

## Mini-Challenge DEL: Hyperparameter Tuning und Model Evaluation

### Ziel

Implementiere, trainiere und validiere mehrere Neuronale Netzwerk Modelle für einen Klassifikationstask mit geeignetem Dataset. Lerne dabei einige Techniken des Deep Learning kennen.

### Vorgehen

Der zu modellierende Task und die dazu passenden Daten sollen am Anfang der Mini-Challenge individuell vereinbart werden. Dabei kannst Du auch Deine Wünsche einbringen und einen Task und Datensatz vorschlagen.

Dazu erstellst Du ein Jupyter Notebook, welches Du in zwei Etappen abgibst:

1. Erste Etappe: Abgabe mit dem Basismodell und evaluierter Lernrate und Batchsize (nach «Schritt 2» unten);
2. Zweite Etappe: Finale Abgabe nach allen Schritten.

### Erste Etappe: Basismodell

#### Auswahl Task / Datensatz

1. Mache Dir Gedanken, welcher Task gelernt werden soll und mit welchen Daten Du arbeiten möchtest.  
*Bemerkung:* Bei der Auswahl des Datensatz ist darauf zu achten, dass der Fokus in der MC auf das Training und die Evaluation gelegt werden kann, und nicht zu viel Zeit mit Datenbeschaffung, -Analyse und -Vorverarbeitung verbraucht wird. Der Datensatz sollte auch nicht zu gross sein, um viele verschiedene Modell- und Parametervarianten evaluieren zu können.
2. Spezifiziere einen Vorschlag für ein Modell, welches Du als Basismodell nehmen möchtest.  
*Bemerkung:* Als Basis-Modell sollte eine Architektur mit ein paar Convolutional Layern für die Feature-Extraktion und wenigen Fully Connected Layern für die Klassifikation verwendet werden. Davon ausgehend, werden dann verschiedene Varianten untersucht (siehe unten).

**Absprache/Beschluss mit Fachcoach über Daten, Task, Basismodell.**

#### Daten Kennenlernen, Aufbau Modellierung, Basismodell

1. Vorbereitung der Daten:
  - a. (kurze) explorative Analyse der Features (z.B. Vergleich der Klassen pro Feature, Balanciertheit der Klassen).
  - b. geeignetes Preprocessing (z.B. Normalisierung der Daten)
2. Festlegen der Metriken für die Evaluation.
3. Implementiere Basisfunktionalität:
  - a. Fürs Trainieren und Evaluieren von Modellen.
  - b. Verwende MLOps Plattform (z.B. [W&B](#))
  - c. Implementiere das vereinbarte Basismodell.

- d. Overfitting Test: Trainiere das Basismodell mit einem einzigen Sample oder einem Batch, um die Basisfunktionalität zu testen.
4. Training des Basismodells:
  - a. Training mit SGD (ohne Momentum), ohne Regularisierung, ohne Batchnorm unter Verwendung von allen Trainingsdaten;
  - b. Tuning der Lernrate und der Batchsize. Zeige die Unterschiede anhand der Lernkurven mit der gewählten Kostenfunktion und den Metriken auf. Diskussion der Ergebnisse und Vergleich mit der Erwartung aus der Theorie. *Bemerkung:* Keine Verfahren zur automatischen Hyperparameter-Suche (z.B. keine Bayes'sche Optimierung) verwenden!
  - c. Schätzung des statistischen Fehlers dieser Metriken (mit cross-validation, zumindest für ein Training durchgeführt).

#### Abstimmung mit Fachcoach:

- **Abgabe** (erste Etappe):
  - Zeige das Notebook dem Fachcoach indem Du ihn durchs Notebook führst.
  - Share mit ihm das Notebook (eingefrorene git-Version).
- **Absprache mit Fachcoach:**
  - Der Fachcoach gibt Dir Feedback zur ersten Abgabe, womit Du eine verbesserte Version erstellen sollst.
  - Fahre fort mit der zweiten Etappe.

### Zweite Etappe: Hyperparameter Tuning

Im Folgenden sollst Du nun den Trainingsverlauf und die Modellperformance unter Änderung verschiedener Hyperparameter (siehe unten) untersuchen. Hierzu sollst Du mind. 10 Hypothesen formulieren, Experimente durchführen, um die Hypothesen zu testen, und die Ergebnisse einordnen und reflektieren.

Pro Hypothese gehe etwa wie folgt vor:

- a) *Hypothese*: Formuliere deine Erwartung dazu, was sich in der Modellperformance und/oder Trainingsverlauf ändert, wenn ein Hyperparameter verändert wird. Versuche diese Hypothese zu begründen mit Hilfe Deines (allenfalls zu erarbeitenden) theoretischen Wissens. Gib eine entsprechende Referenz an.
- b) *Experiment*: Definiere nun ein Experiment, mit welchem die Hypothese getestet werden kann. Führe das Experiment durch.
- c) *Analyse/Reflexion*: Analysiere die Ergebnisse daraufhin, ob die Hypothese sich bestätigt. Stelle eine Reflexion an.

Die Hypothesen sollen aus **folgenden Themengruppen** gebildet werden:

- **Modellkomplexität:**
  - Anzahl Conv Layer und/oder Anzahl Lineare Layer (Modelltiefe)
  - Anzahl Filter pro Conv-Layer oder Anzahl Units in Linearen Layers (Breite)Allenfalls muss die Lernrate nochmals angepasst werden.
- **Settings bei Conv Layer:**
  - Kernel Size
  - Pooling

- Regularisierung:
  - Dropout
  - Data Augmentation
- Verwendung von BatchNorm
- Initialisierung der Gewichte
- Verwendung von Adam Optimizer
- Transfer Learning

Am Schluss soll auch ein Fazit darüber gezogen werden, welche Varianten/Methoden zu einem besser funktionierenden Modell beitragen.

Erstelle dann «Dein bestes Modell» aus einer geeigneten Kombination der obenstehenden Methoden.

**Absprache/Beschluss** mit Fachcoach: Zeige dem Fachcoach die erste formulierte Hypothese und das hierzu durchgeführte Experiment. Diskutiere die weiteren Hypothesen nach Bedarf.

## Präsentation, Bericht

1. **Präsentation** (~10m): Kurze Präsentation (Slides, nicht Notebook) mit Diskussion der wichtigsten Ergebnisse.  
**Q&A** (~10min): Verständnisfragen zu den untersuchten Methoden (SGD, Parameter Tuning, Regularisierung, Batchnorm, Optimizers, Transfer Learning).
2. **Bericht** in Form **eines (!)** gut dokumentierten, übersichtlichen Jupyter Notebooks (typischerweise das von der ersten Etappe erweiterten Notebook). Dieses soll dem Fachexperten erlauben, die Schritte nachzuvollziehen. Die Outputs (z.B. von W&B) sollten im Notebook integriert sein, so dass die Trainings nicht alle nochmals durch den Fachcoach durchgeführt werden müssen.

## Zeitlicher Rahmen:

Wird am Anfang verbindlich festgelegt.

## Beurteilung

Beurteilt wird auf Basis des abgegebenen Notebooks und der Präsentation inkl. Q&A:

- Vollständige und korrekte Umsetzung der vereinbarten Aufgabestellung.
- Klare, gut-strukturierte Umsetzung.
- **Schlüssige Beschreibung und Interpretation der Ergebnisse.** Gut gewählte und gut kommentierten Plots und Tabellen.
- Vernünftiger Umgang mit (Computing-)Ressourcen.
- Verständliche Präsentation der Ergebnisse, korrekte Antworten auf die Fragen.