

Data Mining in technischen Anwendungen

WS 2017/18

Übungsblatt Semester Aufgabe

Betreuer: Janosch Henze, <jhenze@uni-kassel.de>
Jens Schreiber, <jens.schreiber@uni-kassel.de>

Aufgabe 1: Erstellen sie einen Klassifikator für den gegebenen Datensatz.

Diese freiwillige Semesteraufgabe dient dazu, Ihr erlerntes Wissen zu zeigen. Durch Bearbeiten der Aufgabe können Sie sich alleine oder in Gruppen von maximal Personen einen Bonus für die Prüfung am Ende des Semesters erarbeiten. Am Ende der Vorlesungszeit wird ein Termin angesetzt, an dem Sie Ihre Lösung kurz vorstellen, und die beste Lösung mit einem Preis prämiert wird.

In der Semesteraufgabe geht es um die Klassifikation des gegebenen Datensatzes. Der Datensatz besteht dabei aus verschiedenen Datentypen:

- kategorisch,
- kontinuierlich,
- u.v.m.

Ihre Aufgabe besteht darin ein Modell zu entwerfen, das die Daten den entsprechenden Klassen zuordnet. Sie können den kompletten Datensatz zum Training und Testen verwenden. Bei der Vorstellung Ihres Algorithmus werden wir Ihre Lösung mit neuen, unbekannten Daten testen.

a) Machen Sie sich, mit dem Datensatz vertraut.

b) Erarbeiten Sie einen Data-Mining-Prozess.

Speichern Sie Ihre Ergebnisse in einem numpy Array namens `result`. Dabei sollen die entsprechenden Klassenlabels im Array gespeichert werden. Weiterhin können Sie davon ausgehen, dass der übergebene Datensatz in `unknown_data` demselben Format wie der gegebene Datensatz unterliegt.

Beispiel:

```
import numpy as np
...
result = my_fancy_classifier(unknown_data)
# This should print something in the format of [0, 2, 1, 4,... 0, 3, 0, 1]
print(result)
```

Die Ergebnisse der einzelnen Gruppen werden mit Hilfe von Precision, Recall und dem F_1 -Measure miteinander verglichen.

c) Erklären Sie beim Präsentationstermin Ihren Lösungsansatz (max. 5 Minuten).