



Glossar

Glossar I

- **Accuracy:** Misst den Anteil korrekt klassifizierter Beispiele, sinnvoll bei ausbalancierten Datensätzen.
- **Aktivierungsfunktionen:** Entscheiden, welche Signale weitergeleitet werden, um nichtlineare Beziehungen abzubilden. Beispiele sind Sigmoid, ReLU und Tanh.
- **AUC (Area Under Curve):** Fasst die Modellgüte in einer einzelnen Kennzahl zusammen, basierend auf der ROC-Kurve.
- **Autoscaling:** Passt die Anzahl der laufenden Instanzen automatisch basierend auf der aktuellen Nachfrage an.
- **Back-Translation:** Übersetzt einen Satz zunächst in eine andere Sprache und anschließend zurück in die Ausgangssprache, um neue Varianten zu generieren.
- **Backpropagation:** Verfahren, bei dem der am Ausgang gemessene Fehler rückwärts durch das Netzwerk propagiert wird, um die Gewichte iterativ zu korrigieren.
- **Bag-of-Words (BoW):** Zählt die Häufigkeit von Wörtern ohne Rücksicht auf Reihenfolge oder Kontext.
- **Batch Normalization:** Normalisiert die Aktivierungen jeder Schicht, um Schwankungen zu reduzieren und das Training zu beschleunigen.
- **Batch Size:** Bestimmt die Anzahl der Trainingsbeispiele, die in einem Schritt zur Aktualisierung der Modellparameter verwendet werden.
- **Bayesian Optimization:** Nutzt Modellierungen des Parameterraums, um gezielt Bereiche mit hohem Potenzial für Hyperparameter zu erkunden.

Glossar II

- **BERT:** Bidirektionales Transformer-Modell, das den Kontext von links und rechts eines Tokens berücksichtigt, um tiefe semantische Zusammenhänge zu lernen.
- **BLEU (Bilingual Evaluation Understudy):** Bewertet maschinelle Übersetzungen, indem es die Ähnlichkeit der generierten Sätze mit Referenzübersetzungen quantifiziert.
- **Caching:** Speichert bereits verarbeitete Daten im Arbeitsspeicher, um erneute Berechnungen zu vermeiden.
- **Canary Deployments:** Führen neue Modellversionen schrittweise ein, um das Risiko von Fehlern zu minimieren.
- **CIDEr (Consensus-based Image Description Evaluation):** Misst die Übereinstimmung zwischen generierten und Referenzbeschreibungen, besonders relevant für Bildbeschreibungsaufgaben.
- **CNNs (Convolutional Neural Networks):** Extrahieren lokale Muster (Kanten, Texturen) aus Bildern und finden Anwendung in der Bildklassifikation, Objekterkennung.
- **Cross Entropy:** Ein gängiges Maß für Klassifikationsfehler, das Wahrscheinlichkeitsverteilungen betrachtet.
- **Cross-Attention:** Schichten im Decoder beziehen Informationen aus dem Encoder-Ausgaberaum ein, um den generierten Output auf den Input-Kontext abzustimmen.
- **Cyclical Learning Rates:** Periodische Schwankungen der Lernrate helfen, aus lokalen Minima herauszufinden und die Konvergenz zu verbessern.

Glossar III

- **Data Parallelism:** Aufteilen von Batches auf mehrere GPUs, um das Training zu beschleunigen.
- **Datenaugmentation:** Erweitert den Trainingsdatensatz künstlich, um die Datenvielfalt und Robustheit des Modells zu erhöhen.
- **Dependency Parsing:** Identifiziert Abhängigkeitsbeziehungen zwischen Wörtern, etwa welches Wort Subjekt oder Objekt eines Verbs ist.
- **Dropout:** Schaltet Neuronen zufällig aus, um Overfitting zu vermeiden und die Robustheit des Modells zu erhöhen.
- **Early Stopping:** Beendet das Training bevor sich das Modell zu stark an die Trainingsdaten anpasst, um Overfitting zu vermeiden.
- **Elasticsearch:** Verwendet invertierte Indizes für schnelle Volltextsuche.
- **Encoder-Struktur:** Besteht aus gestapelten Attention- und Feed-Forward-Schichten, die Eingabe-Embeddings in kontextreiche Repräsentationen transformieren.
- **Epoche:** Ein vollständiger Durchlauf des gesamten Trainingsdatensatzes.
- **Exploding Gradients:** Der Gradient steigt rasant an, was zu instabilen Gewichtsupdates führt.
- **Extraktive Zusammenfassung:** Wählt wichtige Sätze aus dem Originaltext aus, um eine verkürzte Version zu erstellen.

