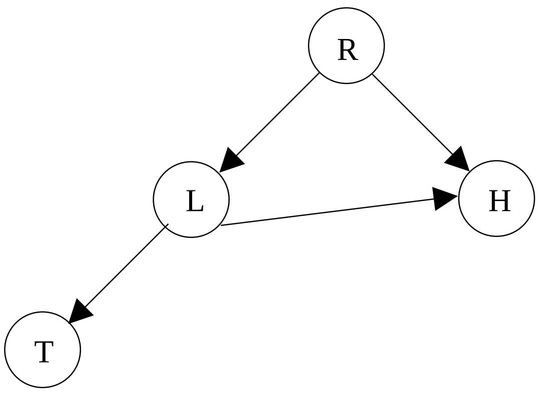
Aufgabe 1 – Probabilistische Inferenz (28 Punkte)

Mit einem neuen Test soll die Genauigkeit von Lungenkrebs-Diagnosen verbessert werden. In einer Studie wird folgendes festgestellt:

- Einer von 1000 Rauchern (R=w) unter den Studienteilnehmern ist an Lungenkrebs erkrankt (L=w). Bei Nichtrauchern (R=f) ist die Wahrscheinlichkeit 100 mal kleiner.
- Bei 70% der an Lungenkrebs (L=w) Erkrankten tritt chronischer Husten (H=w) auf. Bei nicht an Lungenkrebs (L=f) Erkrankten, die rauchen, kann aber auch mit 5% Wahrscheinlichkeit chronischer Husten auftreten. Bei Nichtrauchern nur zu 0.1%.
- Bei Lungenkrebspatienten (L=w) fällt der Test zu 1% trotzdem negativ aus (T=f). Sind die Patienten hingen gesund (L=f) wird der Test zu 99.9% negativ ausfallen.
- Ein Fünftel der Probanden gab an, Raucher zu sein.

(a) Zeichnen Sie ein Bayes-Netz, das zu diesem Modell passt! Die Wahrscheinlichkeitstabellen brauchen Sie hierfür nicht anzugeben. (6 Punkte)



(b) Wie groß ist die Wahrscheinlichkeit, dass der Test bei einem Nichtraucher positiv ausfällt? (6 Punkte)

$$
\begin{aligned}
P(T=w|R=f) &= \sum_x P(T=w, L=x|R=f) \\
&= \sum_x P(T=w|L=x)P(L=x|R=f) \\
&= P(T=w|L=f)P(L=f|R=f) \\
&+ P(T=w|L=w)P(L=w|R=f) \\
&= 0.001 \cdot (1-10^{-5}) + 0.99 \cdot 10^{-5} \\
&\approx 1.0099 \cdot 10^{-3}
\end{aligned}
$$

(c) Wie wahrscheinlich ist es, dass ein Nichtraucher gesund ist, obwohl der Test positiv ausfällt? (6 Punkte)

$$\begin{aligned}
P(L=f|R=f,T=w) &= \frac{P(T=w|R=f,L=f)P(L=f|R=f)}{P(T=w|R=f)} \\
&= \frac{P(T=w|L=f)P(L=f|R=f)}{P(T=w|R=f)} \\
&\overset{b)}{\approx} \frac{0.001\cdot(1-10^{-5})}{1.0099\cdot 10^{-3}} \\
&\approx 0.9902
\end{aligned}$$

(d) Wie wahrscheinlich ist es, dass ein zufällig ausgewählter Proband an chronischem Husten leidet? (10 Punkte)

$$\begin{aligned}
P(H=w) &= \sum_{x,y} P(H=w|L=x,R=y)P(L=x,R=y) \\
&= \sum_{x,y} P(H=w|L=x,R=y)P(L=x|R=y)P(R=y) \\
&= P(H=w|L=f,R=f)P(L=f|R=f)P(R=f) \\
&+ P(H=w|L=f,R=w)P(L=f|R=w)P(R=w) \\
&+ P(H=w|L=w,R=f)P(L=w|R=f)P(R=f) \\
&+ P(H=w|L=w,R=w)P(L=w|R=w)P(R=w) \\
&= 0.001\cdot(1-10^{-5})\cdot 0.8 + 0.05\cdot(1-10^{-3})\cdot 0.2 \\
&+ 0.7\cdot 10^{-5}\cdot 0.8 + 0.7\cdot 10^{-3}\cdot 0.2 \\
&\approx 0.011
\end{aligned}$$

Aufgabe 2 – Hidden-Markov-Modell (28 Punkte)

