

HAW HAMBURG  
INFORMATIK MASTER

GRUNDPROJEKT

---

**Evaluation of TensorFlow Probability**  
for enhanced machine learning

---

*Bearbeiter:*

Tom Schöner (2182801)

*Betreuung:*

Prof. Dr. Olaf Zukunft

28. April 2019

# Inhaltsverzeichnis

<b>1</b>	<b>Abstract</b>	<b>2</b>
<b>2</b>	<b>Tensorflow Probability Komponenten</b>	<b>2</b>
2.1	Schicht 0: Tensorflow . . . . .	2
2.2	Schicht 1: Statistical Building Blocks . . . . .	3
2.3	Schicht 2: Model Building . . . . .	5
2.4	Schicht 3: Probabilistic Inference . . . . .	7
<b>3</b>	<b>Semantik und Projektarchitektur</b>	<b>8</b>
3.1	Fehlerbehandlung . . . . .	8
3.2	Dokumentation . . . . .	9
<b>4</b>	<b>Anwendungsgebiete</b>	<b>9</b>
4.1	Beispiel: Korrelation von Luftverschmutzung und Temperatur	10
4.1.1	Modellierung ohne Unsicherheit . . . . .	12
4.1.2	Modellierung mit Unsicherheit . . . . .	13
<b>5</b>	<b>Fazit</b>	<b>16</b>
<b>6</b>	<b>Materialien</b>	<b>17</b>

# 1 Abstract

Durch die Tensorflow Bibliothek Tensorflow Probability<sup>1</sup> v0.7.0 — fortan mit *TFP* abgekürzt — lässt sich maschinelles Lernen mit probabilistischen Modellen der Wahrscheinlichkeitstheorie vereinen. Mittels einer breiten Masse an bereitgestellten Tools, wie statistischen Verteilungen, Sampling oder verschiedenster probabilistischer Funktionen für neuronale Netze, können einfache bis hin zu komplexen Modellen erstellt werden. Berechnungen werden, wie man es aus Tensorflow gewohnt ist, durch *Dataflow Graphs*<sup>2</sup> abgebildet. Auf die verschiedenen Funktionsweisen und Schichten von TFP gehe ich in Abschnitt 2 detaillierter ein.

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

## 2 Tensorflow Probability Komponenten

Die Struktur von TFP lässt sich, wie der Dokumentation zu entnehmen ist<sup>3</sup>, 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.

### 2.1 Schicht 0: Tensorflow

TFP ist nicht als eigenständige Komponente neben Tensorflow anzusehen, sondern als Bestandteil dessen. Die probabilistischen Berechnungen werden in-

---

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

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

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

nerhalb von Tensorflow Sessions oder im *Eager*-Modus 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.

## 2.2 Schicht 1: Statistical Building Blocks

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 unter Anderem Funktionen für die Berechnung für Korrelationen, Quantilen oder Standardabweichungen. Diese Funktionen sind auf die Verwendung von Tensoren ausgelegt, lassen sich aber generell auch mit normalen n-dimensionalen Python Arrays oder Numpy Arrays aufrufen.

Verschiedenste, für probabilistische Modelle essentielle Verteilungen reihen sich unter dem Modul **tfp.distributions** in dieser Schicht ein: *Normal*-, *Bernoulli*-, *Exponential*- oder *Gammaverteilung*, um einige zu nennen. Generell erben Klassen für Verteilungen, wie etwa die Normalverteilung, von der Klasse **tfp.distributions.Distribution**<sup>4</sup>. Diese gemeinsame Schnittstelle standardisiert die Benutzung. Sie fordert zudem Implementationen mehrerer hilfreicher Methoden, wie etwa für die logarithmische Wahrscheinlichkeit oder dem Mittelwert. Am Beispiel der Normalverteilung mit der Definition  $X \sim \mathcal{N}(\mu, \sigma^2)$  sind im Folgenden einige der Funktionen aufgeführt (siehe Listing 1). Visualisiert man die Verteilung auf Basis der Stichproben *sample* (Zeile 4) ergibt sich 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=25000)
5 # <tf.Tensor: id=188, shape=(25000,), dtype=float32, numpy=array
   ([-0.45733708, -0.19126031, -0.33290815, ..., -1.1285563, -0.6958163,
```

---

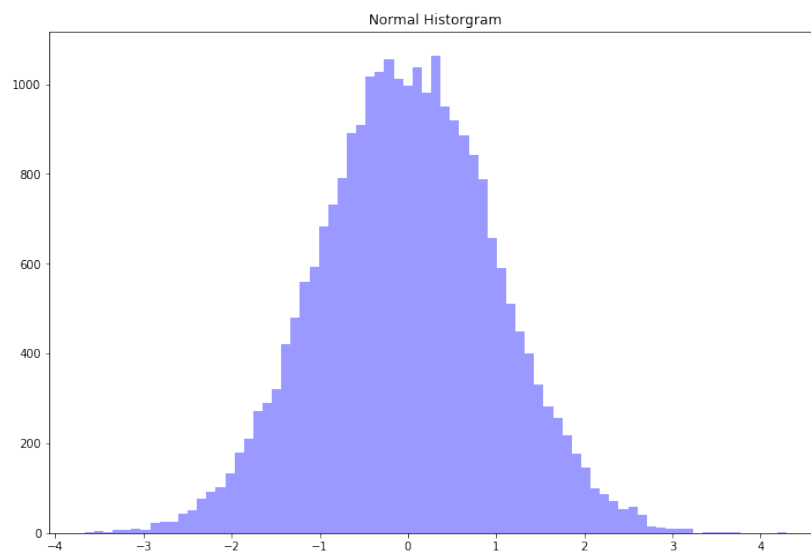
<sup>4</sup>[https://www.tensorflow.org/probability/api\\_docs/python/tfp/distributions/Distribution](https://www.tensorflow.org/probability/api_docs/python/tfp/distributions/Distribution)

