

Künstliche Intelligenz (SS 2021)

Aufgabenblatt 3

zu bearbeiten bis: 21.06.2021

Aufgabe 3.1 (Naive Bayes: Implementierung (Teil 2))

Finalisieren Sie Ihre Naive-Bayes-Implementierung:

- (a) **Anwendung:** Implementieren Sie die Methode `apply()`, die gegebenen Dokumenten jeweils eine Klasse zuweist.
Beachten Sie die Hinweise aus der Vorlesung:
- Glätten Sie die Wahrscheinlichkeiten auf eine Mindest-Wahrscheinlichkeit ϵ .
 - Beachten Sie, dass in Ihr Produkt alle Worte des Vokabulars einfließen sollte. Kommt ein Wort nicht im Dokument vor, gilt die Gegenwahrscheinlichkeit.
 - Da viele einzelne Wahrscheinlichkeiten aufmultipliziert werden, erhalten wir sehr kleine (im Allgemeinen zu kleine) Zahlen. Verwenden Sie – wie besprochen – den Logarithmus.
- (b) **Testen:** Klassifizieren Sie die Dokumente der Testmenge mit folgendem Befehl:
- ```
python classifier.py --apply
```
- (c) **Messen:** Erweitern Sie Ihr Programm, so dass die Genauigkeit – d.h. der Prozentsatz korrekt klassifizierter Testdokumente – gemessen wird. Vergleichen Sie hierzu das Resultat Ihres Klassifikators mit den echten Kategorien in `labels`. Berechnen Sie auch die sogenannte Confusion Matrix. Diese gibt an, welche Klassen besonders häufig miteinander verwechselt werden.
- (d) **Optimieren:** Testen Sie in einer Grid Search verschiedene geeignete Werte für die Minimalwahrscheinlichkeit  $\epsilon$ . Welches  $\epsilon$  funktioniert am besten?