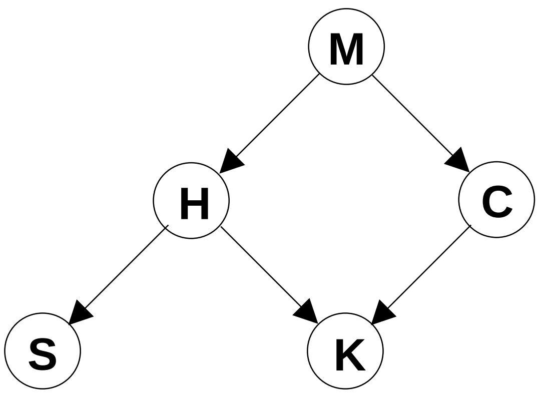
Aufgabe 1 – Probabilistische Inferenz (30 Punkte)

In einer medizinischen Studie wurden die Auswirkungen von Metastasen bildenden Karzinomen untersucht. Dabei wurde folgendes festgestellt:

- Bei den in der Studie untersuchten, an Krebs erkrankten Patienten traten nur in 20% aller Fälle Metastasen ($M = w$) auf. Bei den anderen 80% bildeten sich keine Tochtergeschwulste ($M = f$).
- Metastasen ($M = w$) lassen sich mit einer Wahrscheinlichkeit von 80% an einer stark erhöhten Calcium-Konzentration ($C = w$) im Blut erkennen. Andere Ursachen (d. h. $M = f$) können ebenfalls dieses Symptom ($C = w$) in 20% der Fälle hervorrufen.
- Falls sich ein Tochtergeschwulst bildet ($M = w$), ist es mit Wahrscheinlichkeit 20% ein Hirntumor ($H = w$). Bei 5% der Patienten ohne Metastasen ($M = f$) wurde dagegen ein Hirntumor ($H = w$) gefunden, der unabhängig von der ursprünglichen Krebserkrankung entstanden war.
- Tritt bei einem Patienten ein Hirntumor oder eine stark erhöhte Calcium-Konzentration ($H = w \vee C = w$) auf, so liegt das Risiko, dass er ins Koma ($K = w$) fällt bei 80%. Andernfalls ($H = f \wedge C = f$) beträgt die Wahrscheinlichkeit hierfür nur 5%.
- Hirntumore ($H = w$) führen in 80% der Fälle zu starken Kopfschmerzen ($S = w$). Ohne Hirntumor ($H = f$) tritt dieses Symptom mit 60% Wahrscheinlichkeit auf.

Die angegebenen Wahrscheinlichkeiten beziehen sich alle auf die Teilnehmer der Studie.

(a) Zeichnen Sie ein Bayes-Netz, das zu diesem Modell passt und eine Auswertung ohne zusätzliche Informationen erlaubt! (6 Punkte)



(b) Wie groß ist die Wahrscheinlichkeit, dass ein zufällig ausgewählter Proband einen Hirntumor hat? (6 Punkte)

$$\begin{aligned} P(H=w) &= P(H=w|M=w)P(M=w) + P(H=w|M=f)P(M=f) \\ &= 0.20 \cdot 0.20 + 0.05 \cdot 0.80 = 0.04 + 0.04 = 0.08 \end{aligned}$$

(c) Ein Teilnehmer der Studie klagt über starke Kopfschmerzen. Wie wahrscheinlich ist es, dass gerade er einen Hirntumor hat? (8 Punkte)

$$\begin{aligned} P(H=w|S=w) &= \frac{P(S=w|H=w)P(H=w)}{P(S=w|H=w)P(H=w) + P(S=w|H=f)P(H=f)} \\ &= \frac{0.80 \cdot 0.08}{0.80 \cdot 0.08 + 0.60 \cdot (1-0.08)} = \frac{0.064}{0.064 + 0.552} \approx 0.104 \end{aligned}$$