```

0.552399  ], dtype=float32)>
6
7 normal_dist.mean()
8 # tf.Tensor(0.0, shape=(), dtype=float32)
9
10 normal_dist.prob(1.0)
11 # tf.Tensor(0.24197072, shape=(), dtype=float32)

```

*Listing 1: Verwendung der Klasse `tfp.distributions.Normal`*



*Abbildung 1:  $X \sim \mathcal{N}(0, 1)$  mit 25000 Samples*

*Broadcasting*, *batching* und *shapes* sorgen dafür, dass unabhängige Verteilungen als sogenannter Batch in einer Entität gekapselt werden können. Um beispielsweise ein zweidimensionales Batch für die Normalverteilung aus Listing 1 zu erzeugen, kann als Erwartungswert  $\mu$  `loc=[ 0., 10.]` übergeben werden. TFP unterstützt broadcasting, somit müssen die folgenden Aufrufe nicht angepasst werden. Das Ergebnis wird ebenfalls Zweidimensional sein. Broadcasting verhält sich analog zu dem in Numpy etablierten Konzept<sup>5</sup>.

Die Shapes lassen sich nochmals in drei Kategorien unterteilen, Tabelle 1 verdeutlicht diese an einem Beispiel.

<sup>5</sup><https://docs.scipy.org/doc/numpy/user/basics.broadcasting.html>

sample shape	Ergebnis
1	(1, 2, 3)
2	(2, 2, 3)
[1, 5]	(1, 5, 2, 3)
[3, 4, 5]	(3, 4, 5, 2, 3)

Tabelle 1: Auswirkung der verschiedenen *sample shapes* mit einem *batch shape* von [2] und einem *event shape* von [3]

- *event shape*: Beschreibt die Form einer Stichprobe aus einer Verteilung, abhängig von seiner Dimension. Skalare Verteilungen besitzen die Form [ ], multivariate Verteilungen, zum Beispiel die Form [ 3 ].
- *batch shape*: Beschreibt die Form unabhängiger, nicht identisch verteilter Stichproben.
- *sample shape*: Beschreibt die Form unabhängiger, aber identisch verteilter Stichproben aus *batches* der Verteilung(en).

Ein weiterer Bestandteil von Schicht 1 sind *Bijectors*. Bijectors bilden eine Zahl  $x_1 \subseteq \mathbb{R}^n$  auf  $x_2 \subseteq \mathbb{R}^m$  ab. Aufgrund der im Bijector hinterlegten bijektiven Funktion ist dieser Schritt umkehrbar, daher:  $x = f^{-1}(f(x))$ . Dies ist besonders bei der Transformierung von Stichproben aus Verteilungen nützlich und wird folglich für das Erstellen von Modellen (Abschnitt 2.3) und Berechnungen probabilistischer Inferenzen (Abschnitt 2.4) eingesetzt. TFP bietet bereits einige vorgefertigte Bijectors unter **tfp.bijectors**<sup>6</sup> an.

## 2.3 Schicht 2: Model Building

Keras ist eine modulare high-level API für die Erstellung und Komposition neuronaler Netze. TFP erweitert das Angebot an Keras-kompatiblen Komponenten durch einige probabilistische Schichten im Modul **tfp.layers**. Die

<sup>6</sup>[https://www.tensorflow.org/probability/api\\_docs/python/tfp/bijectors](https://www.tensorflow.org/probability/api_docs/python/tfp/bijectors)

Architektur orientiert sich weiterhin am selben Keras Model<sup>7</sup>. Neben vorgefertigten Schichten wie **tfp.layers.IndependentNormal** lassen sich durch die Klasse **tfp.layers.DistributionLambda** alle Verteilungen, welche von **tfp.distributions.Distribution** erben, in die Sequenz einbinden. Keras für Tensorflow erlaubt nur konkrete Tensoren und keine generischen Verteilungen, daher muss die jeweilige Schicht wissen, wie sie ihren Ausgabevektor definieren soll. Dies geschieht mittels des übergebenen Parameters *convert\_to\_tensor\_fn*, welcher standardmäßig durch die *sample* Funktion repräsentiert wird. Hierdurch ist eine konfigurierbare Transformation in einen konkreten Tensor gewährleistet.

Für ein funktionierendes Modell müssen die internen Ein- und Ausgabevektoren untereinander kompatibel sein. Sobald das Modell mit *compile()* und *train()* kompiliert und trainiert wurde, können Prognosen mit neuen Daten vorgenommen werden. Die Prognose liefert praktischerweise direkt eine Wahrscheinlichkeitsverteilung zurück, dessen batch shape (siehe Abschnitt 2.2) an die Form der zu prognostizierenden Daten angepasst ist. Haben die Daten die Form [ 100 ], so ergibt sich für die Verteilung beispielsweise ein batch shape von [ 100, 1 ] und repräsentiert somit 100 unabhängige, nicht zwangsläufig gleich verteilte Stichproben. Die über das Training ermittelten Gewichte der Eingabevektoren solcher probabilistischen Schichten lassen sich als Parameter der Verteilung interpretieren, natürlich abhängig von der genauen Implementierung.

Die Interpretation der Gewichte sowie die Verwendung und der Umgang mit Ein- und Ausgabevektoren dieser Schichten in Keras werden in Abschnitt 4.1 anhand eines Beispiels verdeutlicht. Listing 2 zeigt die Verwendung von verschiedenen probabilistischen und nicht probabilistischen Keras Schichten. Das Beispiel ist der Dokumentation<sup>8</sup> entnommen.

