

Methoden des maschinellen Lernens:

Aufgabenblatt 3

SS 2018

Abgabe: 16.05.2018

Aufgabe 1 (5 Punkte)

Aus der Drehtür eines Geschäftes kommen die Menschen einzeln heraus. In der folgenden Tabelle steht, wie viel Geld die Personen in dem Geschäft ausgegeben haben.

	Unter 10 €	10 bis 100€	Über 100€
Frauen	0,07	0,2275	0,0525
Männer	0,1625	0,26	0,2275

Mit welcher Wahrscheinlichkeit....

- a) ist die erste Person eine Frau?
- b) kommen zunächst 3 Frauen und dann ein Mann durch die Tür?
- c) kommt ein Kunde durch die Tür, der über 100€ ausgegeben hat?
- d) kommt ein Mann durch die Tür, wenn bekannt ist, dass der Kunde unter 10€ ausgegeben hat?
- e) kommt eine Frau durch die Tür, wenn bekannt ist, dass der Kunde 10€ oder mehr ausgegeben hat?

Aufgabe 2 (5 Punkte)

Eine Testgruppe wird einem Hepatitis-Testverfahren unterzogen. Der Anteil der tatsächlich kranken Personen beträgt 0,1%. Das verwendete Testverfahren erkennt 95% der Kranken richtig und stuft 2% der gesunden Personen als krank ein. Wie hoch ist die Wahrscheinlichkeit, dass eine positiv getestete Person tatsächlich krank ist?

Aufgabe 3 (5 Punkte)

Gegeben sind zwei Beutel, im ersten Beutel befinden sich 60 rote Kugeln und 40 blaue. Im zweiten Beutel befinden sich 30 rote Kugeln und 70 blaue. Einer der beiden Beutel wird zufällig gewählt und eine Kugel wird gezogen. Wenn die gezogene Kugel rot ist, wie hoch ist die Wahrscheinlichkeit, dass aus dem ersten Beutel gezogen wurde?

Aufgabe 4 (5 Punkte)

Gegeben sind folgende Trainingsdaten:

Größe	Gewicht	Haarlänge	Geschlecht
mittel	normal	kurz	männlich
klein	leicht	lang	weiblich
groß	schwer	kurz	männlich
klein	normal	lang	weiblich
groß	normal	lang	weiblich
klein	leicht	kurz	weiblich
klein	schwer	kurz	männlich
mittel	normal	kurz	weiblich
mittel	leicht	lang	weiblich
groß	normal	kurz	männlich

- In welcher Beziehung ist der Naive Bayes „naiv“?
- Klassifizieren Sie folgende Merkmalsvektoren und kommentieren Sie kurz:
 - <groß,leicht,lang>
 - <klein,schwer,kurz>
 - <mittel,leicht,lang>

Aufgabe 5 (5 Punkte)

- Erläutern Sie die Funktionsweise eines Laplace Schätzers
- Wiederholen Sie die Klassifikation der Merkmalsvektoren aus Aufgabe 4 mit einer Laplace Glättung, bei $m=5$.

Aufgabe 6 (10 Punkte)

- Erläutern Sie stichpunktartig die Vorgehensweise des ID3 Algorithmus
- Implementieren Sie ein Programm, das die ID3_Beispieldaten.csv aus dem Ilias einliest, mithilfe des ID3 Algorithmus einen Entscheidungsbaum daraus generiert und diesen ausgeben kann.

Aufgabe 7 (10 Punkte)

Arbeiten Sie sich in den C4.5 Algorithmus ein

- Erläutern Sie die maßgeblichen Veränderungen des C4.5 Algorithmus im Vergleich zum ID3.
- Welchen Vorteil hat die information gain ratio gegenüber dem information gain?
- Erweitern Sie Ihr Programm aus Aufgabe 6 dahingehend, dass es zur Bestimmung des stärksten Attributs die information gain ratio benutzt.
- Erweitern Sie Ihr Programm aus Aufgabe 6 um die Diskreditierungsstrategie des C4.5 Algorithmus
- Verwenden Sie das erweiterte Programm um aus den C45_Beispieldaten.csv einen Entscheidungsbaum zu generieren

Aufgabe 8 (30 Punkte)

Machen Sie sich mit den Grundlagen von Tensorflow vertraut.

Guter Startpunkt: https://www.tensorflow.org/get_started/eager

Schreiben Sie ein Skript das folgende Aufgaben erfüllt

- a) Importieren Sie das Iris Dataset.
- b) Verwenden Sie die Scikit Flow Bibliothek um 5 neue Merkmalsvektoren mit einem Gaussian Naive Bayes Klassifikator zu klassifizieren
- c) Visualisieren Sie die Trainingsdaten und die neu klassifizierten Merkmalsvektoren mit Hilfe von Tensorboard
- d) Implementieren Sie Ihren eigenen Naive Bayes Klassifikator, verwenden Sie hierzu entweder Diskreditierung oder die Gaußsche Dichtefunktion