(d) Berechnen Sie das Risiko, ins Koma zu fallen, für einen Patienten, bei dem die Calcium-Konzentration im Blut nicht erhöht ist! (10 Punkte)

$$\begin{aligned} P(M=w|C=f) &= \frac{P(C=f|M=w)P(M=w)}{P(C=f|M=w)P(M=w) + P(C=f|M=f)P(M=f)} \\ &= \frac{0.20 \cdot 0.20}{0.20 \cdot 0.20 + 0.80 \cdot 0.80} = \frac{0.04}{0.04 + 0.64} \approx 0.059 \end{aligned}$$

$$\begin{aligned} P(K=w|C=f) &= P(K=w|H=f,C=f)P(H=f|M=f)P(M=f|C=f) \\ &+ P(K=w|H=f,C=f)P(H=f|M=w)P(M=w|C=f) \\ &+ P(K=w|H=w,C=f)P(H=w|M=f)P(M=f|C=f) \\ &+ P(K=w|H=w,C=f)P(H=w|M=w)P(M=w|C=f) \\ &= 0.05 \cdot (1-0.05) \cdot (1-0.059) + 0.05 \cdot (1-0.20) \cdot 0.059 \\ &+ 0.80 \cdot 0.05 \cdot (1-0.059) + 0.80 \cdot 0.20 \cdot 0.059 \\ &\approx 0.045 + 0.002 + 0.038 + 0.009 = 0.094 \end{aligned}$$

Aufgabe 2 – Hidden-Markov-Modell (30 Punkte)

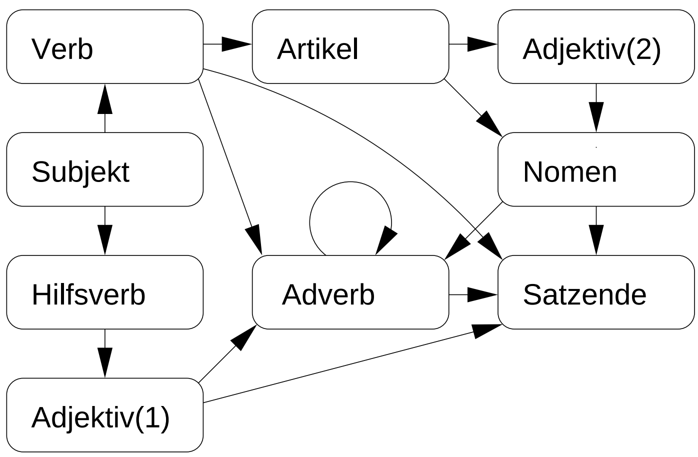
Bei der Spracherkennung werden Wortmodelle eingesetzt, um die Wahrscheinlichkeit von Wortfolgen zu berechnen. Um die Grammatik besser berücksichtigen zu können, wird zwischen einer Folge von Wortarten $x_1, x_2, x_3, \ldots$ und der tatsächlich beobachteten Folge von Wörtern $y_1, y_2, y_3, \ldots$ unterschieden. Ein solches Modell für einen englischen Satz, der mit einem Subjekt beginnt, ist in den folgenden zwei Tabellen (stark vereinfacht) angegeben:

x_i	x_{i+1}	$P(x_{i+1}\|x_i)$
Subjekt	Verb	0.8
Subjekt	Hilfsverb	0.2
Verb	Artikel	0.5
Verb	Adverb	0.2
Verb	Satzende	0.3
Hilfsverb	Adjektiv(1)	1.0
Artikel	Nomen	0.8
Artikel	Adjektiv(2)	0.2
Adjektiv(2)	Nomen	1.0
Adverb	Adverb	0.2
Adverb	Satzende	0.8
Adjektiv(1)	Adverb	0.2
Adjektiv(1)	Satzende	0.8
Nomen	Adverb	0.2
Nomen	Satzende	0.8

x_i	y_i	$P(y_i\|x_i)$
Subjekt	He	0.5
Subjekt	She	0.5
Verb	is	0.5
Verb	shoots	0.5
Hilfsverb	is	0.8
Hilfsverb	seems	0.2
Artikel	the	1.0
Adjektiv(1)	unwell	0.5
Adjektiv(1)	well	0.5
Adverb	well	0.5
Adverb	badly	0.5
Adjektiv(2)	unwell	0.5
Adjektiv(2)	well	0.5
Nomen	duck	0.6
Nomen	well	0.4

Alle nicht angegebenen Wahrscheinlichkeiten $P(x_{i+1}|x_i)$ und $P(y_i|x_i)$ sind Null.