```
1 tfd = tfp.distributions
2 tfpl = tfp.layers
3 tfk = tf.keras
4 tfkl = tf.keras.layers
```

---

<sup>7</sup><https://keras.io/getting-started/sequential-model-guide/>

<sup>8</sup>[https://www.tensorflow.org/probability/api\\_docs/python/tfp/layers/IndependentNormal](https://www.tensorflow.org/probability/api_docs/python/tfp/layers/IndependentNormal)

```

5
6 # Create a stochastic encoder – e.g., for use in a variational auto-encoder
7 input_shape = [28, 28, 1]
8 encoded_shape = 2
9 encoder = tfk.Sequential([
10     tfkl.InputLayer(input_shape=input_shape),
11     tfkl.Flatten(),
12     tfkl.Dense(10, activation='relu'),
13     tfkl.Dense(tfpl.IndependentNormal.params_size(encoded_shape)),
14     tfpl.IndependentNormal(encoded_shape)
15 ])

```

*Listing 2: Beispiel probabilistischer und nicht probabilistischer Keras Schichten*

## 2.4 Schicht 3: Probabilistic Inference

Als letzte Schicht enthält TFP Werkzeuge für probabilistische Inferenz. Dazu gehören *Markov chain Monte Carlo (MCMC)* Algorithmen, *Variational Inference* Algorithmen und stochastische Optimierungsverfahren.

Durch Sampling ermöglicht MCMC das Berechnen statistischer Modelle mit einer hohen Anzahl an Parametern. Eine analytische Herangehensweise ist bereits ab wenigen Parametern nicht mehr sinnvoll, da hierfür die Bestimmung eines multidimensionalen Integrals erforderlich wäre. TFP bietet mehrere MCMC Algorithmen unter **tfp.mcmc** an.

Listing 3 zeigt einen Ausschnitt der Verwendung des Hamiltonian Monte Carlo Algorithmus. Die hier nicht enthaltene Auswertung würde in einer approximierten Verteilung des Parameters *tau* für das gegebene Modell *\_data\_model* resultieren. Ein praxisbezogenes und ausführlicheres Beispiel ist unter <https://github.com/tom-schoener/ml-probability/blob/master/tfp-evaluation/notebooks/mcmc.ipynb> einsehbar.

