

## Übungen zu Ideen der Informatik

<https://www.mpi-inf.mpg.de/departments/algorithms-complexity/teaching/winter19/ideen/>

Blatt 11

Abgabeschluss: 13.1.2020

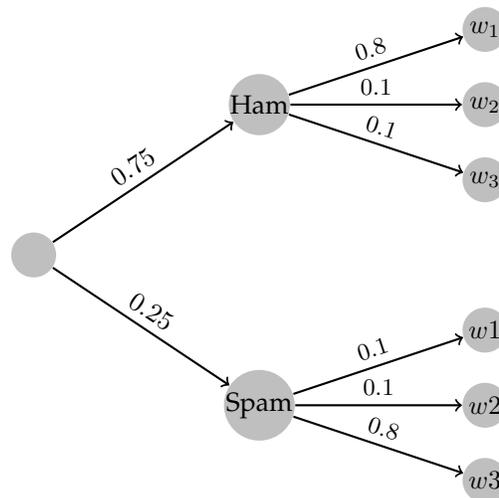
**Aufgabe 1** (10 Punkte) 100 von 10000 (1%) Frauen über vierzig haben Brustkrebs. Wir machen Mammographien um Brustkrebs nachzuweisen. 900 von 1000 (90%) Frauen mit Brustkrebs haben eine positive Mammographie. Nur 90 von 9900 (weniger als 1%) Frauen ohne Brustkrebs haben eine (fälschlicherweise) positive 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.

Die oben angegebenen Wahrscheinlichkeiten entsprechen ungefähr der Wirklichkeit.

**Aufgabe 2** (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.



Die email, die nur aus dem Wort  $w_1$  besteht, wird also mit Wahrscheinlichkeit  $0.75 \cdot 0.8 + 0.25 \cdot 0.1$  erzeugt. Entsprechend für  $w_2$  und  $w_3$ .

Der Filter erklärt eine email, die aus dem Wort  $w_i$  besteht zu Spam, wenn  $prob(Spam|w_i) \geq 0.7$ .

(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?

Hinweis: Sie können diese Aufgabe auf zwei Arten bearbeiten. Entweder wenden Sie die Bayessche Regel an. Oder Sie betrachten eine Kollektion von 1000 Emails. Davon sind nach dem Model 750 Ham und 250 Spam.

