag 16:00 (8 Tage später)
- Bewertung der Übungen u.a. automatisiert durch Unit-Tests.
- Klausurbonus: In Abhängigkeit der erreichten Übungspunkte wird ein Klausurbonus von bis zu 10% der maximal erreichbaren Klausurpunkte gewährt, wenn die Klausur auch ohne die Bonuspunkte als bestanden gewertet würde.

Machine-Learning *Challenge*

- Aufgabe: eine möglichst gute Vorhersage durch ein Machine-Learning Modell auf einem vorgegebenen Datensatz.
- Am Ende des Kurses, zählt wie Übungsblätter.
- um Punkte zu bekommen:
 - ▶ einfache Baseline muss übertroffen werden
 - ▶ kurze Vorstellung des gewählten Ansatzes
- das beste Team erhält außerdem einen kleinen Preis
- Mehr Details im Laufe des Semesters.

Themen

- Einführung in Machine-Learning und Implementierung einfacher Klassifikatoren
 - ▶ Perceptron
 - ▶ Naive Bayes
- Einführung in die Machine-Learning Bibliotheken Numpy/Scikit-Learn
- Grundlagen von Neuronalen Netzwerken für die Sprachverarbeitung, die Bibliothek Keras
- Implementierung typischer NLP-Anwendungen wie z.B.
 - ▶ Spam-detection
 - ▶ Sentiment-Analyse
 - ▶ Paraphrasen-Erkennung
 - ▶ Wortartenerkennung
 - ▶ Erkennung von Entitäten-Typen
 - ▶ Klassifikation von Relationen
- Evtl. Firmenvorträge gegen Ende des Semesters.

Plan für heute: Wiederholung

- Repräsentation von Dokumenten/Trainingsinstanzen in Python
- Klassifikation mit Perzeptron
- Unit Tests