```

1 # Set the chain's start state.
2 initial_chain_state = [ 0.5 * tf.ones([], dtype=tf.float32, name="init_tau")
3 ]
4 def joint_log_prob(_data_model, tau):
5     rv_tau = tfd.Uniform()
6     rv_observation = tfd.Poisson(rate=rv_tau)
7
8     return rv_tau.log_prob(tau) + tf.reduce_sum(rv_observation.log_prob(
9         _data_model))

```



```

9
10 # setup the chain
11 [ tau ], kernel_results = tfp.mcmc.sample_chain(
12     num_results=1000,
13     num_burnin_steps=500,
14     current_state=initial_chain_state ,
15     kernel=tfp.mcmc.TransformedTransitionKernel(
16         inner_kernel=tfp.mcmc.HamiltonianMonteCarlo(
17             target_log_prob_fn=lambda tau: joint_log_prob(_data.model, tau)),
18             bijector=[ tfp.bijectors.Sigmoid() ] # Maps [0,1] to R
19         )
20 )

```

Listing 3: Verwendung des Hamiltonian Monte Carlo Algorithmus (gekürzt)

Die in der API verwendeten Bezeichnungen decken sich hier größtenteils mit der in der Literatur verwendeten Sprache, wodurch sich Methoden anderer Quellen relativ problemlos übertragen lassen.

Zu den Optimierungsverfahren von TFP gehören verschiedenste Algorithmen. Ohne an dieser Stelle zu sehr ins Detail zu gehen, gibt es beispielsweise unter `tfp.optimizer.bfgs_minimize`<sup>9</sup> einen *Optimizer*[2], welcher einen Tiefpunkt einer differenzierbaren Funktion approximiert ermittelt. Vergleichbar sind diese Algorithmen mit dem bereits in Tensorflow enthaltenen Optimierungsverfahren<sup>10</sup>.

*Variational Inference* in TFP habe ich für diese Evaluation nicht weiter untersucht.

## 3 Semantik und Projektarchitektur

### 3.1 Fehlerbehandlung

Ein nicht zu vernachlässigender Teil der Softwareentwicklung ist das Erkennen, Interpretieren und Beheben von semantischen Fehlern. Als Teil von Tensorflow gibt TFP bereits sinnvolle und leserliche Fehlermeldungen aus. Zum Beispiel wird bei falschen Typen innerhalb eines Tensors folgende Fehlermeldung

<sup>9</sup>[https://www.tensorflow.org/probability/api\\_docs/python/tfp/optimizer/bfgs\\_minimize](https://www.tensorflow.org/probability/api_docs/python/tfp/optimizer/bfgs_minimize)

<sup>10</sup>[https://www.tensorflow.org/versions/r2.0/api\\_docs/python/tf/optimizers](https://www.tensorflow.org/versions/r2.0/api_docs/python/tf/optimizers)

mit Stacktrace zurückgegeben: *ValueError: Can't convert Python sequence with mixed types to Tensor*. Auf fehlende, nicht optionale Parameter wird ebenfalls hingewiesen. Befindet man sich nicht im Eager-Execution Modus, findet bereits beim Erstellen des Graphen und somit vor dessen Evaluation ein solches Typechecking statt.

## 3.2 Dokumentation

TFPs offizielle Dokumentation bietet neben der Beschreibung der API eine Auflistung an Beispielen und weiterführenden Ressourcen. Die als Einführung anzusehende und kürzlich nach TFP portierte Version von *Bayesian Methods for Hackers: Probabilistic Programming and Bayesian Inference*<sup>[1]</sup> stellte sich als besonders hilfreich heraus.

Die Bibliothek befindet sich in einem vergleichsweise frühen Stadium der Entwicklung. Es wird darauf hingewiesen, dass es in Zukunft zu Änderungen kommen kann. Teile der API sind entsprechend noch nicht vollständig dokumentiert (Stand März 2019). Undokumentierte Funktionen sind allerdings eher die Ausnahme als die Norm. Ein Großteil der in dieser Evaluation verwendeten Schnittstellen ließ sich anhand von Beispielen, genaueren Erläuterungen der Parameter oder durch Verweise auf relevante Literatur und Paper nachvollziehen. Einzig ein Überblick oder eine Einleitung der individuellen Module und deren Funktionsweisen fehlt fast vollständig. Die Verwendung dieser Module wird dadurch nicht sofort ersichtlich und erschwert deren Benutzung. Häufig musste ich auf Beispiele oder externe Quellen zurückgreifen, um mir einen Überblick verschaffen zu können.

