Grundlagen der Künstlichen Intelligenz - Informatik

Rapp, DHBW Lörrach

10.11.2023

Inhaltsübersicht

Lernziele

Meine 3 Lernziele für heute

1. Ich kenne den Satz von Bayes und kann ihn zur Anwendung bringen.
2. Ich bin in der Lage, ein Bayes-Netz zu interpretieren und kann daraus Wahrscheinlichkeiten für das Eintreten von Ereignissen berechnen.
3. Ich kenne die Voraussetzungen für stationäre Zeitreihen und das Vorgehen zur Bestimmung der Modellordnungen im ARIMA-Modell.

Literaturempfehlung

Bayes-Netz

- Grundkurs Künstliche Intelligenz - Eine praxisorientierte Einführung, *Wolfgang Ertel*

Rechnen mit Wahrscheinlichkeiten

Überblick

Wahrscheinlichkeit

Maß der Erwartung für ein unsicheres Ereignis.

Ereignis

Teil einer Menge von möglichen Ergebnissen *(Ergebnisraum)* eines Zufallsexperiments, dem eine Wahrscheinlichkeit zugeordnet werden kann.

Beispiel: "Würfeln einer geraden Zahl"

- entspricht der Teilmenge $\{2, 4, 6\}$ aus dem Ergebnisraum $\{1, 2, 3, 4, 5, 6\}$
- Dem Ereignis wird die Wahrscheinlichkeit $p = \frac{1}{2}$ zugeordnet.

Spezialfall: Sicheres Ereignis

Das mit dem Ergebnisraum Ω identische Ereignis bezeichnet man als sicheres Ereignis, da es immer eintritt.

Laplace Formel

Laplace Experiment

Zufallsexperiment, bei dem alle elementaren Ergebnisse dieselbe Wahrscheinlichkeit haben:

$$P(\omega) = \frac{1}{|\Omega|}$$

Beispiele: *Werfen von*

- Münzen oder Würfeln
- Nichtbeispiel: Reißzwecken

Laplace Formel

$$P(A) = \frac{\text{"}Anzahl\ der\ Ergebnisse,\ bei\ denen\ das\ Ereignis\ A\ eintritt\text{"}}{\text{"}Anzahl\ aller\ möglichen\ Ergebnisse\text{"}}$$

$$= \frac{|A|}{|\Omega|} \quad \text{: Wahrscheinlichkeit für ein Ereignis A in einem Laplace Experiment}$$

Rechnen mit Wahrscheinlichkeiten

Satz

1. $P(\Omega) = 1$.
2. $P(\emptyset) = 0$, d.h. das unmögliche Ereignis hat die Wahrscheinlichkeit 0.
3. Für paarweise unvereinbare Ereignisse A und B gilt
   $P(A \vee B) = P(A) + P(B)$.
4. Für zwei zueinander komplementäre Ereignisse A und $\neg A$ gilt
   $P(A) + P(\neg A) = 1$.
5. Für beliebige Ereignisse A und B gilt
   $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$.
   Hinweis: $P(A, B)$ ist zu $P(A \wedge B)$ analoge Schreibweise.
6. Sind $A_1, ..., A_n$ die Elementarereignisse, so gilt
   $\sum_{i=1}^{n} P(A_i) = 1$ (Normierungsbedingung).

Aus: Grundkurs Künstliche Intelligenz, Ertel

Erwartungswert

Der **Erwartungswert** beschreibt die Zahl, die eine Zufallsvariable im Mittel bei mehrmaligem Wiederholen eines Experiments annimmt.
Bei einer **diskreten Zufallsvariablen** berechnet sich der Erwartungswert *(normalerweise μ genannt)* als Summe der Produkte von Wert und dessen Wahrscheinlichkeit:

$$\begin{aligned} E(X) &= x_1 \cdot P(X = x_1) + \ldots + x_n \cdot P(X = x_n) \\ &= \sum_{i=1}^{n} x_i \cdot P(X = x_i) \end{aligned}$$

Beispiel: Würfeln von zwei Laplace-Würfeln mit den Werten der Zufallsgröße X gleich der Summe der Augenzahlen.

x	2	3	4	5	6	7	8	9	10	11	12
$P(X = x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

$$\begin{aligned} E(X) \;=\; & 2 \cdot \tfrac{1}{36} + 3 \cdot \tfrac{2}{36} + 4 \cdot \tfrac{3}{36} + 5 \cdot \tfrac{4}{36} + 6 \cdot \tfrac{5}{36} + 7 \cdot \tfrac{6}{36} \\ & + 8 \cdot \tfrac{5}{36} + 9 \cdot \tfrac{4}{36} + 10 \cdot \tfrac{3}{36} + 11 \cdot \tfrac{2}{36} + 12 \cdot \tfrac{1}{36} = 7 \end{aligned}$$

$\Rightarrow$ Im Mittel würfelt man beim Werfen von zwei fairen Würfeln eine „7“.