(a) Stellen Sie das Modell für die Satzstruktur in einem Übergangsdiagramm graphisch dar! Berücksichtigen Sie dabei nur die Wortarten x_i, aber nicht die Wörter y_i! Sie brauchen keine Wahrscheinlichkeiten einzutragen. (6 Punkte)



(b) Mit welcher Wahrscheinlichkeit tritt der Satz „He shoots the duck well well well." in diesem Modell auf? (6 Punkte)

$$\begin{aligned}
& P(\text{„He shoots the duck well well well."}) \\
= \; & P(y_1 = \text{„He"}|x_1 = \text{Subjekt})P(x_2 = \text{Verb}|x_1 = \text{Subjekt}) \\
\times \; & P(y_2 = \text{„shoots"}|x_2 = \text{Verb})P(x_3 = \text{Artikel}|x_2 = \text{Verb}) \\
\times \; & P(y_3 = \text{„the"}|x_3 = \text{Artikel})P(x_4 = \text{Nomen}|x_3 = \text{Artikel}) \\
\times \; & P(y_4 = \text{„duck"}|x_4 = \text{Nomen})P(x_5 = \text{Adverb}|x_4 = \text{Nomen}) \\
\times \; & P(y_5 = \text{„well"}|x_5 = \text{Adverb})P(x_6 = \text{Adverb}|x_5 = \text{Adverb}) \\
\times \; & P(y_6 = \text{„well"}|x_6 = \text{Adverb})P(x_7 = \text{Adverb}|x_6 = \text{Adverb}) \\
\times \; & P(y_7 = \text{„well"}|x_7 = \text{Adverb})P(x_8 = \text{Satzende}|x_7 = \text{Adverb}) \\
= \; & 0.5 \cdot 0.8 \cdot 0.5 \cdot 0.5 \cdot 1.0 \cdot 0.8 \cdot 0.6 \cdot 0.2 \cdot 0.5 \cdot 0.2 \cdot 0.5 \cdot 0.2 \cdot 0.5 \cdot 0.8 \\
= \; & 0.2^3 \cdot 0.5^6 \cdot 0.6 \cdot 0.8^3 \cdot 1.0 \\
\approx \; & 3.84 \cdot 10^{-5}
\end{aligned}$$

(c) Das zweite Wort des aus drei Wörtern bestehenden Satzes „She … well." wurde nicht erkannt. Wie lautet die wahrscheinlichste Ergänzung? Geben Sie auch an, wie sicher dieses Ergebnis ist! (14 Punkte)

- $y_2 = \text{„is"}, x_2 = \text{Verb}, x_3 = \text{Adverb}$

$$\begin{aligned}
p_1 \; = \; & P(y_1 = \text{„She"}|x_1 = \text{Subjekt})P(x_2 = \text{Verb}|x_1 = \text{Subjekt}) \\
\times \; & P(y_2 = \text{„is"}|x_2 = \text{Verb})P(x_3 = \text{Adverb}|x_2 = \text{Verb}) \\
\times \; & P(y_3 = \text{„well"}|x_3 = \text{Adverb})P(x_4 = \text{Satzende}|x_3 = \text{Adverb}) \\
= \; & 0.5 \cdot 0.8 \cdot 0.5 \cdot 0.2 \cdot 0.5 \cdot 0.8 \\
= \; & 0.2 \cdot 0.5^3 \cdot 0.8^2 = 0.016
\end{aligned}$$

