

Mini-Challenge DEL: Hyperparameter Tuning und Model Evaluation

Ziel

Implementiere, trainiere und validiere mehrere Neuronale Netzwerk-Modelle für einen Klassifikationstask mit geeignetem Dataset. 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. Dabei ist aber darauf zu achten, dass der Fokus auf das Training und die Evaluation gelegt wird, 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. Als Basis-Modell sollte eine Architektur mit ein paar (wenigen) Convolutional Layern für die Feature-Extraktion und wenige Dense Layer für die Klassifikation verwendet werden. Davon ausgehend, sollen dann verschiedene Varianten untersucht (siehe unten). Bei einer Variante soll auch ein grösseres vortrainiertes Modell verwendet werden, welches dann fined-tuned werden soll.

Arbeitsschritte

Schritt 1: Auswahl Task / Datensatz

1. Mache Dir Gedanken, welchen Task das Modell lernen soll und mit welchen Daten Du arbeiten möchtest.
2. Diskutiere die Idee mit dem Fachcoach.

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

Schritt 2: Daten Kennenlernen

1. Mache Dich mit dem Datensatz vertraut, indem Du eine (kurze) explorative Analyse der Features durchführst: z.B. Vergleich der Klassen pro Feature, Balanciertheit der Klassen.
2. Führe ein geeignetes Preprocessing durch, z.B. Normalisierung der Daten.

Wie erwähnt, sollte dieser Schritt nicht viel Zeit beanspruchen, da der Datensatz einfach sein soll

Schritt 3: Aufbau Modellierung, Basismodell

1. Implementiere ein Basismodell.
2. Lege fest, wie (mit welchen Metriken) Du Modelle evaluieren möchtest.
3. Implementiere Basisfunktionalität, um Modelle zu trainieren und gegeneinander zu evaluieren. Mache Dich vertraut mit einer MLOps Plattform (z.B. [W&B](#)), welche für die Evaluation verwendet werden soll.
4. Teste das Basismodell und die Trainingsfunktionalität, indem Du nur mit einem einzigen Sample oder einem Batch trainierst. Damit bekommst Du zwar Overfitting, aber auch einen guten Test, der zeigt, dass Information aus den Daten aufgenommen werden kann.
5. Trainiere nun das Basismodell mit SGD, ohne Regularisierung, ohne Batchnorm mit allen Trainingsdaten. Wähle eine Lernrate und eine Batchgrösse, so dass das Training gut funktioniert und zeige das anhand der Lernkurven für die gewählte

Kostenfunktion und die Metriken (immer «train» und «validation» Partition betrachten).

6. Führe Cross-Validation durch, um den statistischen Fehler zu schätzen.

Beachte: In dieser MC sollen **keine Verfahren zur automatischen Hyperparameter-Suche** (z.B. keine Bayes'sche Optimierung) verwendet werden!

Absprache/Beschluss mit Fachcoach: Zeige das erreichte dem Fachcoach. Besprich mit ihm die zu untersuchenden Varianten (Schritte 4).

Schritt 4: 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.

Mindestens eine Hypothese muss aus **jeder der folgenden Themengruppen** stammen:

- Wahl der Lernrate und/oder Batchsize (mit SGD)
- Initialisierung der Gewichte
- 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, Verwendung von MaxPooling
- Verwendung von Dropout
- Verwendung von Data Augmentation
- Verwendung von BatchNorm
- Verwendung von Adam Optimizer

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.

Schritt 5: Transfer Learning

Untersuche schliesslich Transfer Learning: Wähle ein grösseres Modell, welches Dir für die gegebene Problemstellung sinnvoll erscheint und für welches vortrainierte Gewichte verfügbar sind. Führe dann eine sorgfältiges Fine-Tuning durch. Vergleiche die Performance mit dem bisher Erreichten.

Schritt 6: 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. 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 beim Schritt 1 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.