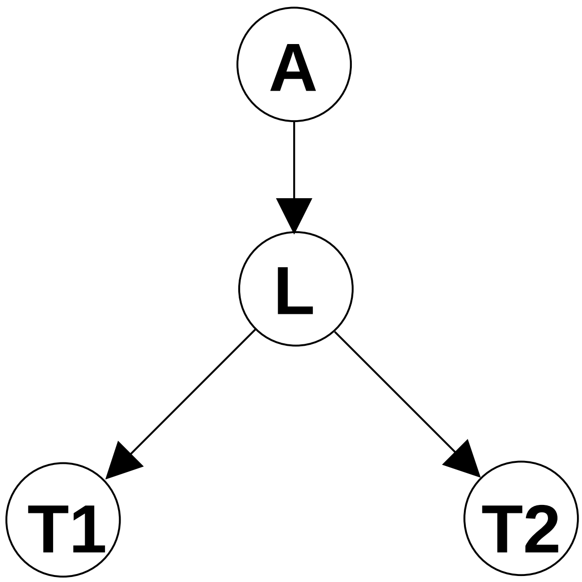
Aufgabe 1 – Probabilistische Inferenz (32 Punkte)

In einer medizinischen Studie werden zwei Tests zur Diagnose von Leberschäden verglichen. Dabei wurde folgendes festgestellt:

- Test 1 erkennt ($T_1 = w$) einen Leberschaden ($L = w$) nur mit 90% Wahrscheinlichkeit. Eine gesunde Leber ($L = f$) führt aber in 99% aller Untersuchungen zu einem negativen Testergebnis ($T_1 = f$).
- Test 2 ist mit 95% Wahrscheinlichkeit positiv ($T_2 = w$), wenn der jeweilige Proband tatsächlich einen Leberschaden hat ($L = w$). Mit derselben Wahrscheinlichkeit fällt der Test negativ ($T_2 = f$) aus, wenn die Leber in Ordnung ist ($L = f$).
- Die Genauigkeit beider Tests wird nicht dadurch beeinflusst, ob die Versuchsperson Alkohol konsumiert ($A = w$) oder nicht ($A = f$).
- In der untersuchten Gruppe lag das Risiko (Wahrscheinlichkeit) für einen Leberschaden bei 10^{-2} für Probanden mit Alkoholkonsum, sonst nur bei 10^{-4}.

Ein Anteil von 25% der Studienteilnehmer trinkt regelmäßig Alkohol ($A = w$).

(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 Leberschaden hat? (8 Punkte)

$$\begin{aligned} P(L=w) &= P(L=w|A=w)P(A=w)+P(L=w|A=f)P(A=f) \\ &= 10^{-2}\cdot 0.25 + 10^{-4}\cdot 0.75 \\ &= 0.002575 \end{aligned}$$

(c) Bei einem Alkoholkonsumenten fällt Test 1 positiv aus. Wie wahrscheinlich ist es, dass er tatsächlich einen Leberschaden hat? (9 Punkte)

$$\begin{aligned} P(L=w|A=w,T_1=w) &= \frac{P(T_1=w|A=w,L=w)P(L=w|A=w)}{\sum_L P(T_1=w|A=w,L)P(L|A=w)} \\ &= \frac{P(T_1=w|L=w)P(L=w|A=w)}{\sum_L P(T_1=w|L)P(L|A=w)} \\ &= \frac{0.90\cdot 0.01}{0.90\cdot 0.01+0.01\cdot 0.99} \\ &\approx 0.4762 \end{aligned}$$

(d) Jetzt soll die Versuchsperson aus Teilaufgabe (c), bei der Test 1 bereits positiv ausgefallen ist, auch mit Test 2 untersucht werden. Welches Ergebnis ist am wahrscheinlichsten und wie sicher ist diese Vorhersage? (9 Punkte)

$$\begin{aligned} P(T_2=w|A=w,T_1=w) &= P(T_2=w|L=w)P(L=w|A=w,T_1=w) \\ &+ P(T_2=w|L=f)P(L=f|A=w,T_1=w) \\ &\approx 0.95\cdot 0.4762 + 0.05\cdot 0.5238 \\ &\approx 0.4786 \\ P(T_2=f|A=w,T_1=w) &= 1-P(T_2=w|A=w,T_1=w) \\ &\approx 0.5214 \end{aligned}$$

Es ist am wahrscheinlichsten, dass der Test 2 negativ ausfällt. Allerdings beträgt die Wahrscheinlichkeit hierfür nur 52%.