- $y_2 = \text{„shoots"}, x_2 = \text{Verb}, x_3 = \text{Adverb}$

$$\begin{aligned}
p_2 \; = \; & P(y_1 = \text{„She"}|x_1 = \text{Subjekt})P(x_2 = \text{Verb}|x_1 = \text{Subjekt}) \\
\times \; & P(y_2 = \text{„shoots"}|x_2 = \text{Verb})P(x_3 = \text{Adverb}|x_2 = \text{Verb}) \\
\times \; & P(y_3 = \text{„well"}|x_3 = \text{Adverb})P(x_4 = \text{Satzende}|x_3 = \text{Adverb}) \\
= \; & 0.5 \cdot 0.8 \cdot 0.5 \cdot 0.2 \cdot 0.5 \cdot 0.8 \\
= \; & 0.2 \cdot 0.5^3 \cdot 0.8^2 = 0.016
\end{aligned}$$

- $y_2 =$ „is", $x_2 =$ Hilfsverb, $x_3 =$ Adjektiv(1)

$$\begin{aligned}
p_3 &= P(y_1 = \text{„She"}|x_1 = \text{Subjekt})P(x_2 = \text{Hilfsverb}|x_1 = \text{Subjekt}) \\
&\times P(y_2 = \text{„is"}|x_2 = \text{Hilfsverb})P(x_3 = \text{Adjektiv(1)}|x_2 = \text{Hilfsverb}) \\
&\times P(y_3 = \text{„well"}|x_3 = \text{Adjektiv(1)})P(x_4 = \text{Satzende}|x_3 = \text{Adjektiv(1)}) \\
&= 0.5 \cdot 0.2 \cdot 0.8 \cdot 1.0 \cdot 0.5 \cdot 0.8 \\
&= 0.2 \cdot 0.5^2 \cdot 0.8^2 \cdot 1.0 = 0.032
\end{aligned}$$

- $y_2 =$ „seems", $x_2 =$ Hilfsverb, $x_3 =$ Adjektiv(1)

$$\begin{aligned}
p_4 &= P(y_1 = \text{„She"}|x_1 = \text{Subjekt})P(x_2 = \text{Hilfsverb}|x_1 = \text{Subjekt}) \\
&\times P(y_2 = \text{„seems"}|x_2 = \text{Hilfsverb})P(x_3 = \text{Adjektiv(1)}|x_2 = \text{Hilfsverb}) \\
&\times P(y_3 = \text{„well"}|x_3 = \text{Adjektiv(1)})P(x_4 = \text{Satzende}|x_3 = \text{Adjektiv(1)}) \\
&= 0.5 \cdot 0.2 \cdot 0.2 \cdot 1.0 \cdot 0.5 \cdot 0.8 \\
&= 0.2^2 \cdot 0.5^2 \cdot 0.8 \cdot 1.0 = 0.008
\end{aligned}$$

- Posterior-Wahrscheinlichkeit

$$\begin{aligned}
P(y_2 = \text{„is"}|\text{„She \dots well."}) &= \frac{p_1 + p_3}{p_1 + p_2 + p_3 + p_4} \\
&= \frac{0.016 + 0.032}{0.016 + 0.016 + 0.032 + 0.008} \\
&= \frac{0.048}{0.048 + 0.024} = \frac{2}{3} \approx 0.667
\end{aligned}$$

- Die wahrscheinlichste Ergänzung ist $y_2 =$ „is".

(d) Geben Sie die Wahrscheinlichkeit an, einen Satz aus nur zwei Wörtern zu bilden! Wie viele Möglichkeiten hierfür gibt es? (4 Punkte)

- Die einzige mögliche Folge von Wortarten bei einem Satz aus zwei Wörtern ist $x_1 =$ Subjekt, $x_2 =$ Verb, $x_3 =$ Satzende:

$$\begin{aligned}
P(L = 2) &= P(x_2 = \text{Verb}|x_1 = \text{Subjekt})P(x_3 = \text{Satzende}|x_2 = \text{Verb}) \\
&= 0.8 \cdot 0.3 = 0.24
\end{aligned}$$

- Für einen Satz aus zwei Wörtern gibt es insgesamt $2 \cdot 2 = 4$ Möglichkeiten.

