

Übungen zu Ideen der Informatik

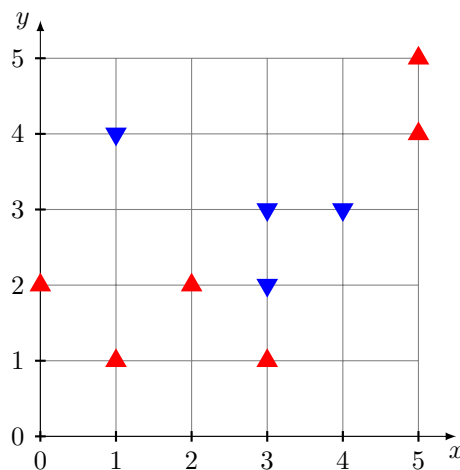
<https://www.mpi-inf.mpg.de/departments/algorithms-complexity/teaching/winter21/ideen/>

Blatt 11: Künstliche Intelligenz I

Abgabeschluss: 24.01.2022

Aufgabe 1 kNN-Klassifikation (10 Punkte)

Betrachten Sie die folgende Abbildung, die Daten mit zwei numerischen Attributen als Punkte in zwei Dimensionen darstellt. Die Klassen der Datenpunkte sind Ihnen bekannt: rote Dreiecke mit Spitze nach oben gehören zu Klasse 1, blaue Dreiecke mit Spitze nach unten gehören zu Klasse 2.



Die *Manhattan-Distanz* L_1 zwischen zwei Punkten a und b mit n Dimensionen ist die Summe der absoluten Differenzen zwischen ihren jeweiligen Koordinaten: $L_1 = \sum_{i=1}^n |a_i - b_i|$. Für zwei Punkte a und b mit jeweils zwei Dimensionen berechnet sich die Manhattan-Distanz also als $L_1 = |a_1 - b_1| + |a_2 - b_2|$.

Beantworten Sie nun die folgenden Fragen (jeweils mit Begründung oder Rechenweg):

a) Berechnen Sie die Manhattan-Distanz zwischen den folgenden Punkten. (3 Punkte)

- (1) (0,2) und (1,1)
- (2) (3,1) und (4,3)
- (3) (1,4) und (5,5)

b) Erklären Sie, warum die Manhattan-Distanz *Manhattan Distance*, *Cityblock Distance* oder *Taxi Cab Metric* genannt wird. (1 Punkt)

c) Klassifizieren Sie die folgenden *neuen* Punkte mit dem 1NN-Klassifikator auf Basis der Manhattan-Distanz (d.h. jeder Punkt erhält die Klasse seines nächsten Nachbarn). (2 Punkte)

- (1) (3,4)
- (2) (5,0)

d) Welches Problem stellt sich bei der Klassifikation des Punkts (2,3), und nach welcher Regel könnten Sie dieses Problem lösen? (2 Punkte)

e) Angenommen, Sie erhalten einen neuen Punkt (2,1) mit *bekannter* Klasse 2, den Sie bei der Klassifikation von Punkten mit unbekanntem Klassen mit dem 1NN-Klassifikator auf Basis der Manhattan-Distanz berücksichtigen wollen.

- (1) Wie wird der Punkt (2,0) klassifiziert, wenn Sie den Punkt (2,1) nicht berücksichtigen, und wie wird er klassifiziert, wenn Sie den Punkt berücksichtigen? (1 Punkt)
- (2) Welches Problem illustriert der in der vorigen Teilaufgabe betrachtete Fall und wie ließe sich dieses lösen? (1 Punkt)

Aufgabe 2 Bayessche Regel (10 Punkte)

2 % der Bevölkerung sind mit einem Virus infiziert.
X-Tests dienen dem Nachweis dieses Virus.

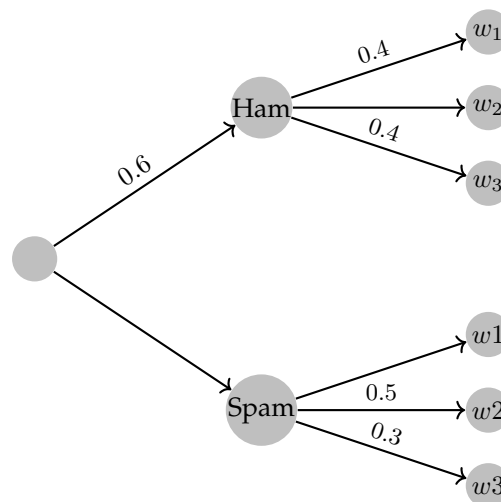
5 % der Infizierten haben einen fälschlicherweise negativen Test.
2 % der Nicht-Infizierten haben einen fälschlicherweise positiven Test.

Geben Sie bei der Bearbeitung der folgenden Aufgabenteile jeweils Ihren vollständigen Rechenweg an.

- a) Nehmen Sie an, Sie lassen sich testen und erhalten ein *positives* Testergebnis.
Was ist die Wahrscheinlichkeit, dass Sie infiziert sind? (5 Punkte)
- b) Nehmen Sie an, Sie lassen sich testen und erhalten ein *negatives* Testergebnis.
Was ist die Wahrscheinlichkeit, dass Sie infiziert sind? (5 Punkte)

Aufgabe 3 Spam (10 Punkte)

Wir benutzen einen Bayes-Filter zur Klassifikation von Emails. Wir nehmen der Einfachheit halber an, dass Emails immer nur aus einem Wort bestehen und zwar aus einem der drei Worte w_1 , w_2 und w_3 . Das generative Modell ist wie in folgender Abbildung.



Der Filter erklärt eine Email, die aus dem Wort w_i besteht, zu Spam, wenn die Spam-Wahrscheinlichkeit, gegeben w_i , mindestens 0.5 beträgt:

$$p(\text{Spam}|w_i) \geq 0.5.$$

Beantworten Sie zu diesem Szenario die folgenden Fragen und geben Sie dabei stets Ihren vollständigen Rechenweg an.

- a) Ergänzen Sie die fehlenden Wahrscheinlichkeiten im oben angegebenen Modell. (3 Punkte)

- b) Geben Sie für w_1 , w_2 und w_3 jeweils an, ob diese zu Spam erklärt werden. (3 Punkte)
- c) Mit welcher Wahrscheinlichkeit wird eine Ham-Email als Spam klassifiziert?
(2 Punkte)
- d) Mit welcher Wahrscheinlichkeit wird eine Spam-Email als Ham klassifiziert?
(2 Punkte)

Ich habe für die Videos, die Nachbereitung und das Übungsblatt etwa Stunden gebraucht.
(Ann-Sophie fertigt aus diesen Zahlen eine Statistik an. Kurt und Corinna sehen nur diese Statistik. Wir möchten wissen, ob der Schwierigkeitsgrad in etwa richtig ist.)

Künstliche Intelligenz I war spannend okay langweilig
schwierig okay einfach