



Checklist Training von Neuronalen Netzen

1. Datenvorbereitung

- **Unzureichende Datenqualität:** Daten enthalten Rauschen, fehlende Werte oder Outlier, die nicht behandelt wurden.
- **Unbalancierte Datensätze:** Eine Klasse ist stark überrepräsentiert (Class Imbalance).
- **Falsche Skalierung/Normalisierung:** Merkmale sind nicht ordnungsgemäß skaliert (Standardscaler verwenden).
- **Falsche Datenaufteilung:** Trainings-, Validierungs- und Testdaten wurden nicht korrekt getrennt (z. B. Leckage zwischen Sets bei Zeitreihendaten).
- **Inkonsistente Labels:** Labels sind fehlerhaft oder ungenau.

2. Modellarchitektur

- **Falsche Modellgröße:** Modell ist zu klein (Underfitting) oder zu groß (Overfitting).
- **Unpassende Aktivierungsfunktionen:** Auswahl von Aktivierungsfunktionen, die nicht zur Aufgabe passen (z. B. sigmoid statt ReLU in tiefen Netzwerken).
- **Fehlende Regularisierung:** Dropout, L1/L2-Regularisierung oder Batch-Normalisierung wurden nicht eingesetzt.

3. Hyperparameter

- **Falsche Lernrate:** Lernrate ist zu hoch (instabile Konvergenz) oder zu niedrig (langsames Lernen).
- **Ungeeigneter Optimierer:** Verwendung eines nicht optimalen Optimierungsalgorithmus (z. B. SGD statt Adam für komplexe Modelle).
- **Batchgröße:** Zu kleine oder zu große Batchgrößen führen zu ineffizientem Training.
- **Fehlerhafte Epochenzahl:** Zu viele oder zu wenige Epochen führen zu Overfitting oder Underfitting.

4. Trainingsprozess

- **Fehlende Kontrolle über den Verlust:** Verlustfunktion passt nicht zur Aufgabe (z. B. Mean Squared Error für Klassifikationsprobleme).
- **Mangelhafte Überwachung:** Keine Verwendung von Validierungsdaten, um den Trainingsfortschritt zu überwachen.
- **Fehlender Early Stopping:** Training läuft weiter, obwohl keine Verbesserung mehr stattfindet.
- **Gradientenexplosion/-verschwinden:** Keine Mechanismen zur Kontrolle von Gradientenproblemen (z. B. durch Gradienten-Clipping).

5. Evaluierung und Testen

- **Overfitting ignorieren:** Hohe Leistung auf Trainingsdaten, aber schlechte Generalisierung auf Validierungs-/Testdaten.
- **Unrealistische Metriken:** Verwendung von Metriken, die die Leistung nicht sinnvoll bewerten (z. B. Accuracy bei unbalancierten Daten).
- **Testdaten-Leckage:** Testdaten wurden im Training verwendet, wodurch die Ergebnisse verfälscht werden.