Nullhypothese

Ziel eines statistischen Tests

Falsifizieren *(Verwerfen)* der Nullhypothese, so dass die Arbeitshypothese als Möglichkeit übrig bleibt ($\rightarrow$ disjunktes Gegensatzpaar).

- Nullhypothese sagt z.B. aus, dass *kein* Effekt bzw. Unterschied vorliegt oder dass ein bestimmter Zusammenhang *nicht* besteht.

Beispiel

- **Arbeitshypothese:** Es besteht der Verdacht, dass es einen *prinzipiellen Unterschied* zwischen Frauen und Männern in Bezug auf ein bestimmtes (z.B. medizinisches) Testergebnis gibt.
- **Formulierung als Nullhypothese** $H_0 : \mu_1 = \mu_2$, mit Erwartungswerten der Testergebnisse für μ_1 : Frauen und μ_2 : Männer
- Spricht das Stichprobenergebnis gegen die Annahme, so wird die **Nullhypothese abgelehnt,** andernfalls wird sie **beibehalten.**

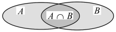
Definition

Bedingte Wahrscheinlichkeit

Wahrscheinlichkeit, dass Ereignis A eintritt unter der Bedingung, dass das Eintreten eines anderen Ereignisses B bereits bekannt ist:

$$P(A|B) = \frac{P(A \wedge B)}{P(B)}$$

Unabhängige Ereignisse



Zwei Ereignisse A und B sind unabhängig, wenn $P(A|B) = P(A)$ gilt.

Es folgt:

$$P(A|B) = \frac{P(A \wedge B)}{P(B)} \overset{unabh.}{=} P(A)$$

$$\Rightarrow P(A \wedge B) = P(A) \cdot P(B)$$

Beispiel unabhängige Würfel:

$P(W_1 = 6 \wedge W_2 = 6) = P(W_1) \cdot P(W_2) \overset{Laplace}{=} \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$

Beispiel abhängige Würfel, d.h. 2. Würfel fällt immer gleich wie der erste:

$P(W_1 = 6 \wedge W_2 = 6) = P(W_2 = 6 | W_1 = 6) \cdot P(W_1 = 6) = \frac{1}{6}$

Marginalisierung

Für zweiwertige Zufallsvariablen A und B gilt:

$$\begin{aligned} P(A) &= P(A \wedge (B \vee \neg B)) \\ &= P((A \wedge B) \vee (A \wedge \neg B)) \\ &= P(A \wedge B) + P(A \wedge \neg B) \end{aligned}$$

Beispiel aus der medizinischen Diagnostik

Hat ein Patient Schmerzen im rechten Unterbauch und erhöhten Leukozytenwert, so besteht der Verdacht auf eine akute Blinddarmentzündung (Appendizitis):

$$\textit{Bauchschmerzen re. u.} \wedge \textit{Leukozyten} > 10.000 \rightarrow \textit{Blinddarmentzündung}$$

Ereignis 1: Leukozytenwert größer als 10.000 *(Leuko)*

Ereignis 2: Patient hat Appendizitis *(App)*

$P(App, Leuko)$	App	$\neg App$
$Leuko$	0.23	0.31
$\neg Leuko$	0.05	0.41

$$\Rightarrow P(Leuko) = P(App, Leuko) + P(\neg App, Leuko) = 0.54$$

$$\Rightarrow P(Leuko|App) = \frac{P(Leuko, App)}{P(App)} = 0.82$$

Satz von Bayes

Satz von Bayes: KI-Anwendungen

Thomas Bayes (1701-1761)

Anwendungen in der künstlichen Intelligenz

- **Spam-Erkennung**
  - Von charakteristischen Wörtern in Emails (Ereignis A) wird auf die Eigenschaft, Spam zu sein (Ereignis B), geschlossen
  - Mit Maschinellem Lernen lässt sich der Filter verfeinern
- **Entscheidungsfindung**
  - Schließen mit Unsicherheit
  - nicht deduktiv und somit auch nicht immer korrekt, hat sich aber zur Hypothesenbildung und zum Lernen in Systemen als nützlich erwiesen
- **Qualitätsmanagement**
  - Beurteilung der Aussagekraft von Testreihen
  - Medizinische Diagnosen

Bayessches Schließen ist in der KI essentiell und weit entwickelt.

Bayes Formel

Satz von Bayes

Für zwei Zufallsvariablen A und B besteht ein Verhältnis zwischen der bedingten Wahrscheinlichkeit $P(A|B)$ und der umgekehrten Form $P(B|A)$, welches durch die Bayes-Formel zum Ausdruck gebracht wird.

