

HAW HAMBURG
INFORMATIK MASTER

GRUNDPROJEKT

TensorFlow Probability

Evaluation der Bibliothek für
probabilistische und statistische Analysen

Bearbeiter:

Tom Schöner (2182801)

Betreuung:

Prof. Dr. Olaf Zukunft

9. April 2019

Inhaltsverzeichnis

1	Abstract	3
2	Tensorflow Probability Komponenten	3
2.1	Layers	3
2.1.1	Layer 0: Tensorflow	4
2.1.2	Layer 1: Statistical Building Blocks	4
2.1.3	Layer 2: Model Building	4
2.1.4	Layer 3: Probabilistic Inference	4
3	Pragmatik und Semantik	5
4	Integration in Tensorflow	5
5	Beispiel: Korrelation von Luftverschmutzung und Temperatur	5
6	Fazit	5

Abbildungsverzeichnis

1	$X \sim \mathcal{N}(0, 1)$ mit 2500 Samples	5
---	---------------------------------------------	---

Tabellenverzeichnis

1 Abstract

Die auf Tensorflow basierende Bibliothek Tensorflow Probability¹ - fortan mit *TFP* abgekürzt - erweitert das Framework um eine probabilistische Komponente. Mittels einer breiten Masse an bereitgestellten Tools, wie statistischen Verteilungen, Sampling oder verschiedenster probabilistischer Erweiterungen für neuronale Netze, können einfache bis hin zu komplexen Modellen erstellt werden. Berechnungen werden, wie man es aus Tensorflow gewohnt ist, durch *Dataflow Graphs*² abgebildet. Auf die verschiedenen Funktionsweisen und Schichten von TFP wird in Abschnitt 2 detaillierter eingegangen.

In dieser Evaluation soll die Bibliothek auf ihre Semantik und Pragmatik, Effektivität beim Erstellen von statistischen Modellen und Integration in das Framework Tensorflow untersucht werden. Das maschinelle Lernen mit Hilfe von neuronalen Netzen und deren Abstraktion durch Keras ist hierbei als Schwerpunkt anzusehen.

2 Tensorflow Probability Komponenten

2.1 Layers

Die Struktur von TFP lässt sich, wie aus der Dokumentation zu entnehmen ist³, in die folgenden vier Schichten einteilen. Die Schichten bauen hierarchisch aufeinander auf, abstrahieren die unterliegenden Schichten aber nicht zwangsläufig. Möchte man beispielsweise durch *MCMC* in Schicht 3 Parameter seines probabilistischen Modells mittels Sampling ermitteln, sollten *Bijectors* aus Schicht 1 kein Fremdwort sein.

¹<https://www.tensorflow.org/probability>

²<https://www.tensorflow.org/guide/graphs>

³<https://www.tensorflow.org/probability/overview>

2.1.1 Layer 0: Tensorflow

TFP ist nicht als eigenständige Komponente neben Tensorflow anzusehen, sondern als Bestandteil dessen. Die probabilistischen Berechnungen werden mit demselben Berechnungsmodell mittels Tensorflow Sessions oder im *Eager*-Modus für sofortige Berechnungen ausgeführt. Tensorflow wird von mehreren Programmiersprachen wie Python, JavaScript oder C++ unterstützt. Die Bibliothek TFP ist aktuell nur für die primär unterstützte Programmiersprache Python implementiert.

2.1.2 Layer 1: Statistical Building Blocks

(Verteilungen / Bijections)

Als Fundament statistischer Modelle sind mehrere, in Python Module aufgeteilte, Klassen und Funktionen gegeben. Diese können in der API Dokumentation der TFP Website eingesehen werden. Ein Beispiel hierfür ist das Modul **tfp.stats**. Unter **tfp.stats** finden sich Funktionen für die Berechnung für Korrelationen, Quantilen oder Standardabweichungen.

Auch verschiedenste, für probabilistische Modelle essentielle Verteilungen im Modul **tfp.distributions** lassen sich dieser Schicht zuordnen: *Normal*-, *Bernoulli*-, *Exponential*- oder *Gammaverteilung*, um einige zu nennen.

Abbildung 1

```
1 normal_dist = tfd.Normal(name="N", loc=0., scale=1.)
2 # -> tfp.distributions.Normal("N", batch_shape=(), event_shape
   =(), dtype=float32)
3
4 sample = normal_dist.sample(sample_shape=normal_sample_size)
```

2.1.3 Layer 2: Model Building

Edward2 / Probabilistic Layers with Keras / Trainable Distributions

2.1.4 Layer 3: Probabilistic Inference

MCMC / VI / Optimizers

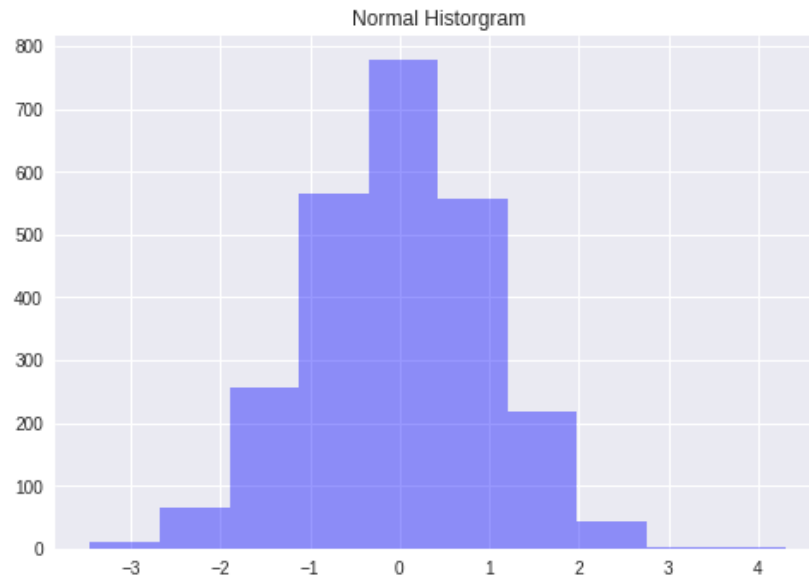


Abbildung 1: $X \sim \mathcal{N}(0, 1)$ mit 2500 Samples

3 Pragmatik und Semantik

4 Integration in Tensorflow

5 Beispiel: Korrelation von Luftverschmutzung und Temperatur

Das Jupyter Notebook ist unter https://github.com/tom-schoener/ml-probability/blob/master/tfp-evaluation/notebooks/air_quality.ipynb einsehbar.

6 Fazit