Glossar IV

- **F1-Score:** Harmonisches Mittel aus Precision und Recall, besonders nützlich bei unausgewogenen Datensätzen.
- **FastAPI:** Ein Python-Framework zur Erstellung von RESTful APIs.
- **Fine-Tuning:** Feinabstimmung eines vortrainierten Modells auf eine neue Aufgabe.
- **FP16 (16-Bit Floating Point):** Verwendet 16-Bit-Gleitkommazahlen, was das Training beschleunigt, aber die numerische Präzision verringert.
- **FP32 (32-Bit Floating Point):** Verwendet 32-Bit-Gleitkommazahlen für genauere Berechnungen, ist aber ressourcenintensiver.
- **GANs (Generative Adversarial Networks):** Bestehen aus einem Generator, der künstliche Daten erzeugt, und einem Diskriminator, der zwischen echten und künstlichen Daten unterscheidet.
- **Generative Zusammenfassung:** Erstellt neue, komprimierte Textfassungen, die den Inhalt des Originals wiedergeben.
- **Glorot-Initialisierung (Xavier):** Sorgt für ausgeglichene Varianzverteilungen in den Schichten.
- **Gradientenabstieg:** Verfahren, bei dem die Gewichte in Richtung des abnehmenden Fehlers angepasst werden.
- **Grid Search:** Durchsucht systematisch den Parameterraum für Hyperparameter-Tuning.

Glossar V

- **GRU (Gated Recurrent Unit):** Eine erweiterte RNN-Variante mit internen Gating-Mechanismen zur besseren Erfassung von Langzeitabhängigkeiten.
- **GPT (Generative Pre-trained Transformer):** Autoregressives Modell für Textgenerierung, das das nächste Wort auf Basis der bisherigen Tokens vorhersagt.
- **Haystack:** Ein Open-Source-Framework zur Entwicklung von Fragebeantwortungssystemen.
- **He-Initialisierung:** Speziell für ReLU-Schichten konzipiert und unterstützt schnelles, stabiles Training.
- **Hidden-Layer:** Extrahieren schrittweise abstrakte Merkmale aus den Eingangsdaten.
- **Hugging Face Transformers:** Bietet eine breite Auswahl an vortrainierten Modellen und Tools für NLP.
- **Input-Layer:** Nimmt die Rohdaten auf und gibt sie an die nächsten Schichten weiter.
- **Iteration:** Einzelner Schritt der Gewichtsaktualisierung pro Batch.
- **L1-Regularisierung:** Fügt Strafen für große Gewichte hinzu, um Overfitting zu vermeiden.
- **L2-Regularisierung:** Fügt Strafen für große Gewichte hinzu, um Overfitting zu vermeiden.
- **Layer Normalization:** Normalisiert Eingaben jeder Schicht, stabilisiert Training und begünstigt Konvergenz.