## 4 Anwendungsgebiete

Das Ziel maschinellen Lernens ist das Erkennen und Übertragen von Mustern. Reale Daten bringen allerdings immer eine gewisse Unsicherheit bei der Auswertung mit sich. Die Wahrscheinlichkeitstheorie wirkt diesem Problem, unter Anderem mithilfe der bayesschen Inferenz, entgegen. Prognosen lassen sich infolgedessen auf mathematischer Grundlage einschätzen und verifizieren.

Auch die (Hyper-)Parameter des Modells können als Zufallsvariable einer Wahrscheinlichkeitsverteilung dargestellt werden. Dieser Ansatz schränkt die Anwendungsgebiete nicht ein, sondern bietet lediglich eine neue Sichtweise auf maschinelles Lernen. Das folgende Beispiel soll diesen Vorteil und die Anwendung von TFP für probabilistische Modelle zeigen. In *Machine Learning: A Probabilistic Perspective*[3] Kapitel 1.2.1.2 ff. (*The need for probabilistic predictions*) werden weitere konkrete Beispiele, wie die Klassifizierung von Blumen, erläutert.

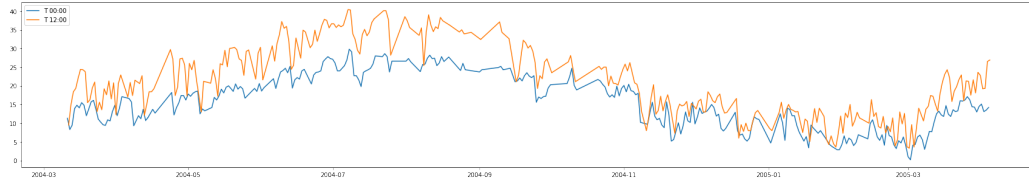
## 4.1 Beispiel: Korrelation von Luftverschmutzung und Temperatur

Das Jupyter Notebook für das folgende Beispiel ist unter [https://github.com/tom-schoener/ml-probability/blob/master/tfp-evaluation/notebooks/air\\_quality.ipynb](https://github.com/tom-schoener/ml-probability/blob/master/tfp-evaluation/notebooks/air_quality.ipynb) einsehbar. Die Daten stammen aus dem *UC Irvine Machine Learning Repository*<sup>11</sup>.

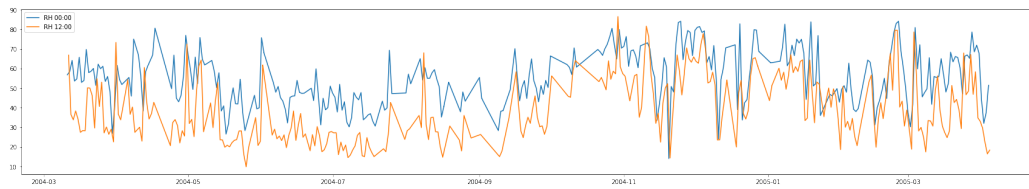
Von März 2004 bis Februar 2005 wurden in einer italienischen Stadt nahe einer stark befahrenen Straße verschiedene Wetterdaten wie Temperatur, Luftfeuchtigkeit oder auch Stickstoffdioxidgehalt ( $NO_2$ ) in der Luft stündlich gemessen. Die Daten für Temperatur, relativer Luftfeuchtigkeit und  $NO_2$  sind aggregiert in den Abbildungen 2, 3 und 4 visualisiert. Die blaue Linie stellt den gemittelten Wert um 00:00Uhr und die orangene Linie um 12:00Uhr dar.

---

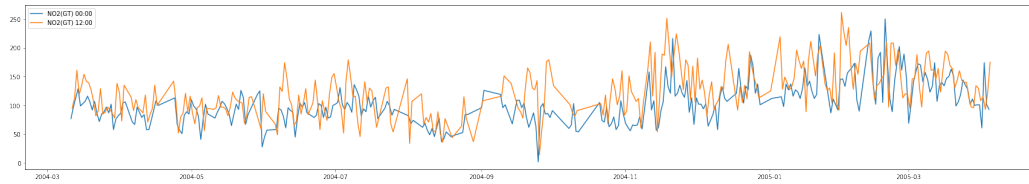
<sup>11</sup><https://archive.ics.uci.edu/ml/datasets/Air+quality>



*Abbildung 2: Temperatur in °C*



*Abbildung 3: Relative Luftfeuchtigkeit in %*



*Abbildung 4: Stickstoffdioxid  $NO_2$  in  $\mu g/m^3$*

In Europa ist der Grenzwert von  $NO_2$  gesetzlich auf max  $200\mu g/m^3$  festgelegt. Die EU-Richtlinie 2008/50/EG verschärfte dieses Gesetz insofern, dass das Jahresmittel  $40\mu g/m^3$  nicht überschreiten darf. In diesem Beispiel soll ein probabilistisches Modell für Aussagen über die Korrelation zwischen Temperatur und Luftverschmutzung auf Basis eines neuronalen Netzes in TFP erstellt werden. Da bei solchen Modellen Verlust (*loss*) mitmodelliert werden kann, wird nicht nur die Relation der Daten, sondern auch die Aussagekraft des jeweiligen ermittelten Wertes, zum Beispiel über dessen Standardabweichung, verdeutlicht.

### 4.1.1 Modellierung ohne Unsicherheit

Um das Modell erstellen zu können, müssen die Daten zunächst bereinigt werden. Null- oder fehlende Werte werden der Einfachheit halber interpoliert. In Listing 4 ist das Erstellen eines einfachen linearen Modells mit TFP und Keras zu sehen. Das Modell besteht aus zwei Schichten: Einem *dense layer* mit einem eindimensionalen Ausgabevektor und eine durch eine Normalverteilung beschriebene Schicht, deren Eingabevektor durch den Ausgabevektor des *dense layers* befüllt wird. Die Normalverteilung benutzt diesen Input, um den Mittelwert zu bestimmen. Beim Trainieren des Modells wird der Verlust (*loss*) durch die Funktion *neg\_log\_lik*, die den negativen Logarithmus einer Probe der Verteilung bestimmt, festgelegt. Es sollte auffallen, dass die Standardabweichung konstant bei 1 liegt, wodurch das resultierende Modell ebenfalls eine konstante Standardabweichung für jede Temperatur aufweisen wird. Abschnitt 4.1.2 erweitert das Modell um diesen Aspekt. Die durch das Modell ermittelten Gewichte und eine Visualisierung der sich daraus ergebenden Wahrscheinlichkeitsverteilung sind im Jupyter Notebook für beide Beispiele hinterlegt.