Bayes-Formel

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

- $P(A)$: a-priori *(lat. frühere)* Wahrscheinlichkeit
- $P(A|B)$: a-posteriori *(lat. spätere)* Wahrscheinlichkeit

Umkehrung der Schlussfolgerung: Der Nutzer geht von einem bekannten Wert $P(B|A)$ aus, ist aber eigentlich an dem Wert $P(A|B)$ interessiert.

Probabilistisches Schließen

Der Satz von Bayes kann so interpretiert werden, dass zwischen den bedingten Wahrscheinlichkeiten für **Ursache** und **Wirkung** ein Verhältnis besteht:

$$P(Ursache|Wirkung) \textit{ steht im Verhältnis zu } P(Wirkung|Ursache).$$

Man kann auf zwei Arten probabilistisch schließen:

1. **Kausale Inferenz: von Ursache auf Wirkung**
   z.B. $P(Symptom|Krankheit)$, $P(JohnCallsBob|Alarm)$
2. **Diagnostische Inferenz: von Wirkung auf Ursache**
   z.B. $P(Krankheit|Symptom)$, $P(Alarm|JohnCallsBob)$

Der Satz von Bayes ist die Grundlage von KI-Systemen, die mit probabilistischer Inferenz arbeiten.

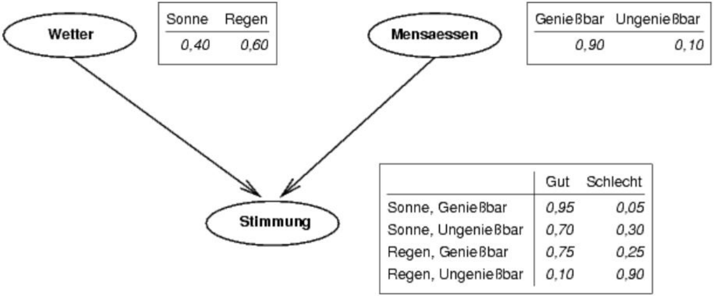
Bayes-Netz

Definition

Bayes-Netz

Ein Bayes-Netz ist ein **gerichteter azyklischer Graph** (engl. directed acyclic graph), in dem die **Knoten** Zufallsvariablen und die **Kanten** bedingte Abhängigkeiten zwischen den Variablen beschreiben.

Beispiel:



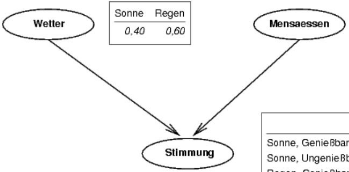
Sonne	Regen
0,40	*0,60*

Genießbar	Ungenießbar
0,90	*0,10*

	Gut	Schlecht
Sonne, Genießbar	*0,95*	*0,05*
Sonne, Ungenießbar	*0,70*	*0,30*
Regen, Genießbar	*0,75*	*0,25*
Regen, Ungenießbar	*0,10*	*0,90*

Kompakte Darstellung im Bayes-Netz

Jedem Knoten des Netzes ist eine **bedingte Wahrscheinlichkeitsverteilung** der durch ihn repräsentierten **Zufallsvariable,** gegeben die Zufallsvariablen an den **Elternknoten,** zugeordnet.
Diese werden durch **Tabellen der bedingten Wahrscheinlichkeit** *(engl. conditional probability table, CPT)* beschrieben, d.h. die Tabelle der bedingten Wahrscheinlichkeiten $P(A|Eltern(A))$:



Sonne	Regen
0,40	0,60

Genießbar	Ungenießbar
0,90	0,10

	Gut	Schlecht
Sonne, Genießbar	0,95	0,05
Sonne, Ungenießbar	0,70	0,30
Regen, Genießbar	0,75	0,25
Regen, Ungenießbar	0,10	0,90

Ein Bayes-Netz dient dazu, die **gemeinsame Wahrscheinlichkeitsverteilung** aller beteiligten Variablen unter Ausnutzung bekannter bedingter Unabhängigkeiten möglichst kompakt zu repräsentieren.

