



Aufgabe 2: Zeitreihenanalyse mit RNN und Jena Climate Datensatz

In dieser Aufgabe sollen Sie ein grundlegendes Verständnis für Recurrent Neural Networks (RNNs), insbesondere LSTMs, erlangen und deren Anwendung auf den Jena Climate Datensatz erproben. Das Ziel ist es, ein RNN zu erstellen und zu trainieren, um Temperaturentwicklungen vorherzusagen, und dabei die verschiedenen Schritte der Datenvorbereitung, Modellbildung und Evaluierung zu durchlaufen.

1. Konzeption und Mathematik eines RNNs

- Konzept eines Recurrent Neural Networks (RNN) und speziell eines LSTMs erläutern: Wie funktionieren Rückkopplungen, Gedächtniszellen und Gates?
- Grundlegende Mathematik hinter LSTMs erklären (z. B. Zellenzustand, Input-, Forget- und Output-Gates).

2. Datenanalyse und Vorbereitung

- Laden des Jena Climate Datensatzes für den Zeitraum **01.01.2014 bis 31.12.2016** (WS Beutenberg): <https://www.bgc-jena.mpg.de/wetter/>.
- Daten sinnvoll reduzieren, um die Bearbeitbarkeit sicherzustellen:
 - **Reduktion der Messfrequenz:** Zum Beispiel von 10-Minuten-Daten auf stündliche oder tägliche Durchschnittswerte.
- Explorative Datenanalyse (EDA): Visualisierung der Temperatur- und anderen relevanten Zeitreihen.
- Aufteilung der Daten in Trainings-, Validierungs- und Testdatensatz.

3. Modellaufbau und Training

- Erstellung und Training eines einfachen LSTMs zur Temperaturvorhersage mit einer Library bzw. einem Framework ihrer Wahl (z. B. TensorFlow, PyTorch, Keras).
- Festlegung der Netzwerkarchitektur: Anzahl der LSTM-Zellen, Anzahl der Schichten, Dropout etc.
- Das Modell für eine geeignete Anzahl an Epochen trainieren und die Entwicklung der Loss- und Accuracy-Werte überwachen.

4. Evaluierung des Modells

- Evaluierung der Modellleistung auf dem Testdatensatz (z. B. durch Berechnung des Mean Squared Error, MSE).
- Visualisierung der tatsächlichen vs. vorhergesagten Temperaturen zur besseren Beurteilung der Modellgüte.

5. Hyperparameter-Tuning

- Variation von Hyperparametern (z. B. Anzahl der LSTM-Zellen, Lernrate, Batch-Größe), um die Modellleistung zu verbessern.
- Dokumentation der Ergebnisse und Vergleich der Modelle nach Anpassungen.

6. Visualisierung der Ergebnisse

- Darstellung der Trainings- und Validierungsverluste über die Trainingsperioden.
- Visualisierung der Temperaturvorhersagen im Vergleich zu den tatsächlichen Werten (z. B. als Linienplot).
- Analyse von Perioden mit hohen Vorhersagefehlern und Diskussion möglicher Gründe.