

Prof. Dr. Kurt Mehlhorn  
Michael Dirnberger

WiSe 2015/16

## Übungen zu Ideen der Informatik

<http://www.mpi-inf.mpg.de/departments/algorithms-complexity/teaching/winter15/ideen/>

Blatt 8

Abgabeschluss: 4.1.2016

**Aufgabe 1** (5 Punkte) 100 von 10000 (1%) Frauen über vierzig haben Brustkrebs. Wir machen Mammographien um Brustkrebs nachzuweisen. 900 von 1000 (90%) Frauen mit Brustkrebs haben eine positive Mammographie. Nur 495 von 9900 (5%) Frauen ohne Brustkrebs haben eine (fälschlicherweise) positive Mammographie.

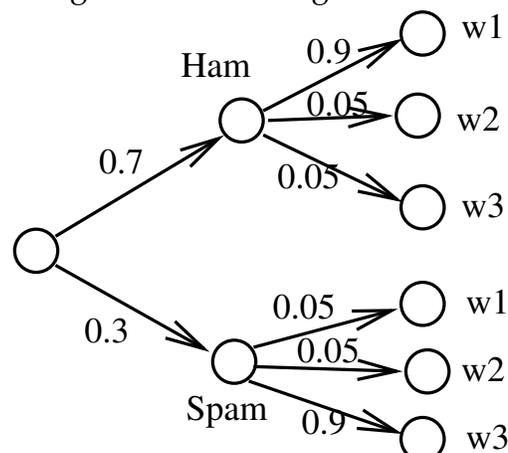
Nehmen Sie an, sie seien eine Frau über vierzig und erhalten ein positives Mammographieergebnis. Was ist die Wahrscheinlichkeit, dass Sie Brustkrebs haben?

Männliche Hörer ersetzen Brustkrebs durch Prostatakrebs.

**Aufgabe 2** (5 Punkte) Wir wollen, dass unser Klassifikator nicht nur aus der anfänglichen Trainingsmenge lernt, sondern jedes neue Beispiel, das zu klassifizieren ist, auch in die Trainingsmenge aufgenommen wird. Dadurch verbessert sich unser Klassifikator kontinuierlich.

- Wir verwenden den Klassifikator, der die Zentren vorberechnet. Welche zusätzlichen Daten (neben den Klassenzentren) müssen wir speichern um nicht für jedes neue Beispiel alle Trainingsdaten erneut verarbeiten zu müssen um die Zentren anzupassen?
- Wie könnte ein Update-Verfahren für einen k-Means Klassifikator aussehen?

**Aufgabe 3** (10 Punkte) Wir benutzen einen Bayes-Filter zur Klassifikation von emails. Wir achten auf das Vorkommen von drei Worten  $w_1$ ,  $w_2$  und  $w_3$  in den emails. Wir nehmen der Einfachheit halber an, dass in den emails genau eines dieser drei Worte vorkommt. Das generative Modell ist wie in folgender Abbildung.



Die email, die nur aus dem Wort  $w_1$  besteht, wird also mit Wahrscheinlichkeit  $0.7 \cdot 0.9 + 0.3 \cdot 0.05$  erzeugt. Entsprechend für  $w_2$  und  $w_3$ .

Der Filter erklärt eine email, die aus dem Wort  $w_i$  besteht zu Spam, wenn  $\text{prob}(\text{Spam}|w_i) \geq 0.8$ .

- a) Wird eine email, die aus dem Wort  $w_1$  besteht zu Ham erklärt? Ja/Nein. Analog für  $w_2$  und  $w_3$ .
- b) Welche Wege im Modell stehen für emails, die falsch klassifiziert werden.
- c) Mit welcher Wahrscheinlichkeit wird eine Ham-Email als Spam klassifiziert?
- d) Mit welcher Wahrscheinlichkeit wird eine Spam-Email als Ham klassifiziert?

**Aufgabe 4** (10 Punkte) Das TCAS System zur Vermeidung von Kollisionen im Luftverkehr ist so zuverlässig, dass Piloten angewiesen sind, TCAS Anweisungen zu befolgen auch wenn sie sich mit den menschlichen Anweisungen des Towers widersprechen.

Moderne Autos haben bereits viele ähnliche Features. Sie parken automatisch ein, halten konstante Geschwindigkeiten und Abstände<sup>1</sup>, warnen beim Überfahren von Markierungen, geben Alarm wenn sich ein Auto im toten Winkel befindet und helfen bei Vollbremsungen.

Diskutieren Sie die Vorteile und Risiken autonom fahrender Autos. Was wenn die Autopiloten, wie TCAS, zuverlässiger sind als Menschen? Sie sind weder betrunken, noch werden sie müde, oder lassen sich vom Handy ablenken. Aber wer haftet im Falle eines Unfalls?

**Aufgabe 5** (5 Punkte) Kurts kleine Knochelei: Alice und Bob möchten sich auf das Ergebnis eines Münzwurfs einigen. Sie sind allerdings räumlich getrennt, kennen aber die Segnungen der modernen Kryptographie.

Hinweis: Alice und Bob werfen jeweils eine Münze (oder denken sich ein Bit aus). Das Ergebnis des gemeinsamen Münzwurfs soll das Exclusive-Oder der beiden Würfe sein, also 0, wenn beide das gleiche Bit gewählt haben, und 1, wenn sie verschiedene Bits gewählt haben.

Alice wählt einen zufälligen Bitstring  $k$  der Länge 128 und benutzt ihn als Schlüssel in einem symmetrischen Verfahren um ihr Bit zu verschlüsseln. Sie schickt das verschlüsselte Bit an Bob.

Wie geht es weiter? Kann man auf diese Art auch Karten spielen?

---

<sup>1</sup>z. B. <http://www.youtube.com/watch?v=FP71EaXDT1I>