Glossar VI

- **Learning Rate:** Bestimmt die Größe der Gewichtsänderungen pro Schritt während des Trainings.
- **Learning Rate Schedules:** Steuern die Anpassung der Lernrate während des Trainings.
- **Lemmatisierung:** Reduziert Wörter auf ihre Grundform (Lemma).
- **Load Balancer:** Verteilen den eingehenden Traffic gleichmäßig auf mehrere Server, um Überlastungen zu verhindern.
- **Loss-Funktion:** Bestimmt über ihren Gradienten die Richtung des Lernschrittes. Beispiele sind MSE und Cross Entropy.
- **LSTM (Long Short-Term Memory):** Eine erweiterte RNN-Variante, die durch interne Gating-Mechanismen Langzeitabhängigkeiten besser erfasst.
- **MAE (Mean Absolute Error):** Betrachtet den absoluten Fehler und ist robuster gegenüber Ausreißern, oft bei Regressionen gebräuchlich.
- **Masked Language Modeling:** Maskiert Tokens im Input, um das Modell zu zwingen, kontextuelle Hinweise intensiver zu nutzen.
- **Masking:** Markiert gepaddete Bereiche, sodass das Modell diese ignorieren kann.
- **METEOR (Metric for Evaluation of Translation with Explicit Ordering):** Berücksichtigt Wortstämme, Synonyme und Wortreihenfolge stärker als BLEU.
- **MLflow:** Bietet ein zentrales UI, um Experimente, Parameter, Metriken und Artefakte zu verwalten.
- **MLOps:** Bezeichnet die Praktiken zur Automatisierung und zum Management von Machine-Learning-Modellen.

Glossar VII

- **Model Parallelism:** Zerlegung großer Modelle auf verschiedene GPUs oder Maschinen, um den Speicherbedarf je Gerät zu reduzieren.
- **MSE (Mean Squared Error):** Wird häufig bei Regressionsaufgaben als Fehlermaß eingesetzt.
- **Multi-Task Learning:** Trainiert mehrere verwandte Aufgaben gleichzeitig, um gemeinsame Strukturen besser auszunutzen.
- **MUSE:** Mehrsprachige Embeddings, die Wörter aus verschiedenen Sprachen in einen gemeinsamen Vektorraum projizieren.
- **N-Gramme:** Bilden Folgeeinheiten von mehreren Wörtern und berücksichtigen damit rudimentär die Wortreihenfolge.
- **Named Entity Recognition (NER):** Automatisierte Extraktion und Klassifizierung von benannten Entitäten wie Personen, Orten und Organisationen.
- **Next-Token Prediction:** Modell sagt das nächste Wort auf Basis des bisherigen Kontextes vorher.
- **NLTK:** Ein umfangreiches Tool für NLP, stark im akademischen Kontext.
- **Output-Layer:** Liefert die finale Vorhersage, beispielsweise eine Klasse oder einen numerischen Wert.
- **Overfitting:** Modell passt sich zu stark an die Trainingsdaten an und verfehlt die Generalisierung.
- **Padding:** Fügt künstliche Platzhalter hinzu, um alle Sequenzen auf eine einheitliche Länge zu bringen.

Glossar VIII

- **Perceptron:** Ein einfaches, lineares Modell für binäre Klassifikationsaufgaben.
- **Perplexity:** Misst, wie gut ein Sprachmodell Wahrscheinlichkeitsverteilungen von Wortfolgen abbildet.
- **Pooling Layer:** Reduziert die räumlichen Dimensionen und fasst Informationen zusammen.
- **POS-Tagging (Part-of-Speech-Tagging):** Weist jedem Wort seine Wortart zu (Verb, Nomen, Adjektiv usw.).
- **Positional Encoding:** Kodiert relative Wortpositionen, um den Transformer ein Verständnis für Wortreihenfolgen zu vermitteln.
- **Precision:** Definiert den Anteil der als positiv vorhergesagten Beispiele, die tatsächlich positiv sind.
- **Prefetching:** Lädt Daten im Hintergrund, während das Modell trainiert, um Leerlaufzeiten der GPU zu minimieren.
- **Prompts:** Dienen als Eingabeaufforderungen, die das Verhalten und die Antworten des Modells steuern.
- **Random Search:** Probiert zufällig gewählte Parameterkombinationen für Hyperparameter-Tuning.
- **Recall:** Misst, wie viele der tatsächlich positiven Beispiele korrekt erkannt werden.
- **ReLU (Rectified Linear Unit):** Eine Aktivierungsfunktion, die einfach ist, zum Teil das Verschwinden von Gradienten verhindert und sehr verbreitet ist.
- **Residual Connections:** Leiten Informationen um tiefe Schichten herum und ermöglichen so stabileres Training.

