

# KI Praktikum 2: Neuronales Netz

## Aufgabenstellung

Modellierung und Training eines neuronalen Netzes für den „Forest Cover-Type“-Datensatz, wobei verschiedene Netzarchitekturen ausprobiert werden sollen, um eine möglichst optimale Netzarchitektur zu finden.

## Der Forest Cover-Type Datensatz

Die Stichproben in diesem Datensatz entsprechen 30×30 m großen Waldstücken in den USA, die für die Vorhersage des Bewuchses jedes Stücks, d. h. der vorherrschenden Baumart, gesammelt wurden. Details s.:

<http://archive.ics.uci.edu/ml/datasets/Covertime>

## Trainings-, Validierungs- und Testdaten

Mit der Datei „train.csv“ ist Ihnen ein Trainingsdatensatz gegeben. Mit diesem Datensatz trainieren Sie das neuronale Netz. Sobald Sie ein geeignetes Modell trainiert haben, werden Sie auf den Trainingsdaten gute Ergebnisse erhalten. Ob Sie allerdings auch für neue Daten gute Ergebnisse mit dem Modell erzielen, d.h. später, wenn das Modell in einer Anwendung genutzt wird, wissen Sie nicht. Um das herauszufinden, gibt es Testdaten. Mit den Testdaten testen Sie ganz am Ende das trainierte Modell. Da die Testdaten auf keinen Fall während des Trainings des Modells genutzt werden dürfen, erhalten Sie erst ein paar Tage vor der Praktikumsabgabe den Testdatensatz.

Da Sie bei der Suche nach der optimalen Netzarchitektur das neuronale Netz auf ungesehenen Daten „testen“ müssen, müssen Sie den erhaltenen Trainingsdatensatz in einen Trainings- und Validierungsdatensatz trennen. Mit dem neuen Trainingsdatensatz trainieren Sie das Modell und mit dem Validierungsdatensatz „testen“ Sie das Modell, man spricht von dem Validieren des Modells.

Das Modell, das auf den Validierungsdaten das beste Ergebnis erzielt, ist Ihr optimales Modell. Dieses sollten Sie ganz am Ende auch auf den Testdaten testen.

## Suche nach der optimalen Netzarchitektur

Die optimale Netzarchitektur können Sie mit der Python-Bibliothek optuna finden, s. die Datei „breast\_cancer\_template.ipynb“.

## Hardware für das Training

Sie können das Praktikum komplett auf Ihrem Rechner ausführen (Installation: s. „Environment\_mit\_Anaconda\_aufsetzen.docx“ und „environment.yml“). Falls Sie merken, dass Ihr Rechner die benötigte Rechenleistung nicht mitbringt, dann gibt es die folgenden zwei Alternativen:

1. Server der Hochschule: s. Link in ilias: „Web-Server für Jupyter Notebooks“.

2. Google Colab: Wenn Sie einen Google Account haben oder kostenlos erstellen möchten, können Sie Google Colab kostenlos nutzen. Details s.: <https://colab.research.google.com/>

## Aufgabenstellung im Detail

1. Schauen Sie sich das Notebook „data exploration basics for students.ipynb“ an, um eine Idee von dem Forest Cover-Type Datensatz zu erhalten und schauen Sie sich die oben verlinkte Webseite zu dem Datensatz an.
2. Modellieren und trainieren Sie ein bestmögliches neuronales Netz für den Datensatz. Schauen Sie sich dazu das Beispiel-Notebook „breast\_cancer\_template.ipynb“ an (gerne auch: „digits\_template.ipynb“) und erstellen Sie selbst ein ähnliches Notebook.
3. Sobald wir Ihnen den Testdatensatz geben, testen Sie Ihr bestes Modell an dem Testdatensatz.
4. Erstellen Sie eine kurze Präsentation (10 Min.), die Ihr finales Modell zeigt, Ihre Ergebnisse auf Trainings-, Validierungs- und Testdaten, Ihren Weg zum Ziel und Ihre Interpretation der Ergebnisse, bzw. Ihre Erkenntnisse, die Sie im Praktikum gewonnen haben.