Bedingte Wahrscheinlichkeit bei Unabhängigkeit

Unabhängige Zufallsvariablen

$$P(X_1, ..., X_n) = P(X_1) \cdot ... \cdot P(X_n)$$

$\Rightarrow$ bedingte Wahrscheinlichkeiten werden trivial:

$$P(A \mid B) = \frac{P(A, B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$$

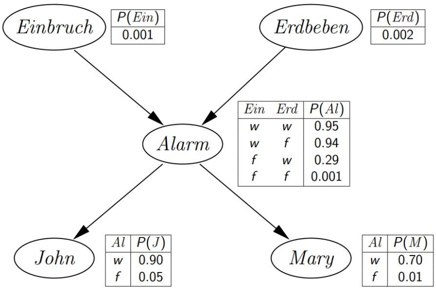
Interessanter wird die Welt, wenn nur ein Teil der Zufallsvariablen unabhängig beziehungsweise unter bestimmten Bedingungen unabhängig sind. Für das Schließen in der KI sind nämlich gerade die Abhängigkeiten zwischen Variablen wichtig und müssen genutzt werden.

Gemeinsame Wahrscheinlichkeitsverteilung der im Bayes-Netz vorkommenden Zufallsvariablen $X_1, ..., Xn$ wird ausgedrückt durch:

$$P(X_1, ..., X_n) = \prod_{i=1}^{n} P(X_i|Eltern(X_i))$$

Alarm-Beispiel

Der **Hausalarm von Bob** geht an bei **Einbruch** und manchmal bei **Erdbeben.**
$\rightsquigarrow$ Nachbarn John und Mary sollen bei Bob im Büro anrufen, wenn sie tagsüber Alarm hören.



Graphische Darstellung der Wissensbasis als Bayes-Netz modelliert auf Basis von Bobs jahrelanger Erfahrung

Anfrage: John ruft im Büro an, aber nicht Mary. Wie wahrscheinlich ist es, dass gerade eingebrochen wird? $P(J, \neg M, Al, Ein, \neg Erd)$

Wissensbasis im Beispiel

Wissensbasis modelliert basierend auf Bobs jahrelanger Erfahrung:

$$P(J|Al) = 0.90 \qquad P(M|Al) = 0.70$$
$$P(J|\neg Al) = 0.05 \qquad P(M|\neg Al) = 0.01$$

$$\begin{aligned} P(Al|Ein, Erd) &= 0.95 \\ P(Al|Ein, \neg Erd) &= 0.94 \\ P(Al|\neg Ein, Erd) &= 0.29 \\ P(Al|\neg Ein, \neg Erd) &= 0.001, \end{aligned}$$

$$P(Erd) = 0.002.$$

Beispielanfragen

$$P(Ein|J \vee M) \qquad P(J|Ein) \qquad P(M|Ein)$$

J	„John ruft an"	M	„Mary ruft an"	Al	„Alarmsirene ertönt
Ein	„Einbruch"	Erd	„Erdbeben"		

Lösung

Anfrage: John ruft im Büro an, aber nicht Mary. Wie wahrscheinlich ist es, dass gerade eingebrochen wird?

$$
\begin{aligned}
P(J, \neg M, Al, Ein, \neg Erd) &= P(J \mid Al) P(\neg M \mid Al) P(Al \mid Ein, \neg Erd) P(Ein) P(\neg Erd) \\
&= 0.9 \cdot 0.3 \cdot 0.94 \cdot 0.001 \cdot 0.998 \\
&\approx 0.00025 \\
&\approx 0.03\%
\end{aligned}
$$

Bedingte Unabhängigkeit

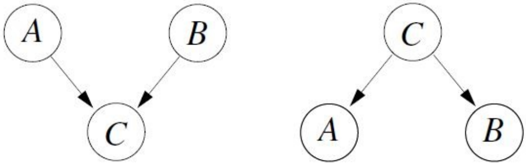


Abb.: Zwischen zwei Knoten A und B wird keine Kante eingetragen, wenn sie unabhängig (links, ohne Eltern) oder bedingt unabhängig sind (rechts)

Zwei Variablen A und B heißen **bedingt unabhängig,** gegeben C, wenn

$$P(A, B|C) = P(A|C) \cdot P(B|C)$$

Beispiele:

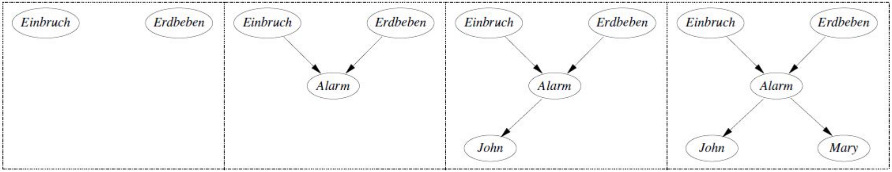
$$\mathbf{P}(J, M|Al) = \mathbf{P}(J|Al) \cdot \mathbf{P}(M|Al)$$

$$\mathbf{P}(J, Ein|Al) = \mathbf{P}(J|Al) \cdot \mathbf{P}(Ein|Al)$$

Aufbau eines Bayes-Netzes

Vorgehen

- Entwurf der Netzstruktur: meist manuell
- Eintragen der Wahrscheinlichkeiten in die CPTs: meist automatisch



Schrittweiser Aufbau des Alarm-Netzes unter Berücksichtigung der Kausalität

Beim Aufbau eines Bayes-Netzes sollten die Variablen im Sinne der **Kausalität** angeordnet werden. Zuerst die **Ursachen**, dann die **verdeckten Variablen** und zuletzt die **Diagnosevariablen** *(Symptome)*.

