

Special Engineering – IKT (M)

3. Projekt – Neuronale Netze

Für den Datensatz “Digits Dataset” soll ein Klassifizierungsmodell für handgeschriebene Ziffern mithilfe neuronaler Netzen erstellt werden. Das Modell soll ausschließlich mithilfe der Pakete `keras` und `sklearn` erstellt, trainiert und ausgewertet werden. Weiterhin können die Pakete `numpy`, `matplotlib` und `seaborn` verwendet werden.

Abzugeben sind eine Dokumentation und Programmcode des Modells und aller Auswertungen in Form

- eines PDFs und eines lauffähigen Python-Skripts **oder**
- eines lauffähigen Jupyter-Notebooks.

Das Projekt ist bis zum Abgabetermin des aktuellen Prüfungszeitraums auf Moodle einzureichen. Es gelten die allgemeinen Kriterien für eine schriftliche Arbeit. Achten Sie z.B. darauf, ganze Sätze zu formulieren und Ihre Entscheidungen zu begründen, bzw. zu belegen.

Im folgenden soll der Datensatz für handgeschriebene Ziffern mithilfe neuronaler Netze bearbeitet werden. Ziel ist die Vorhersage, um welche Ziffer (“0”, “1”, ..., “9”) es sich bei einem Bild handelt. Gehen Sie wie folgt vor:

1. Importieren Sie den Datensatz für handgeschriebene Ziffern mithilfe folgender Anweisungen.

```
from sklearn.datasets import load_digits
digits = load_digits()
```

und erstellen Sie einen geeigneten Zielvektor. Teilen Sie ihn in Trainings- (60%) Validierungs- (20%) und Testdaten (20%) ein.

2. Erstellen Sie ein geeignetes neuronales Netz und werten Sie die Genauigkeit auf den Validierungsdaten aus. Verwenden Sie die Early Stopping Strategie bei der Berechnung.
3. Optimieren Sie Ihr Neuronales Netz, z.B. hinsichtlich Standardisierung, Anzahl der Neuronen pro Schicht, Anzahl der Schichten, Batch Normalization und Dropout anhand der Genauigkeit auf den Validierungsdaten. Sie können z.B. die Klasse `sklearn.model_selection.ParameterGrid` verwenden.
4. Geben Sie für das optimale Modell die Genauigkeit, die Konfusionsmatrix und Kennzahlen für jede Klasse an. Diskutieren Sie Ihre Ergebnisse.
5. Untersuchen Sie Ihr Modell mithilfe einer Lernkurve. Diskutieren Sie, wie Sie Ihr Modell verbessern könnten.
6. Werten Sie das optimale Modell auf den Testdaten aus (Genauigkeit, Konfusionsmatrix, Kennzahlen für jede Ziffer, etc.). Visualisieren Sie falsch klassifizierte Ziffern. Diskutieren Sie Ihr Modell.