Aufgabe 3 – Generatives Modell (22 Punkte)

Ein Freund behauptet, das Ergebnis eines Münzwurfs (Kopf oder Zahl) mittels seiner magischen Kräfte beeinflussen zu können. Zum Beweis wirft er die Münze mehrmals. Zunächst erzielt er 9-mal hintereinander das Ergebnis Zahl. Als danach einmal das Ergebnis Kopf kommt, hört er auf.

(a) Welche Wahrscheinlichkeitsverteilung hat das Auftreten von k Erfolgen hintereinander und danach eines Fehlschlags in $k+1$ unabhängigen Versuchen, wenn in jedem Versuch die Erfolgswahrscheinlichkeit p beträgt? (4 Punkte)

$$P(k|p) = p^k(1-p)$$

(b) Wie viele Erfolge erwarten Sie in 10 Versuchen mit einer normalen Münze, bei der das Ergebnis Zahl mit der Wahrscheinlichkeit $p = 0.5$ auftritt? (2 Punkte)

$$\langle k \rangle = n \cdot p = 10 \cdot 0.5 = 5$$

(c) Bei einem Versandhändler werden Zaubermünzen mit $p = 0.7$ angeboten. Wie wahrscheinlich ist das Resultat, das Ihr Freund erzielt hat, wenn es sich um eine solchen Münze handelt? (4 Punkte)

$$P(k=9|p=0.7) = 0.7^9(1-0.7) \approx 1.21 \cdot 10^{-2}$$

(d) Zeigen Sie, dass die Wahl $p = k/(k+1)$ die Wahrscheinlichkeit maximiert, genau k Erfolge hintereinander und dann einen Fehlschlag zu beobachten! (6 Punkte)

- Logarithmus der Likelihood

$$\log \mathcal{L} = k \log p + \log(1-p)$$

- Berechnung der Ableitung

$$\frac{\mathrm{d}}{\mathrm{d}p} \log \mathcal{L} = \frac{k}{p} - \frac{1}{1-p}$$

- Bedingung für ein Maximum

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}p} \log \mathcal{L} = 0 \iff \frac{k}{p} = \frac{1}{1-p} &\iff k - k\,p = p \\ \iff p + k\,p = k &\iff p = \frac{k}{k+1} \end{aligned}$$

(e) Vergleichen Sie die beiden Hypothesen $p = 0.5$ und $p = 0.7$ für das Ergebnis Zahl beim Münzwurf. Wie wahrscheinlich sind diese nach Beobachtung der 10 Münzwürfe Ihres Freundes, wenn Sie beide vorher als gleichwahrscheinlich angesehen haben und andere Möglichkeiten ausschließen? (6 Punkte)

- Likelihood für $p = 0.5$

$$P(k = 9|p = 0.5) = 0.5^9(1 - 0.5) \approx 9.77 \cdot 10^{-4}$$

- Likelihood für $p = 0.7$

$$P(k = 9|p = 0.7) = 0.7^9(1 - 0.7) \approx 1.21 \cdot 10^{-2}$$

- Posterior-Wahrscheinlichkeiten

$$\begin{aligned} P(p = 0.7|k = 9) &= \frac{P(k = 9|p = 0.7)}{P(k = 9|p = 0.7) + P(k = 9|p = 0.5)} \\ &\approx \frac{1.21 \cdot 10^{-2}}{1.21 \cdot 10^{-2} + 9.77 \cdot 10^{-4}} \approx 0.925 \\ P(p = 0.5|k = 9) &= 1 - P(p = 0.7|k = 9) \\ &\approx 1 - 0.925 = 0.075 \end{aligned}$$

Aufgabe 4 – Neuronales Netz (18 Punkte)