Glossar IX

- **Retriever-Reader Workflow:** Teilt das Fragebeantwortungssystem in zwei Komponenten: den Retriever zur Dokumentsuche und den Reader zur Antwortextraktion.
- **ROC-Kurve:** Illustriert das Verhältnis zwischen True Positive Rate und False Positive Rate.
- **ROUGE (Recall-Oriented Understudy for Gisting Evaluation):** Fokussiert auf den N-Gramm-Overlap zwischen maschineller Zusammenfassung und Referenz.
- **RNNs (Recurrent Neural Networks):** Verarbeiten Sequenzen schrittweise, indem sie Informationen über zeitliche Zusammenhänge in einem internen Zustand speichern.
- **Self-Attention:** Ermöglicht, alle Tokens einer Sequenz parallel zu betrachten und relevante Zusammenhänge direkt zu gewichten.
- **seq2seq-Transformer:** Nutzen ein Encoder-Decoder-Design, um Eingabesequenzen in Ausgabesequenzen zu transformieren.
- **Sigmoid:** Wandelt Eingaben in Werte zwischen 0 und 1 um, ideal für Wahrscheinlichkeitsinterpretationen.
- **SpaCy:** Bietet schnelle, produktionsreife Pipelines für Tokenisierung, POS-Tagging und Parsing.
- **Stemming:** Verkürzt Wörter auf ihre Wortstämme, ohne auf grammatische Korrektheit zu achten.
- **Step Decay:** Die Lernrate wird nach festgelegten Epochen reduziert.
- **Stanza:** Ein NLP-Tool, das moderne, neuronale Modelle nutzt und mehrere Sprachen abdeckt.

Glossar X

- **Statische Wort-Embeddings:** Projizieren jedes Wort in einen kontinuierlichen Vektorraum, in dem semantisch ähnliche Wörter nah beieinander liegen. Beispiele sind Word2Vec und GloVe.
- **Synonym Replacement:** Ersetzt ausgewählte Wörter durch ihre Synonyme.
- **Tanh:** Gibt Werte zwischen -1 und 1 aus, ist um 0 zentriert und oft leistungsfähiger als Sigmoid.
- **Tensor Cores:** NVIDIA Tensor Cores sind für schnelle Matrixmultiplikationen optimiert und profitieren stark von Mixed Precision.
- **TensorBoard:** Tool zur Visualisierung von Loss- und Metrikverläufen.
- **TF-IDF (Term Frequency - Inverse Document Frequency):** Gewichtet seltene, aber aussagekräftige Wörter höher.
- **Tokenisierung:** Die Aufspaltung von Texten in Tokens (Wörter oder Subwörter).
- **Transformer-Modelle:** Setzen auf Self-Attention und umgehen die Nachteile von RNNs.
- **Transfer Learning:** Nutzt bereits erlernte Merkmalsrepräsentationen aus einem vortrainierten Modell.
- **Underfitting:** Das Modell ist zu einfach und erkennt die zugrundeliegenden Muster nicht ausreichend.
- **Validierungsmetriken:** Zeigen, wie gut das Modell auf unbekannte Daten generalisiert.
- **Vanishing Gradients:** Der berechnete Gradient nimmt mit zunehmender Schichttiefe stark ab.

Glossar XI

- **Warmup:** Zu Beginn wird die Lernrate langsam angehoben.
- **Weights & Biases (W&B):** Eine cloudbasierte Lösung für das Tracken und Visualisieren von Experimenten.
- **WordPiece:** Ein Subwortverfahren, das seltene oder komplexe Wörter in häufig vorkommende Bestandteile zerlegt.
- **Word2Vec:** Ein Modell zur Erzeugung statischer Wort-Embeddings.
- **XLNet:** Mehrsprachige Embeddings, die einen direkten Vergleich von Bedeutungen zwischen Sprachen ermöglichen.