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

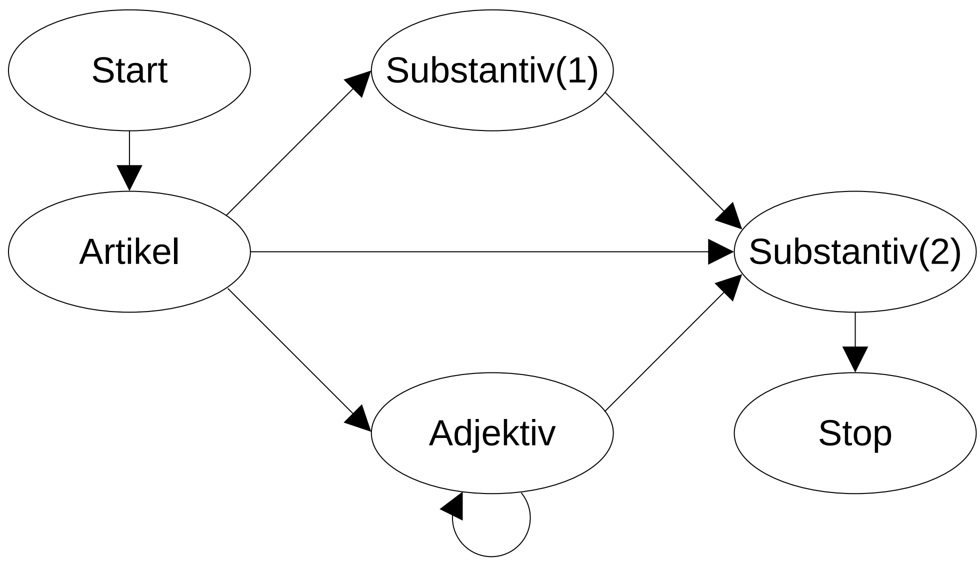
Bei der Spracherkennung werden Wortmodelle eingesetzt, um die Wahrscheinlichkeit von Wortfolgen zu berechnen. Um die Grammatik besser berücksichtigen zu können, wird zwischen einer Folge von Wortarten $x_1, x_2, x_3, \dots$ und der tatsächlich beobachteten Folge von Wörtern $y_1, y_2, y_3, \dots$ unterschieden.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

x_i	x_{i+1}	$P(x_{i+1}\|x_i)$
Start	Artikel	1.0
Artikel	Adjektiv	0.3
Artikel	Substantiv(1)	0.4
Artikel	Substantiv(2)	0.3
Adjektiv	Adjektiv	0.5
Adjektiv	Substantiv(2)	0.5
Substantiv(1)	Substantiv(2)	1.0
Substantiv(2)	Stop	1.0

x_i	y_i	$P(y_i\|x_i)$
Artikel	the	0.8
Artikel	a	0.2
Adjektiv	small	0.5
Adjektiv	green	0.5
Substantiv(1)	village	0.6
Substantiv(1)	green	0.4
Substantiv(2)	village	0.6
Substantiv(2)	green	0.4

Die Markovkette für die Nominalgruppe beginnt immer mit $x_0 = \textit{Start}$ und endet mit $x_{k+1} = \textit{Stop}$. Diesen beiden Zuständen ist keine Ausgabe zugeordnet, da sie Verbindungen zu anderen Teilen des Sprachmodells darstellen. 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! Sie brauchen keine Wahrscheinlichkeiten einzutragen. (6 Punkte)



(b) Mit welcher Wahrscheinlichkeit tritt die Nominalgruppe „the village green“ in diesem Modell auf? (6 Punkte)

$$\begin{aligned} P(\text{„the village green“}) &= P(x_1 = \text{Artikel} | x_0 = \text{Start}) \\ &\cdot\; P(y_1 = \text{„the“} | x_1 = \text{Artikel}) \\ &\cdot\; P(x_2 = \text{Substantiv(1)} | x_1 = \text{Artikel}) \\ &\cdot\; P(y_2 = \text{„village“} | x_2 = \text{Substantiv(1)}) \\ &\cdot\; P(x_3 = \text{Substantiv(2)} | x_2 = \text{Substantiv(1)}) \\ &\cdot\; P(y_3 = \text{„green“} | x_3 = \text{Substantiv(2)}) \\ &\cdot\; P(x_4 = \text{Stop} | x_3 = \text{Substantiv(2)}) \\ &= 1.0 \cdot 0.8 \cdot 0.4 \cdot 0.6 \cdot 1.0 \cdot 0.4 \cdot 1.0 \\ &= 0.0768 \end{aligned}$$