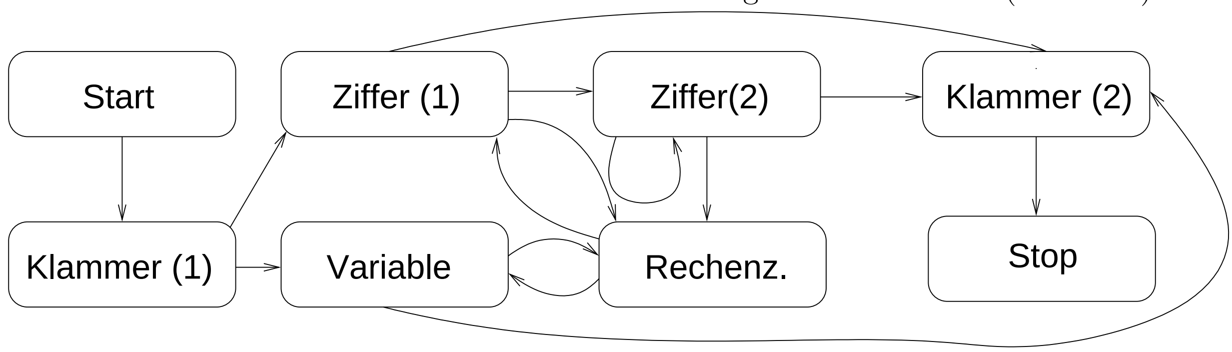
Wir benutzen ein Modell zur Erkennung von mathematischen Ausdrücken. Wir unterscheiden zwischen einer Folge von Zeichentypen $x_1, x_2, x_3, \ldots$ und der tatsächlich beobachteten Folge von Zeichen $y_1, y_2, y_3, \ldots$.

x_i	x_{i+1}	$P(x_{i+1}\|x_i)$
Start	Klammer(1)	1.0
Klammer(1)	Variable	0.6
Klammer(1)	Ziffer(1)	0.4
Variable	Rechensymbol	0.8
Variable	Klammer(2)	0.2
Ziffer(1)	Ziffer(2)	0.5
Ziffer(1)	Rechensymbol	0.3
Ziffer(1)	Klammer(2)	0.2
Ziffer(2)	Ziffer(2)	0.3
Ziffer(2)	Rechensymbol	0.4
Ziffer(2)	Klammer(2)	0.3
Rechensymbol	Ziffer(1)	0.7
Rechensymbol	Variable	0.3
Klammer(2)	Stop	1.0

x_i	y_i	$P(y_i\|x_i)$
Klammer(1)	(	1.0
Variable	x	0.6
Variable	y	0.3
Variable	z	0.1
Ziffer(1)	1	0.5
Ziffer(1)	2	0.3
Ziffer(1)	3	0.2
Ziffer(2)	0	0.4
Ziffer(2)	1	0.3
Ziffer(2)	2	0.2
Ziffer(2)	3	0.1
Rechensymbol	+	0.7
Rechensymbol	-	0.3
Klammer(2)	)	1.0

Die Markovkette beginnt immer mit $x_0 = \textit{Start}$ und endet mit $x_{k+1} = \textit{Stop}$. Diesen Zuständen ist keine Ausgabe zugeordnet, da sie Verbindungen zu anderen Teilen des Modells darstellen. Alle nicht angegebenen Wahrscheinlichkeiten $P(x_{i+1}|x_i)$ und $P(y_i|x_i)$ sind Null. Sie können die Zeichentypen abkürzen, um Platz zu sparen.

(a) Stellen Sie das Modell für die Ausdrücke in einem Übergangsdiagramm graphisch dar! Sie brauchen keine Wahrscheinlichkeiten einzutragen. (8 Punkte)



(b) Mit welcher Wahrscheinlichkeit tritt der Ausdruck „(x+30)“ in diesem Modell auf? (6 Punkte)

$$\begin{aligned}
P((x+30)) &= P(V|K1) \cdot P(RZ|V) \cdot P(Z1|RZ) \cdot P(Z2|Z1) \cdot P(K2|Z2) \\
&\cdot\ P('('|K1) \cdot P('x'|V) \cdot P('+'|RZ) \cdot P('3'|Z1) \cdot P('0'|Z2) \cdot P(')'|K2) \\
&= 0.6 \cdot 0.8 \cdot 0.7 \cdot 0.5 \cdot 0.3 \\
&\cdot\ 1.0 \cdot 0.6 \cdot 0.7 \cdot 0.2 \cdot 0.4 \cdot 1.0 \\
&\approx 0.00169
\end{aligned}$$

(c) Zwischen zwei Rechensymbolen wurde ein einzelnes Zeichen nicht erkannt. Welches Zeichen ist die wahrscheinlichste Ergänzung und wie sicher ist diese Vorhersage? (14 Punkte)

