

## **Special Engineering – IKT (M)**

### **2. Projekt – Handgeschriebene Ziffern**

Für den Datensatz “Digits Dataset” soll ein Klassifizierungsmodell für handgeschriebene Ziffern mithilfe einer logistischen Regression erstellt werden. Das Modell soll ausschließlich mithilfe der Routinen aus den Übungsblättern 1 bis 6 trainiert und ausgewertet werden. Weiterhin können die Pakete `numpy` und `matplotlib` verwendet werden.

Abzugeben sind

1. ein Paket `logistische_regression.py` mit den benötigten Routinen des 1. bis 6. Übungsblattes
2. 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 der logistischen Regression bearbeitet werden. Ziel ist die Vorhersage, ob eine Ziffer eine “9” ist (Klasse 1, d.h. positive Klasse) oder nicht (Klasse 0, d.h. negative Klasse). 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 die Daten in Trainings- (80%) und Testdaten (20%) auf.

2. Visualisieren Sie verschiedene Ziffern mittels `plt.matshow` und beschreiben Sie Ihre Beobachtungen.
3. Erstellen Sie auf den Trainingsdaten eine logistische Regression.
4. Evaluieren Sie Ihr Modell auf den Trainings- und Testdaten. Geben sie die Genauigkeit aus, erstellen Sie eine Konfusionsmatrix und geben Sie geeignete Kennzahlen an. Visualisieren Sie falsch klassifizierte Ziffern des Testdatensatzes. Diskutieren Sie Ihre Ergebnisse.
5. Untersuchen Sie Ihr Modell mithilfe einer Lernkurve. Diskutieren Sie, wie Sie Ihr Modell verbessern könnten. Würde eine Regularisierung helfen?

**Bonusaufgabe (+20 Punkte):** Erweitern Sie die Erkennung mittels One-vs-Rest-Methode auf alle Ziffern und werten Sie Ihr Modell geeignet aus.