

# Logistische Regression

Thomas Sine-Nomine

## Überblick

Die logistische Regression wird häufig zum Abschätzen der Wahrscheinlichkeit eingesetzt, dass ein Datenpunkt einer bestimmten Kategorie angehört (z. B.: Wie hoch ist die Wahrscheinlichkeit, dass eine E-Mail Spam enthält?). Wenn die geschätzte Wahrscheinlichkeit mehr als 50 % beträgt, sagt das Modell vorher, dass der Datenpunkt zu dieser Kategorie gehört. Somit lässt sich die logistische Regression im Gegensatz zu anderen Regressionsmodellen zur Klassifikation einsetzen.

## Modellfunktion

Als Modellfunktion kommt die Sigmoid-Funktion zum Einsatz:

$$h_{\theta}(x) = \frac{1}{1 + e^{-z}} \text{ mit } z = \theta^T x$$

Es handelt sich hierbei um eine nicht-lineare Funktion. Die folgende Abbildung zeigt den Funktionsgraphen.

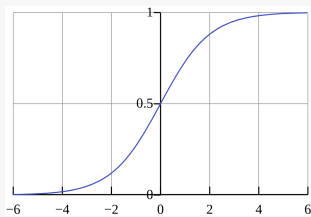


Abb. 1: Logistische Funktion

## Kostenfunktion

Für die Kostenfunktion wird die mittlere quadratische Abweichung (englisch *mean squared error*) verwendet:

$$\text{MSE}(\theta) = \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

$\text{MSE}(\theta)$  ist nicht konvex und besitzt in der Regel lokale Minima. Das Gradientenverfahren lässt sich auf diese Funktion nicht anwenden.

## Anwendungsgebiete

- Vorhersage von Wahlergebnissen
- medizinische Diagnosen
- Kaufverhalten
- Prüfung von Kreditanträgen auf Risiken
- Nutzerverhalten im Internet

## Beispiel-Usecase

Eine Anwendungsmöglichkeit ist die Vorhersage des Bestehens bzw. Nicht-Bestehens einer Prüfung in Abhängigkeit von der investierten Vorbereitungszeit.

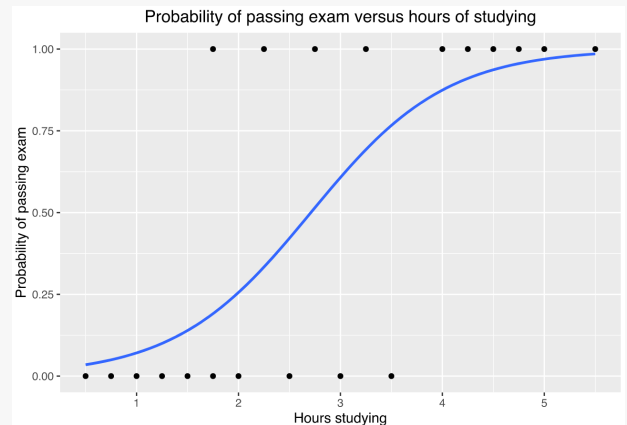


Abb. 2: Vorhersage von Prüfungsergebnissen mittels logistischer Regression.

Bei einem Funktionswert von  $> 0,5$  ist ein Bestehen der Prüfung wahrscheinlicher.

## Programmbeispiel

In scikit-learn ist die logistische Regression im Modul `LogisticRegression` verfügbar. Das Training erfolgt wie gewohnt mit `fit()`.

```
from sklearn.linear_model import
    LogisticRegression
log_reg = LogisticRegression()
log_reg.fit(X, y)
```

Vorhersagen werden mittels `predict` gemacht:

```
log_reg.predict([[1.7], [1.5]])
```

## Literatur

- [1] Data Basecamp. *Was ist eine Logistische Regression?* 2022. URL: <https://databasecamp.de/ki/logistische-regression> (besucht am 26.02.2024).
- [2] Aurélien Géron, Kristian Rother und Thomas Demmig. *Praxiseinstieg Machine Learning mit Scikit-Learn, Keras und TensorFlow. Konzepte, Tools und Techniken für intelligente Systeme*. 2. Aufl. Heidelberg: O'Reilly, 2020.
- [3] Wikipedia. *Logistic regression*. URL: [https://en.wikipedia.org/w/index.php?title=Logistic\\_regression&oldid=1210273561](https://en.wikipedia.org/w/index.php?title=Logistic_regression&oldid=1210273561) (besucht am 26.02.2024).