Youtube-Video Empfehlung

Bayesian Thinking Visual Guide

Julia Galef, Co-Founder of Center for Applied Rationality

Naive Bayes Klassifikator

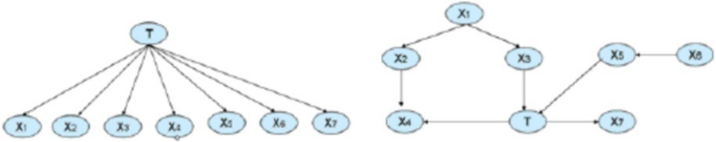
Übersicht

Bayes-Klassifikator

- aus dem Satz von Bayes hergeleiteter Klassifikator
- ordnet jedes Objekt der Klasse zu, zu der es mit der größten Wahrscheinlichkeit gehört

Naive Bayes-Klassifikator

- beliebt aufgrund schneller Berechenbarkeit und guter Erkennungsrate
- kann als sternförmiges Bayes-Netz betrachtet werden

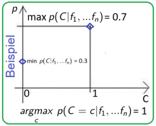


Naive Bayes From: pinterest.de Bayesian Network

Mathematische Funktion

Der **Bayes-Klassifikator** bildet Feature-Vektoren aus dem f-dimensionalen reellwertigen Merkmalsraum auf eine Menge von z.B. 2 Klassen C ab:

$$\hat{c}^{Bayes} : \mathbb{R}^f \to C = \{0, 1\} \quad , \qquad (f_1, \dots f_n) \mapsto c$$



$$\hat{c}^{Bayes}(f_1, \dots f_n) = \underset{c}{argmax}\ p(C = c|f_1, \dots f_n)$$

$$\overset{Bayes-Formel}{=} \underset{c}{argmax} \frac{p(C = c)p(f_1, \dots f_n|C = c)}{p(f_1, \dots f_n)}$$

$p(C)$: apriori Auftrittswahrscheinlichkeit der Klasse c
$p(f_1, \dots f_n|C = c)$ bedingte Auftrittswahrscheinlichkeit der Features $f_1, \dots f_n$ für die gegebene Klasse.

Naive Bayes-Klassifikator

$$\hat{c}^{Bayes}(f_1, \dots f_n) \overset{p(f_1,\dots f_n|C=c)=\prod_{i=1}^{n} p(f_i|C=c)}{=} \underset{c}{argmax} \frac{p(C = c)\prod_{i=1}^{n} p(f_i|C = c)}{p(f_1, \dots, f_n)}$$

Aposteriori Verteilung ist aus Wahrscheinlichkeiten aufgebaut, bei denen bei gegebener Klasse die Features unabhängig voneinander sind.

Beispiel: Entwicklung eines Spam-Filters

Gegeben: Wort W_i einer Email W

$$p(W_i|Spam) = \frac{\text{Anzahl Spam Emails mit dem Wort } W_i}{\text{Anzahl Spam Emails}}$$

$$p(W_i|Ham) = \frac{\text{Anzahl Ham Emails mit dem Wort } W_i}{\text{Anzahl Ham Emails}}$$

Gesucht: Ist die Wahrscheinlichkeit $p(Spam|W)$ größer oder kleiner als $p(Ham|W)$? ($C = \{Spam, Ham\}$)

$$Q = \frac{p(Spam|W)}{p(Ham|W)} \overset{Bayes-Formel}{=} \frac{p(W|Spam)p(Spam)}{p(W)} \cdot \frac{p(W)}{p(W|Ham)p(Ham)}$$

$$= \frac{p(W|Spam)p(Spam)}{p(W|Ham)p(Ham)} \overset{\text{Wörter unabh.}}{=} \frac{\prod_{i=1}^{n} p(W_i|Spam) \cdot p(Spam)}{\prod_{i=1}^{n} p(W_i|Ham) \cdot p(Ham)} \overset{?}{>} 1$$

mit $p(Spam) = \frac{\text{Anzahl Spam Emails}}{\text{Anzahl aller Emails}}$ und $p(Ham) = \frac{\text{Anzahl Ham Emails}}{\text{Anzahl aller Emails}}$.