Betrachten Sie ein Perzeptron mit zwei reellwertigen Eingabeneuronen e_1 und e_2, zwei Gewichten w_1 und w_2 sowie einem Bias w_0. Für das lokale Feld h gilt

$$h = w_1 e_1 + w_2 e_2 - w_0$$

und zur Berechnung der Ausgabe a wird die Aktivierungsfunktion

$$a = sgn(h) = \begin{cases} +1 & \text{für } h > 0 \\ -1 & \text{für } h \leq 0 \end{cases}$$

verwendet. Alle Gewichte und der Bias werden mit 1 initialisiert. Das Perzeptron soll nun die folgenden Beispiele (Soll-Ausgabe y) lernen:

e_1	$+2$	$+1$	$+3$	-1
e_2	$+2$	-2	-1	$+3$
y	$+1$	-1	$+1$	-1

(a) Welche der vier Beispiele werden vom Perzeptron ohne eine Anpassung der Gewichte falsch klassifiziert? (4 Punkte)

Berechnung der Ausgaben ohne Training:

$$\begin{aligned} a_1 &= \operatorname{sgn}(2+2-1) &= \operatorname{sgn}(3) &= +1 = y_1 \\ a_2 &= \operatorname{sgn}(1-2-1) &= \operatorname{sgn}(-2) &= -1 = y_2 \\ a_3 &= \operatorname{sgn}(3-1-1) &= \operatorname{sgn}(1) &= +1 = y_3 \\ a_4 &= \operatorname{sgn}(-1+3-1) &= \operatorname{sgn}(1) &= +1 \neq y_4 \end{aligned}$$

Es wird nur Beispiel 4 falsch klassifiziert.

(b) Wie ändern sich die Gewichte, wenn das neuronale Netz gemäß der Perzeptron-Lernregel mit jedem Beispiel einmal trainiert wird? Verwenden Sie $\lambda = 0.1$ als Lernrate und passen Sie auch den Bias w_0 an. (4 Punkte)

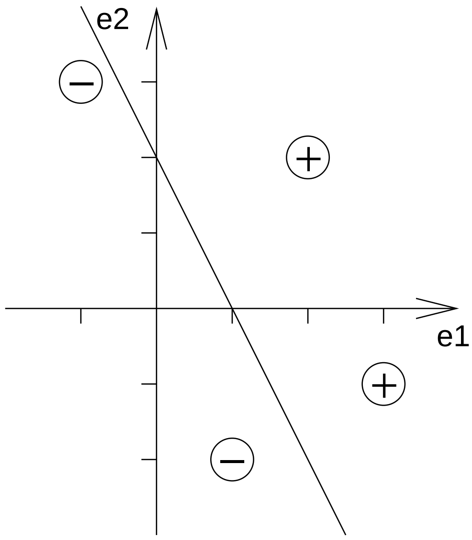
Perzeptron-Lernregel:

$$\left.\begin{array}{ll} a_1 = y_1 & a_2 = y_2 \\ a_3 = y_3 & a_4 \neq y_4 \end{array}\right\} \implies \text{Training nur für Beispiel 4 nötig}$$

Für die neuen Gewichte gilt:

$$\begin{aligned} w_0^+ = 1 + 0.1 \cdot (-1) \cdot (-1) &= 1.1 \\ w_1^+ = 1 + 0.1 \cdot (-1) \cdot (-1) &= 1.1 \\ w_2^+ = 1 + 0.1 \cdot (+3) \cdot (-1) &= 0.7 \end{aligned}$$

(c) Stellen Sie das Klassifikationsproblem in der e_1-e_2-Ebene graphisch dar und lösen Sie es durch möglichst wenige lineare Entscheidungsgrenzen! (6 Punkte)



(d) Kann ein Perzeptron die Beispiele bei genügend langem Training exakt lernen? Begründen Sie Ihre Antwort! (2 Punkte)

Da die Trainingsmenge linear separabel ist, kann ein Perzeptron diese bei genügend langem Training exakt lernen.

(e) Verändern Sie die Lernbarkeit der Trainingsmenge (für ein Perzeptron) durch Hinzufügen oder Weglassen eines Beispiels! (2 Punkte)

Wird zur Trainingsmenge das Beispiel $e_1 = -2, e_2 = -2, y = +1$ hinzugefügt, so ist diese nicht mehr linear separabel und kann nicht mehr von einem Perzeptron gelernt werden.