```
1 def model_no_uncertainty(feature, label, epochs=100):
2     # feature
3     x = np.array(df[feature])
4     x = x[... , np.newaxis]
5
6     # label
7     y = np.array(df[label])
8
9     # model
10    model = tf.keras.Sequential([
11        tf.keras.layers.Dense(1),
12        tfp.layers.DistributionLambda(lambda t: tfd.Normal(loc=t, scale=1)),
13    ])
14
15    # inference
16    neg_log_lik = lambda y_: -rv_y.log_prob(y_)
17    model.compile(optimizer=tf.optimizers.Adam(learning_rate=0.01), loss=
18        neg_log_lik)
19    model.fit(x, y, epochs=epochs)
20
21    return (x, y, model)
```

Listing 4: Modell mit Keras ohne Unsicherheit

Verwendet man dieses Modell, um anhand der Temperatur eine Prognose für jeweils die Luftfeuchtigkeit und den Stickstoffgehalt in der Luft zu bestimmen, ergeben sich die Abbildungen 5 und 6.

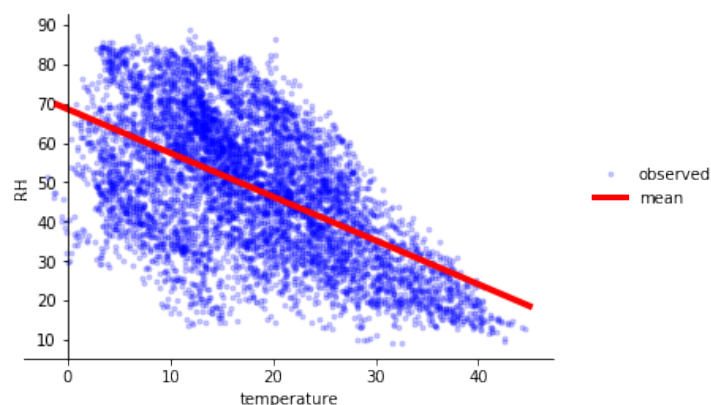


Abbildung 5: Prognose der relativen Luftfeuchtigkeit

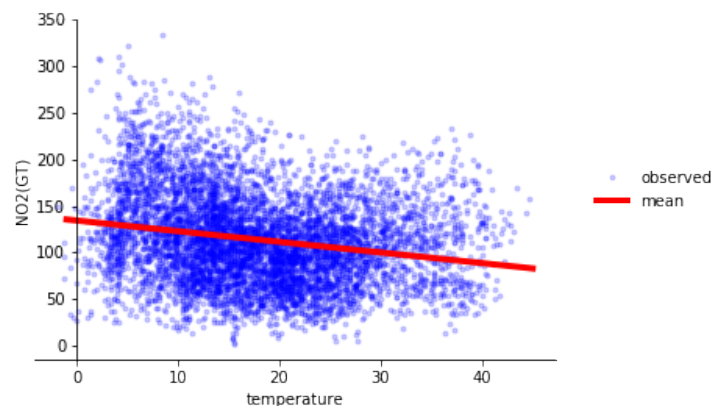


Abbildung 6: Prognose des Stickstoffgehalts in der Luft

#### 4.1.2 Modellierung mit Unsicherheit

Betrachtet man Abbildung 5 fällt auf, dass die Unsicherheit des hier linearen Ansatzes bei geringen Temperaturen stark zunimmt. Mit anderen Worten:

Die Genauigkeit lässt nach. Das Modell kann dieses Problem aber noch nicht repräsentieren, daher wird im folgenden, verbesserten Modell die Unsicherheit integriert (siehe Listing 5). Der Aufbau ist dem Ersten sehr ähnlich und unterscheidet sich im Wesentlichen nur in zwei Aspekten:

- Die erste Schicht hat einen zweidimensionalen Ausgabevektor, um die Unsicherheit mit darzustellen.
- Die zweite Schicht verwendet den zweiten Wert des Vektors, um seine Standardabweichung zu definieren.

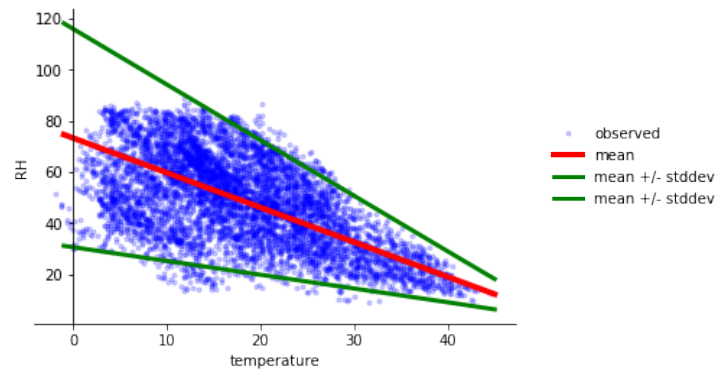
```
1 def model_with_uncertainty(feature, label, epochs=5000):
2     # feature
3     x = np.array(df[feature])
4     x = x[... , np.newaxis]
5
6     # label
7     y = np.array(df[label])
8
9     model = tf.keras.Sequential([
10         tf.keras.layers.Dense(2),
11         tfp.layers.DistributionLambda(lambda t: tfd.Normal(
12             loc=t[... , :1],
13             scale=1e-3 + tf.math.softplus(5e-3 * t[... , 1:])))
14     ])
15
16     # inference
17     negloglik = lambda y_: -rv-y.log_prob(y_)
18     model.compile(optimizer=tf.compat.v2.optimizers.Adam(learning_rate=0.01),
19                   loss=negloglik)
19     model.fit(x, y, epochs=epochs);
20
21     return (x, y, model)
```

*Listing 5: Modell mit Keras mit Unsicherheit*