Zeitreihenanalyse

Literaturempfehlung Zeitreihenanalyse

Übersicht

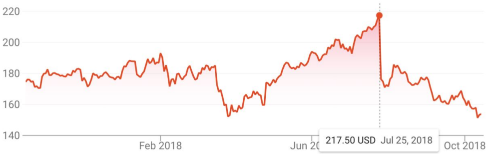
- ARIMA Zeitreihenanalyse, T. Reiter, URL: http://www.reiter1.com/Glossar/ARIMA.htm

Weiterführend

- Kapitel Zeitreihenanalyse in *Angewandte Statistik mit SPSS*, Prof. Dr. Peter P. Eckstein, Gabler Verlag, Springer, 2012

Zeitreihe

Eine **Zeitreihe** ist eine zeitlich geordnete Folge von Datenpunkten.



- Die einzelnen Zeitpunkte werden zu einer Menge von Beobachtungszeitpunkten T zusammengefasst, bei der für jeden Zeitpunkt $t \in T$ eine Beobachtung vorliegt.
- Ein Datenpunkt kann aus einer einzelnen Zahl (skalare Werte, univariate Zeitreihe) oder aus einer Mehrzahl (Tupel) von Zahlenwerten (vektorielle Werte, multivariate Zeitreihe) bestehen.

Beispiele

- Finanzmathematik: Börsenkurse, Liquiditätsentwicklungen
- Ökonometrie: Bruttosozialprodukt, Arbeitslosenquote
- Meteorologie: Temperatur, Windgeschwindigkeit

Zeitreihenanalyse und -ursachen

Zeitreihenanalyse

Inferenzstatistische Analyse von Zeitreihen und Vorhersage von Trends zu ihrer künftigen Entwicklung.

Typische Zeitreihen entstehen aus dem Zusammenwirken regelhafter und zufälliger Ursachen:

Regelhafte Ursachen

- variieren perdiodisch bzw. saisonal
- können langfristige Trends enthalten

Zufällige Ursachen

- Rauschen

Mathematische Definition von Zeitreihen

Gegeben sei ein T-dimensionaler Vektor von Zufallsvariablen $x_1, x_2, ..., x_T$ mit einer zugehörigen **univariaten** Verteilung.

Dies kann auch als eine Folge von Zufallsvariablen $\{x_t\}_{t=1}^{T}$ oder als **stochastischer Prozess** aufgefasst werden.
Eine Stichprobe daraus ergibt als ein mögliches Ergebnis die T reellen Zahlen.
Statt stochastische Prozesse der Dimension T anhand ihrer T-dimensionalen Verteilungsfunktion zu beschreiben, kann man sie durch die **Momente** erster und zweiter Ordnung erfassen, also durch

$$T \textbf{ Erwartungswerte: } E(x_i), \quad i = 1, 2, ..., T,$$

$$T \textbf{ Varianzen: } Var(x_i) = E((x_i - E(x_i))^2), \; i = 1, 2, ..., T,$$

$$\frac{T(T-1)}{2} \textbf{ Kovarianzen: } Cov(x_i, x_j) = E((x_i - E(x_i))(x_j - E(x_j))), \; i < j.$$

ARMA-Modell

ARMA: Auto Regressive Moving Average

- Autoregressiver gleitender Mittelwert
- Zeitdiskrete Modelle für die Abbildung linearer stochastischer Prozesse

Box-Jenkins-Ansatz

- Spezifikation
- Schätzung
- Validierung

Allgemeine Form

- Lineare Differenzengleichungen

ARMA Modelltypen

Auto Regressives Modell

- Lineares Regressionsmodell, das den Wert der abhängigen Variablen y_t basierend auf den Werten der p vorangegangenen Zeitintervallen *(=erklärende Variablen y_{t-i} an z.B. Tag t)* vorhersagt:

$$y_t = \alpha_1 y_{t-1} + \ldots + \alpha_p y_{t-p} \qquad ,\alpha_i : \textit{Regressionskoeffizienten}$$

Moving Average Modell *("Gleitender Mittelwert")*

- Aktueller und vergangene zeitlich voneinander unabhängige Rauschterme ϵ_{t-j} als erklärende Variablen:

$$y_t = \epsilon_t + \beta_1 \epsilon_{t-1} + \ldots + \beta_q \epsilon_{t-q} \qquad ,\beta_j : \textit{Regressionskoeffizienten}$$

ARMA: Summe der beiden Modelle

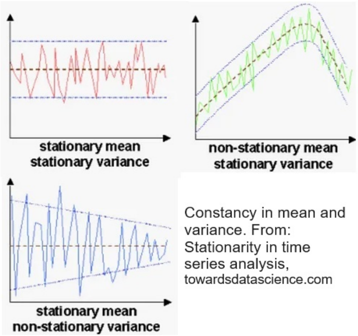
$$y_t = \alpha_1 y_{t-1} + \ldots + \alpha_p y_{t-p} + \epsilon_t + \beta_1 \epsilon_{t-1} + \ldots + \beta_q \epsilon_{t-q}$$