Zwischen zwei Rechensymbolen kann entweder eine Variable oder eine Ziffer(1) stehen.

$$\begin{aligned}
P(„RSVRZ“) &\propto P(V|RS) \cdot P(RS|V) = 0.3 \cdot 0.8 = 0.24 \\
P(„RSZ1RZ“) &\propto P(Z1|RS) \cdot P(RS|Z1) = 0.7 \cdot 0.3 = 0.21
\end{aligned}$$

Normierung:

$$\begin{aligned}
P(„RSVRZ“) &= \frac{0.24}{0.24+0.21} \approx 0.533 \\
P(„RSZ1RZ“) &= \frac{0.21}{0.24+0.21} \approx 0.467
\end{aligned}$$

Wahrscheinlichkeiten für die einzelnen Zeichentypen

$$\begin{aligned}
'x' &: 0.533 \cdot 0.6 = 0.3198 \\
'y' &: 0.533 \cdot 0.3 = 0.1599 \\
'z' &: 0.533 \cdot 0.1 = 0.0533 \\
'1' &: 0.467 \cdot 0.5 = 0.2335 \\
'2' &: 0.467 \cdot 0.3 = 0.1401 \\
'3' &: 0.467 \cdot 0.2 = 0.0934
\end{aligned}$$

Damit ist „x“ mit einer Wahrscheinlichkeit von ungefähr 32% die wahrscheinlichste Ergänzung.

Aufgabe 3 – Generatives Modell (24 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. Der erste Versuch hat das Ergebnis Kopf, aber dann erzielt er 7-mal in Folge das Ergebnis Zahl. Als danach zweimal das Ergebnis Kopf kommt, hört er auf. Sie glauben jedoch nicht an übernatürliche Fähigkeiten, sondern daran, dass die Wahrscheinlichkeit für das Ereignis Zahl bei dieser Münze auf $p \neq 0.5$ geändert wurde.

(a) Welche Wahrscheinlichkeitsverteilung hat das Auftreten einer Folge von einem Fehlschlag, k Erfolgen und zwei weiteren Fehlschlägen in $k+3$ unabhängigen Versuchen, wenn in jedem Versuch die Erfolgswahrscheinlichkeit p beträgt? (4 Punkte)

$$P(k|p) = p^k(1-p)^3$$

(b) Ihr Freund nimmt mit seiner Münze an einem Glücksspiel teil. Sein Einsatz pro Münzwurf beträgt 1 Euro; beim Ereignis Zahl werden 2 Euro ausgezahlt, sonst nichts. Wie hoch ist sein erwarteter Gewinn pro Runde, wenn für die Erfolgswahrscheinlichkeit $p = 0.7$ gilt? (4 Punkte)

$$\langle G \rangle = (2-1) \cdot 0.7 + (0-1) \cdot (1-0.7) = 0.4$$

(c) Ihr Vorwissen über mögliche Manipulationen an der Münze beschreiben Sie durch die Beta-Verteilung $\text{Beta}(p; \alpha, \beta) = B(\alpha, \beta)\, p^{\alpha-1}(1-p)^{\beta-1}$. $B(\alpha, \beta)$ hängt nicht von p ab. Wie müssen Sie die Hyperparameter α und β wählen, wenn Sie gar kein Vorwissen über p haben, so dass Maximum-a-posteriori- und Maximum-Likelihood-Hypothese übereinstimmen? Begründen Sie! (4 Punkte)

Ohne Vorwissen sollte $\alpha = 1$ und $\beta = 1$ gewählt werden. In diesem Fall gilt

$$\text{Beta}(p; 1, 1) = 1$$

unabhängig von p, so dass die Maxima von Likelihood $P(k|p)$ und Posterior

$$P(p|k) = C P(k|p) P(p)$$

an der gleichen Stelle liegen. Die Normierungskonstante

$$C = \frac{1}{P(k)}$$

hängt nicht von p ab und hat somit auch keinen Einfluss auf die Lage des Maximums.

(d) Zeigen Sie, dass die Maximum-a-posteriori Hypothese für das Würfelkunststück Ihres Freundes (eine Folge bestehend aus einem Fehlschlag, k Erfolgen und zwei weiteren Fehlschlägen) durch $p = (k + \alpha - 1)/(k + \alpha + \beta + 1)$ gegeben ist! (8 Punkte)

- Der Logarithmus des Posteriors ist durch

$$\begin{aligned}
\log P(p|k) &= \frac{P(k|p)\mathrm{Beta}(p;\alpha,\beta)}{P(k)} \\
&= \log C + k \log p + 3 \log(1-p) \\
&+ (\alpha - 1)\log p + (\beta - 1)\log(1-p) \\
&= \log C + (k + \alpha - 1)\log p + (3 + \beta - 1)\log(1-p)
\end{aligned}$$

  mit der Normierungskonstanten $C = \mathrm{Beta}(p;\alpha,\beta)/P(k)$ gegeben.