Das Ergebnis lässt sich den Abbildungen 7 und 8 entnehmen. Die Standardabweichung ist nun Teil des Modells. Akzeptiert man hier das lineare Modell, lässt sich für den Stickstoffgehalt in der Luft allgemein aussagen, dass die Prognosesicherheit mit steigenden Temperaturen sinkt und der Stickstoffgehalt in der Luft interessanter Weise ebenfalls fällt. Das Modell der relativen Luftfeuchtigkeit zeigt, dass bei steigenden Temperaturen die Genauigkeit der

Prognosen stark zunimmt, was sich mit der Erwartungshaltung gegenüber der Relation von Temperatur und Luftfeuchtigkeit deckt.

Auffällig ist hier, dass die Prognose der Unsicherheit des Stickstoffgehalts in der Luft die Daten nicht sehr genau abbildet. Das kann mehrere Gründe haben. Hier wurde nur ein einfaches lineares Modell für Demonstrationszwecke verwendet. Einige der Hyperparameter könnten weiter angepasst werden und die Trainingszeit oder Datenmenge ist nicht unbedingt ausreichend. Natürlich kann es auch sein, dass die Temperatur kein sinnvoller Maßstab für den Stickstoffgehalt in der Luft ist, wodurch das Modell keine große Aussagekraft besäße.



*Abbildung 7: Prognose der relativen Luftfeuchtigkeit mit variabler Standardabweichung*



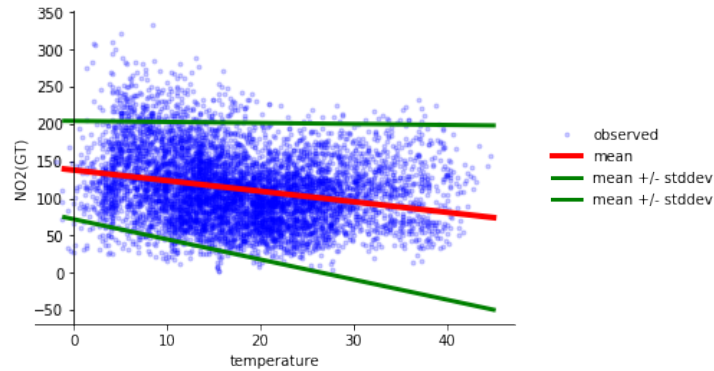


Abbildung 8: Prognose des Stickstoffgehalts in der Luft mit variabler Standardabweichung

## 5 Fazit

Die Bibliothek TFP ist eine praktische Erweiterung des Tensorflow Frameworks, um sein Modell um eine probabilistische Ebene zu erweitern. Die bereitgestellten Funktionen sind sinnvoll aufgeteilt und es sind im Normalfall Beispiele für deren Verwendung vorhanden. Auch beim Release von Tensorflow 2.0 wurden Anpassungen an TFP vorgenommen, was das Interesse von Google und der OpenSource Community an der Weiterentwicklung von TFP zeigt.

Die Einarbeitungszeit ist, auch unter Ansicht der Komplexität des Themengebietes, nicht zu vernachlässigen. Ein Beispiel hierfür sind *Dataflow Graphs*. *Dataflow Graphs* von Tensorflow ermöglichen ein effizientes Berechnungsmodell, welches auf GPUs und auf neuere spezialisierte Hardware wie TPUs ausgelegt ist. Einfache Anwendungsfälle profitieren aber nicht unbedingt von der Abstraktionsebene von Tensorflow. Eine alternative und bekannte Bibliothek für probabilistische Programmierung in Python, anstelle von TFP, ist beispielsweise *PyMC3*<sup>12</sup>.

Besonders im Hinblick auf maschinelles Lernen eignet sich der Einsatz von TFP. Es reiht sich relativ problemlos in das bestehende Framework ein. Die Integration in Keras ist eine der nennenswertesten Stärken von TFP,

<sup>12</sup><https://docs.pymc.io/>

da die Komplexität an der richtigen Stelle gelockert wird. Probabilistische Modelle zu erstellen wird hierdurch nicht trivial, aber deutlich zugänglicher. Ein fundamentales Wissen über die Domäne und das breite Gebiet der Wahrscheinlichkeitstheorie ist natürlich noch immer erforderlich.