Voraussetzung Stationarität für Zeitreihenanalyse

Stationärer Prozess

Ein stochastischer Prozess $(x_t)_{t \in \mathbb{T}}$ heißt (schwach) **stationär,** wenn

1. der Erwartungswert $E(x_t) = \mu$ konstant ist
2. die Varianz $Var(x_t) < \infty$ endlich und konstant ist
3. die Autokovarianz nicht von der Verschiebung $s \in \mathbb{T}$ abhängt



Constancy in mean and variance. From: Stationarity in time series analysis, towardsdatascience.com

- **ARIMA-Modelle** für nichtstationäre Zeitreihen werden aus **schwach stationären ARMA-Modellen** konstruiert.
- **Beispiel zur Herstellung von Stationarität:** Hat ein Trend die Form eines Polynoms n-ter Ordnung $trend(x) = a_0 + a_1 x + a_2^2 x^2 + \ldots + a_n x^n$ dann lässt er sich durch n-faches Differenzieren beseitigen.

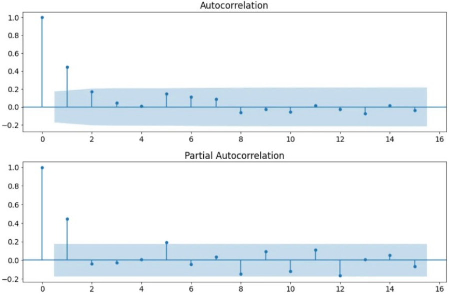
Bestimmung der Modellordnungen

Autokorrelationsfunktion (engl. ACF)

- Korrelation zwischen einer Zeitreihe mit einer zeitlich verschobenen *(engl. lagged)* Version derselben Zeitreihe.
- Wird für die Bestimmung der Ordnung q des MA-Modells verwendet.

Partielle Autokorrelationsfunktion (engl. PACF)

- Autokorrelation berücksichtigt zusätzlich indirekte Korrelationen.
- Wird für die Bestimmung der Ordnung p des AR-Modells verwendet.



ACF and PACF „Lollipop“ plot (Kaggle)

- Sowohl ACF und PACF Plot beginnen mit lag 0 (=Korrelation der Zeitreihe mit sich selbst), was eine Korrelation von 1 ergibt.
- Blaue Bereiche: 95% Konfidenzintervalle

Bestimmung der Modellordnungen

- Zählen der Lollipops ober- und unterhalb des Konfidenzintervalls, bevor der nächste Lollipop in den blauen Bereich eintritt. **Beispiel links:** $q = p = 1$

Bestimmung Regressionskoeffizienten in Python

```
from statsmodels.tsa.ar_model import AutoReg
ar_model = AutoReg(X_train, lags = 1).fit()
ar_model.summary()
from statsmodels.tsa.arima_model import ARMA
ma_model = ARMA(X_train, order = (0, 1)).fit()
ma_model.summary()
```

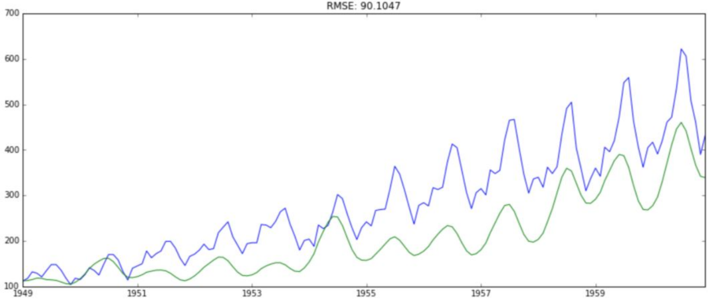
Demo

A comprehensive beginner's guide to create a Time Series Forecast (with Codes in Python and R)

Author: Aarshay Jain, Analytics Vidhya, 2016

Hinweis zur Installation:

```
python -m pip uninstall statsmodels
python -m pip install statsmodels==0.12.2
```



ARIMA-Modell für eine nicht-stationäre Zeitreihe von Flugpassagierdaten

Lernkontrolle

Meine 3 Lernziele für heute waren

1. Ich kenne den Satz von Bayes und kann ihn zur Anwendung bringen.
2. Ich bin in der Lage, ein Bayes-Netz zu interpretieren und kann daraus Wahrscheinlichkeiten für das Eintreten von Ereignissen berechnen.
3. Ich kenne die Voraussetzungen für stationäre Zeitreihen und das Vorgehen zur Bestimmung der Modellordnungen im ARIMA-Modell.

