

Datenanalyse auf Basis von KI-Methoden

Zeitreihenanalyse

Eine Zeitreihe ist eine sequentielle Sammlung von Datenpunkten, die typischerweise über aufeinanderfolgende Zeiten gemessen werden. Sie ist mathematisch definiert als:

eine Sammlung von Vektoren $y(t), t=0,1,2,3,\dots$ wobei t die Zeit darstellt. Die Variable $y(t)$ wird als Zufallsvariable behandelt.

Zeitreihenanalyse

Zeitreihen können univariat oder multivariat sein.

Univariate Zeitreihen sind Zeitreihen, die Datensätze einer einzelnen Variablen enthalten,

während multivariate Zeitreihen Datensätze von mehr als einer Variablen berücksichtigen.

Zeitreihenanalyse

Die Zeitreihenanalyse beschäftigt sich mit statistischen Methoden zur Analyse und Modellierung einer geordneten Folge von Beobachtungen (Zeitreihe).

Diese Modellierung führt zu einem Prozessmodell für das System, das die Daten erzeugt hat.

Somit können anhand dieses Modells zukünftige Ereignisse vorhergesagt werden.

Zeitreihenanalyse

Die Prognoseverfahren der Zeitreihenanalyse lassen sich grob in zwei Gruppen einteilen:

- die Komponenten-Modelle
- und die ARIMA-Modelle

Zeitreihenanalyse

Komponenten-Modelle

Die Komponenten-Modelle sind ein traditioneller Ansatz der Zeitreihenanalyse und ihre Anpassung erfolgt mit Hilfe deterministischer mathematischer Modelle, die die systematische Komponente der Zeitreihe umfassen.

Diese Komponenten sind:

Trend-,
periodische oder saisonale Schwankungen und
Restkomponenten.

Zeitreihenanalyse

ARIMA-Modelle

Eines der wichtigsten und am weitesten verbreiteten Zeitreihenmodelle ist das AutoRegressive Integrated Moving Average (ARIMA)-Modell

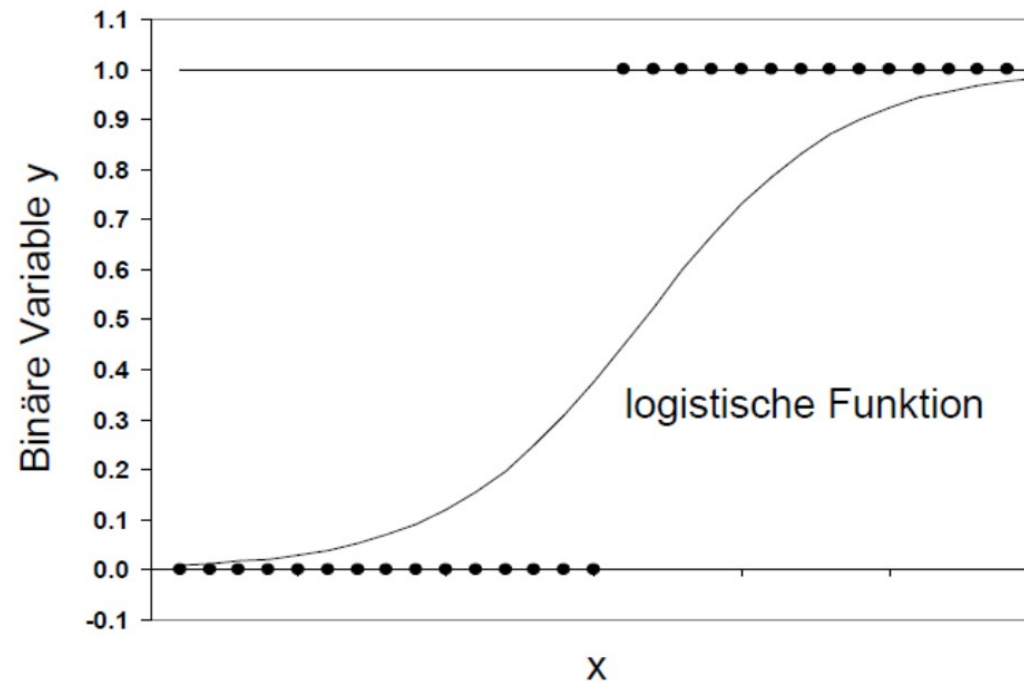
Ein ARIMA-Modell stellt mittels einer linearen Funktion die aktuellen Werte einer Zeitreihe in Bezug

zu seinen eigenen Vergangenheitswerten (die autoregressive Komponente -AR-) und

die Vergangenheitswerte des Fehler-Terms (Komponente des gleitenden Mittelwerts -MA-) dar.

Logistische Regression

Die (binär) logistische Regressionsanalyse testet, ob ein Zusammenhang zwischen mehreren unabhängigen und einer binären abhängigen Variable besteht.



Logistische Regression - Modelgüte

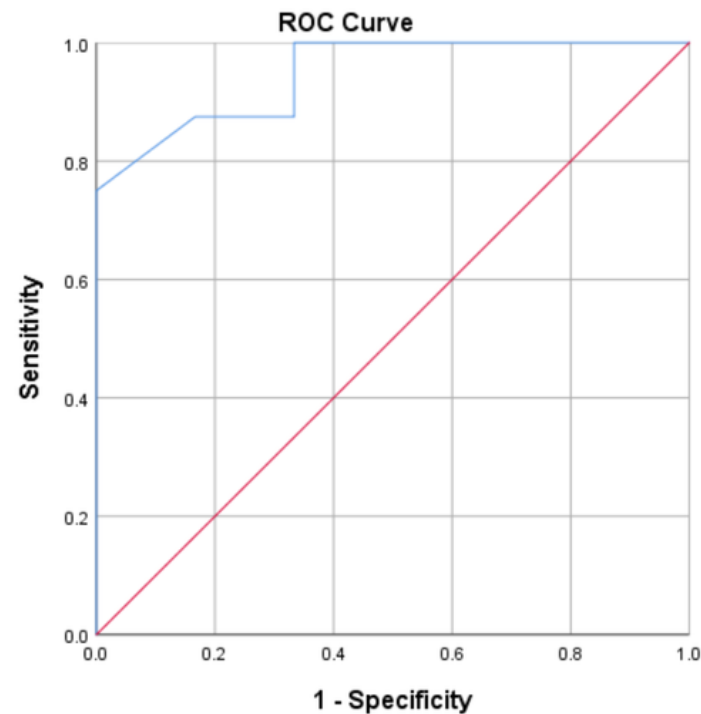
Um zu beurteilen, wie gut ein logistisches Regressionsmodell zu einem Datensatz passt, können wir die folgenden zwei Metriken betrachten:

- **Sensitivität:** Die Wahrscheinlichkeit, dass das Modell ein positives Ergebnis für eine Beobachtung vorhersagt, wenn das Ergebnis tatsächlich positiv ist.
- **Spezifität:** Die Wahrscheinlichkeit, dass das Modell ein negatives Ergebnis für eine Beobachtung vorhersagt, wenn das Ergebnis tatsächlich negativ ist.

Eine einfache Möglichkeit, diese beiden Metriken zu visualisieren, besteht darin, eine **ROC-Kurve** zu erstellen.

Logistische Regression – ROC-Kurve

ROC-Kurve ist ein Diagramm, das die Sensitivität und Spezifität eines logistischen Regressionsmodells anzeigt.



Diagonal segments are produced by ties.

***Vielen Dank
Für Ihre
Aufmerksamkeit!***