## 6 Materialien

An dieser Stelle möchte ich eine Auswahl an hilfreichen Materialien zum Thema *bayesian programming* aufführen, welche mir bei der Recherche positiv aufgefallen sind:

- Jährliche Konferenz über Bayesian Deep Learning (NIPS) <http://bayesiandeeplearning.org/>
- Machine Learning: A Probabilistic Perspective (Kevin P. Murphy)[3]
- Präsentation über Tensorflow Probability des Entwicklers Josh Dillon <https://youtu.be/GqxmRKplj4w> (Folien: [https://docs.google.com/presentation/d/1BWhNVHzhFfYiFL8ynX1wFmag8YjzvHeqKkd8ZOG8H7Y/edit#slide=id.g3f2acb32c4\\_0\\_0](https://docs.google.com/presentation/d/1BWhNVHzhFfYiFL8ynX1wFmag8YjzvHeqKkd8ZOG8H7Y/edit#slide=id.g3f2acb32c4_0_0))

## Literatur

- [1] Cameron Davidson-Pilon. *Bayesian Methods for Hackers: Probabilistic Programming and Bayesian Inference*. Addison-Wesley Professional, 1st edition, 2015. ISBN 0133902838, 9780133902839. URL <https://camdavidsonpilon.github.io/Probabilistic-Programming-and-Bayesian-Methods-for-Hackers/#tensorflow>.
- [2] Stephen Wright Jorge Nocedal. Numerical optimization. In *Springer Series in Operations Research*, pages 136–140. Springer, 2006. ISBN 0387303030. URL [http://pages.mtu.edu/~struther/Courses/OLD/Sp2013/5630/Jorge\\_Nocedal\\_Numerical\\_optimization\\_267490.pdf](http://pages.mtu.edu/~struther/Courses/OLD/Sp2013/5630/Jorge_Nocedal_Numerical_optimization_267490.pdf).
- [3] Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012. ISBN 0262018020, 9780262018029. URL [https://doc.lagout.org/science/Artificial%20Intelligence/Machine%20learning/Machine%20Learning\\_%20A%20Probabilistic%20Perspective%20%5BMurphy%202012-08-24%5D.pdf](https://doc.lagout.org/science/Artificial%20Intelligence/Machine%20learning/Machine%20Learning_%20A%20Probabilistic%20Perspective%20%5BMurphy%202012-08-24%5D.pdf).

## Abbildungsverzeichnis

1	$X \sim \mathcal{N}(0, 1)$ mit 25000 Samples . . . . .	4
2	Temperatur in °C . . . . .	11
3	Relative Luftfeuchtigkeit in % . . . . .	11
4	Stickstoffdioxid $NO_2$ in $\mu g/m^3$ . . . . .	11
5	Prognose der relativen Luftfeuchtigkeit . . . . .	13
6	Prognose des Stickstoffgehalts in der Luft . . . . .	13
7	Prognose der relativen Luftfeuchtigkeit mit variabler Standard- abweichung . . . . .	15
8	Prognose des Stickstoffgehalts in der Luft mit variabler Stan- dardabweichung . . . . .	16

## Listings

1	Verwendung der Klasse <code>tfp.distributions.Normal</code> . . . . .	3
2	Beispiel probabilistischer und nicht probabilistischer Keras Schichten . . . . .	6
3	Verwendung des Hamiltonian Monte Carlo Algorithmus (gekürzt)	7
4	Modell mit Keras ohne Unsicherheit . . . . .	12
5	Modell mit Keras mit Unsicherheit . . . . .	14