

Programmmentwurf – Maschinelles Lernen

Im Modul Maschinelles Lernen & Computational Intelligence

Prof. Dr. Dirk Reichardt

Aufgabe : Cycling Data - B

Es liegen aufgezeichnete Daten von einigen hundert Radfahrten eines Fahrers über mehrere Jahre vor. Teils sind diese unvollständig, da nicht immer alle Sensoren verfügbar waren. Auch sind Messungen manchmal fehlerbehaftet.

Die Daten enthalten Datum und Uhrzeit einer Fahrt, die Kilometer, Höhenmeter, Herzfrequenz, Trittfrequenz, Geschwindigkeit, die GPS Koordinaten der „Bounding Box“. Dazu werden Informationen darüber gespeichert, welches Rad genutzt wurde und weitere zum Fahrer selbst. Bitte betrachten Sie dazu den Ihnen zugeteilten individuellen Datensatz. Dieser kann bzgl. der Anzahl der Attribute abweichen.

Es sollen zwei Lernverfahren eingesetzt werden. Dazu ist es zulässig, eine Bibliotheksfunktion (z.B. aus sklearn) für das MLP zu verwenden, u.a. auch die in der Vorlesung vorgestellten Jupyter Notebooks oder der C/C++ Code. *Zugelassene Programmiersprachen: C/C++, Java, Python*

Folgende Aufgaben sind zu erfüllen:

- (A) Konfigurieren Sie
 - a. ein MLP und
 - b. ein Entscheidungsbaumum folgende Funktion zu erlernen:
Der benutzte Radtyp soll auf Basis der übrigen Daten bestimmt werden (mind. 50% der Attribute einbeziehen, Ausnahmen begründen)
- (B) Bereiten Sie dazu die Daten geeignet vor, um sie durch die Lernverfahren gut nutzbar zu machen.
- (C) Bewerten Sie das Lernergebnis bzgl. der diskutierten Qualitätskriterien und vergleichen Sie die beiden Modelle und verschiedene Konfigurationen zur Optimierung des Ergebnisses

Abgabe:

1. Programmcode zum Einlesen der Datei(en) sowie der Datenvorbereitung, des Lernverfahrens und der Ergebnisausgabe idealerweise als Jupyter Notebook.
2. Dokument (max.3-4 Seiten) in welchem die Konfiguration begründet wird, sowie die Analyse und ein Fazit des Lernergebnisses beschrieben und interpretiert werden. Dies kann in das Jupyter Notebook aus 1.) integriert werden.

Die Abgabe erfolgt über das Moodle System.

Die Bewertung setzt sich aus den folgenden Kriterien zusammen:

- Korrektheit und Strukturiertheit (incl. Kommentierung) des Programmcodes
- Plausibilität und Qualität des Lernergebnisses
- Angemessene Datenvorbereitung
- Passende und gut begründete Konfiguration und Aufgabenanpassung der Lernverfahren
- Darstellung und Interpretation der Lernergebnisse