

Aufgabe 6

1) Optimierung des LC-CRF-Tagger-Trainings

In der letzten Aufgabe haben Sie einen CRF-Tagger trainiert. Nun sollen Sie das Trainingsprogramm verbessern durch

- Early-Stopping
- L_1 -Regularisierung

2) Implementierung eines Taggerprogrammes

Außerdem sollen Sie ein Programm schreiben, welches die gespeicherten Parameter einliest und dann Eingabesätze mit dem Viterbi-Algorithmus annotiert. Die Eingabesätze werden aus einer Datei eingelesen, die ein Wort pro Zeile enthält, wobei auf jeden Satz eine Leerzeile folgt. Die Ausgabe erfolgt im gleichen Format wie die Trainingsdaten.

Vorüberlegungen

- Wie implementieren Sie Early Stopping?
- Wie kann die L_1 -Regularisierung effizienter gemacht werden?
- Welche Teilaufgaben umfasst das Taggerprogramm?

Optimieren Sie den L_1 -Regularisierungsfaktor und die Lernrate auf Developmentdaten und berechnen Sie zum Schluss die Tagginggenauigkeit auf den Testdaten.

Schicken Sie alle Programme, den optimalen Metaparameter und die erzielte Genauigkeit auf den Testdaten an `schmid@cis.lmu.de`.