- Berechnung der Ableitung:

$$\frac{\mathrm{d}}{\mathrm{d}p}\log P(p|k) = \frac{k+\alpha-1}{p} - \frac{\beta+2}{1-p}$$

- Bedingung für ein Maximum

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}p}\log P(p|k) = 0 \quad &\Longleftrightarrow \quad \frac{k+\alpha-1}{p} = \frac{\beta+2}{1-p} \\
&\Longleftrightarrow \quad (k+\alpha-1)(1-p) = (\beta+2)p \\
&\Longleftrightarrow \quad p = \frac{k+\alpha-1}{k+\alpha+\beta+1}
\end{aligned}$$

(e) Welchen Wert p hat die Maximum-a-posteriori-Hypothese für die Beobachtung $k = 7$, wenn Sie die Hyperparameter auf $\alpha = 3$ und $\beta = 4$ setzen? (4 Punkte)

$$p = \frac{k+\alpha-1}{k+\alpha+\beta+1} = \frac{7+3-1}{7+3+4+1} = \frac{9}{15} \approx 0.60$$

Aufgabe 4 – Neuronales Netz (20 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	-1.0	$+2.0$	-1.5	$+1.0$
e_2	-2.0	-0.5	-0.5	-0.5
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:

$$\begin{aligned}
a_1 = &\ \mathrm{sgn}((-1.0) + (-2.0) - 1) &= &\ \mathrm{sgn}(-4.0) &= -1 = y_1 \\
a_2 = &\ \mathrm{sgn}((+2.0) + (-0.5) - 1) &= &\ \mathrm{sgn}(+0.5) &= +1 = y_2 \\
a_3 = &\ \mathrm{sgn}((-1.5) + (-0.5) - 1) &= &\ \mathrm{sgn}(-3.0) &= -1 = y_3 \\
a_4 = &\ \mathrm{sgn}((+1.0) + (-0.5) - 1) &= &\ \mathrm{sgn}(-0.5) &= -1 \neq y_4
\end{aligned}$$

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 $\lambda = 0.2$ als Lernrate und passen Sie auch den Bias w_0 an. (4 Punkte)

Perzeptron-Lernregel:

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

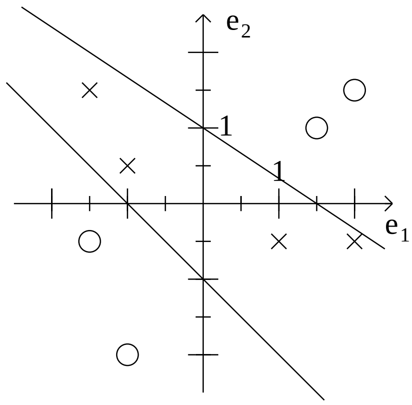
Für die neuen Gewichte gilt:

$$\begin{aligned} w_0^+ = 1 + 0.2 \cdot (-1.0) \cdot (+1) &= 0.8 \\ w_1^+ = 1 + 0.2 \cdot (+1.0) \cdot (+1) &= 1.2 \\ w_2^+ = 1 + 0.2 \cdot (-0.5) \cdot (+1) &= 0.9 \end{aligned}$$

(c) Warum kann ein einzelnes Perzeptron die Beispiele in der unten abgebildeten Grafik nicht erlernen? (Kreuze entsprechen $y = +1$, Kreise $y = -1$.) Begründen Sie Ihre Antwort! (2 Punkte)

Da die Beispiele nicht linear separabel sind, kann ein Perzeptron die Trainingsmenge nicht korrekt lernen. Es wird immer mindestens ein Beispiel geben, das nicht richtig klassifiziert wird.

(d) Zeichnen Sie die Klassifikationsgrenzen zweier Perzeptrone so ein, dass ein neuronale Netz aus drei Neuronen alle Beispiel korrekt klassifiziert, wenn das dritte Neuron eine UND-Verknüpfung der Einzelausgaben a_1 und a_2 realisiert! (4 Punkte)



(e) Geben Sie für eine der beiden Klassifikationsgrenzen die zugehörigen Gewichte des Perzeptrons an! (6 Punkte)

- $w_1 = +1, w_2 = +1, w_0 = -1$
- $w_1 = -2, w_2 = -3, w_0 = -3$