1 Morde

Es begab sich einst in einer kleinen Stadt, dass eine Reihe von Morden die Bürger in Angst und in Schrecken versetzte. (Nein, wir spielen nicht das Mafia-Spiel.) Der findige Detektiv Miller war auch jedes mal nach dem Mord am Tatort, um nach Hinweisen auf den Täter zu suchen. Er fand jedesmal welche, doch dummerweise waren diese Hinweise nie eindeutig. Es waren immer Gegenstände, die genau drei Personen hinterlassen haben könnten. Ausserdem wusste Miller nicht, wieviele Mörder sich nun genau in der Stadt rumtrieben... Er versucht nun, einen Weg zu finden, für jeden einzelnen Menschen, die Wahrscheinlichkeit zu berechnen, dass dieser Mensch ein Mörder ist.

So ungefähr sieht es bei mir aus. Ich habe also an so etwas gedacht: Wir numerieren alle Menschen durch:

Sei $m_i \in M$ die Erkenntnis, dass der Mensch i ein Mörder ist. Die Stadt hat ca 10000 Einwohner. Eine Erkenntnis

$$H_x = m_a \vee m_b \vee m_c$$

enthält drei mögliche Menschen, die die Tat begangen haben können. Da Miller nach einiger Zeit viele solcher Hinweise gesammelt hat und Personen sucht, auf die alle Hinweise zutreffen, definieren wir noch

$$H = H_1 \wedge H_2 \wedge H_3 \ldots$$

Was ich suche, ist die Wahrscheinlichkeit $p(m_i|H)$ und mein erster Ansatz war der Satz von Bayes:

$$p(m_i|H) = \frac{p(m_i) \cdot p(H|m_i)}{p(H)}$$

Dann ist nämlich

$$p(m_i) = \frac{1}{10000}$$

und

$$p(H|m_i) = \begin{cases} 1, \text{ wenn } m_i \text{ in jedem Hinweise } H_1 \ldots H_n \text{ vorkommt} \\ 0, \text{ andernfalls.} \end{cases}$$

Das Problem ist: Was ist $p(H)$? Logisch gesehen, würde ich sagen: Wir müssen die Einsen zählen. Ich stelle also eine Wertetabelle auf, trage für alle möglichen Kombinationen von m_a, $\overline{m_a}$ das Ergebnis von $H_1 \wedge H_2 \wedge H_3 \ldots$ ein und zähle, in wieviel Prozent der Zeilen jetzt eine 1 steht. Im Grunde ist das theoretisch möglich, praktisch müsste ich wohl $2^1 0000$ kombinationen durchprobieren...

Wenn es da nicht einen guten Algorithmus gibt, muss ich wohl einfach eine Unabhängigkeitsannahme machen. Beim Naive-Bayes funktioniert das ja auch. (Ich glaube, wenn ich richtig überlege: Wenn ich einen effizienten Algorithmus entwickeln würde, würde ich vermutlich den Turingpreis bekommen, weil man damit dann auch das 3KNF-SAT Problem lösen könnte und somit $P = NP$ wäre. Aber das ist ja nicht Thema meiner Diplomarbeit.) Ich werde also einfach annehmen, dass H_1 und H_2 und H_3 etc. unabhängig voneinander sind. Dann kommt vielleicht keine exaktes Ergebnis raus, aber als Abschätzung sollte es langen.

Wenn das Problem gelöst ist, müsste ich am Ende die Wahrscheinlichkeit für jeden Menschen berechnen können, und zwar unter der Annahme, dass es 1 Mörder gibt,

$$p(m_i|H)$$

oder unter der Annahme, dass es 2 Mörder gibt,

$$p(m_i \wedge m_j|H) = p(m_i|H) \cdot p(m_j|H)$$

oder beliebig viele...

$$p(\bigwedge_{m_i \in M} m_i|H) = \prod_{m_i \in M} p(m_i|H)$$

wobei ich da wohl gerade auch wieder die Annahme getroffen habe, dass m_i unabhängig von m_j ist, was ja im Prinzip nicht ganz stimmt. Zumindest $p(m_i)$ verändert sich von $\frac{1}{10000}$ auf $\frac{1}{9999}$.