Python für Maschinelles Lernen

Übersicht

Python

Universelle, üblicherweise interpretierte, höhere Programmiersprache.

Programmierparadigmen

- objektorientiert
- aspektorientiert
- funktional

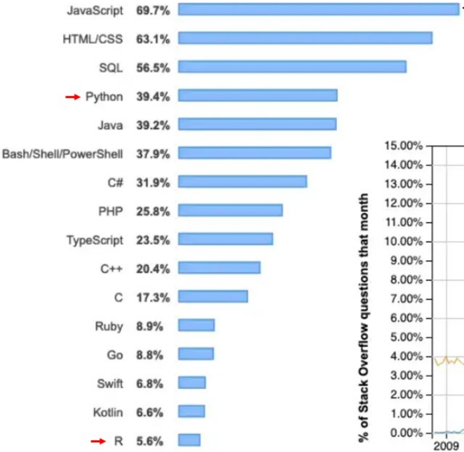
Typisierung

- dynamisch
- stark

Warum Python?

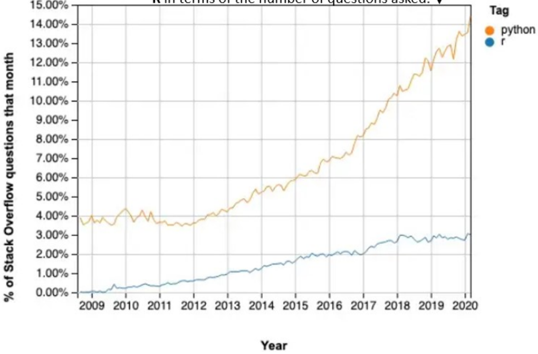
- **Einfachheit**
  - Entwickler können schnell Konzepte und Funktionalitäten lernen
    $\Rightarrow$ finanzieller Vorteil aus Unternehmenssicht
- **Geschwindigkeit und Performanz**
  - für Skalierbarkeit,
  - Kundenzufriedenheit bei z.B. Echtzeit Anwendungsfällen und
  - geringe Trainingszeit
    $\rightarrow$ Cython für C-ähnliche Performanz
- **Schnelles Deployment**
  - beim Übergang von Dev auf Prod (Dev/Ops) kann das Modell den Kunden schnell zur Verfügung gestellt werden
- **Etablierung** der Programmiersprache
- große und aktive **Community** für Lösungsansätze und Unterstützung bei z.B. Bug-Fixing
- **3rd Party Integration** für mehr Funktionalität und Kompatibilität

Python vs. R

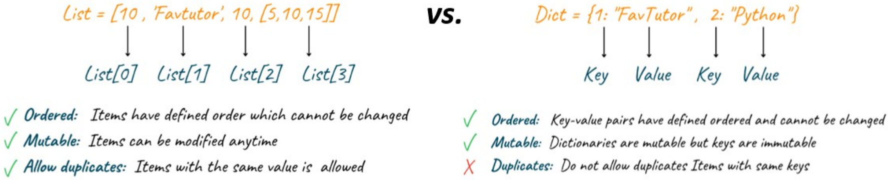


Stack Overflow 2019 Developer Survey: **Python** is the 4th most popular programming language among approx. 70.000 professional developers, even more popular than Java recently. In the same survey, **R** is in the 16th position.

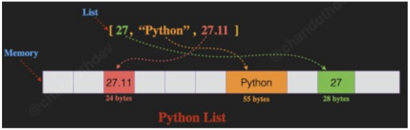
Stack Overflow Trends: **Python** is more popular than **R** in terms of the number of questions asked.



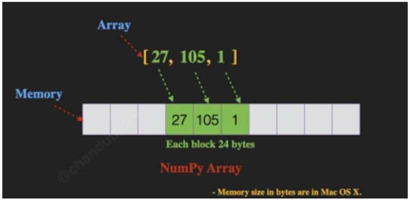
Listen, Dictionaries und Arrays in Python



From: www.favtutor.com



VS.



From: www.dev.to, Chanduthedev

Python Libraries für maschinelles Lernen

- Pandas: Datenmanipulation and -Analyse
- NumPy: Lineare Algebra
- Scikit-Learn: Klassifikation, Regression,...
- Matplotlib: Datenvisualisierung
- Plotly: Datenvisualisierung (z.B. Charts)
- Seaborn: Statistische Graphen
- Tensorflow: Multidimensionale Arrarys (z.B. Künstliche neuronale Netze)
- Keras: Künstliche neuronale Netze
- NLTK: Natural Language Toolkit (z.B. NLP)
- PyTorch: analog zu NumPy (z.B. Computer Vision)

Aus: www.fireblazeaischool.in