(c) Das zweite Wort in der aus drei Wörtern bestehenden Nominalgruppe „a … village“ wurde nicht erkannt. Wie lautet die wahrscheinlichste Ergänzung? (12 Punkte)

$$\begin{aligned} P(\text{„a green village“}) &= P(\text{„a green village“}, x_2 = \text{Adjektiv}) \\ &+\; P(\text{„a green village“}, x_2 = \text{Substantiv(1)}) \\ &= 1.0 \cdot 0.2 \cdot 0.3 \cdot 0.5 \cdot 0.5 \cdot 0.6 \cdot 1.0 \\ &+\; 1.0 \cdot 0.2 \cdot 0.4 \cdot 0.4 \cdot 1.0 \cdot 0.6 \cdot 1.0 \\ &= 0.0282 \\ P(\text{„a small village“}) &= 1.0 \cdot 0.2 \cdot 0.3 \cdot 0.5 \cdot 0.5 \cdot 0.6 \cdot 1.0 \\ &= 0.0009 \\ P(\text{„a village village“}) &= 1.0 \cdot 0.2 \cdot 0.4 \cdot 0.6 \cdot 1.0 \cdot 0.6 \cdot 1.0 \\ &= 0.0288 \end{aligned}$$

Die Ergänzung der Nominalgruppe zu „a village village“ ist in diesem Sprachmodell am wahrscheinlichsten.

(d) Geben Sie die Wahrscheinlichkeit an, eine Nominalgruppe zu beobachten, die aus genau k Wörtern ($k = 0, 1, 2, \ldots$) besteht! (4 Punkte)

$$P(k) = \begin{cases} 0.0 & \text{für } k \leq 1 \\ 0.3 & \text{für } k = 2 \\ 0.4 + 0.3 \cdot 0.5 & \text{für } k = 3 \\ 0.3 \cdot 0.5^{k-2} & \text{für } k \geq 4 \end{cases}$$

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 10-mal und erzielt 8-mal das Ergebnis Zahl. Sie vermuten allerdings, dass dies ein Trick mit einer präparierten Münze ist.

(a) Welche Wahrscheinlichkeitsverteilung hat die Zahl der Erfolge k (hier das Ergebnis Zahl) in einer Reihe von n unabhängigen Versuchen, wenn in jedem Versuch die Erfolgswahrscheinlichkeit p beträgt? (4 Punkte)

$$P(k|p) = \binom{n}{k} p^k \, (1-p)^{n-k}$$

Die gesuchte Verteilung ist die Binomialverteilung.

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

$$\begin{aligned} \langle k \rangle &= n\,p \\ &= 10 \cdot 0.5 \\ &= 5 \end{aligned}$$

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

$$\begin{aligned} P(k=8|p=0.7) &= \binom{10}{8} 0.7^8 \, (1-0.7)^{10-8} \\ &= 45 \cdot 0.7^8 \cdot 0.3^2 \\ &\approx 0.2335 \end{aligned}$$

(d) Zeigen Sie, dass die Wahl $p = k/n$ die Wahrscheinlichkeit maximiert, k Erfolge in n Versuchen zu beobachten! (6 Punkte)

$$\begin{aligned}
\frac{\partial}{\partial p} \log P(k|p) = 0 \quad &\Longleftrightarrow \quad \frac{\partial}{\partial p} \log \left[\binom{n}{k} p^k \, (1-p)^{n-k} \right] = 0 \\
&\Longleftrightarrow \quad \frac{k}{p} - \frac{n-k}{1-p} = 0 \\
&\Longleftrightarrow \quad p = \frac{k}{n}
\end{aligned}$$

(e) Vergleichen Sie die beiden Hypothesen $p = 0.5$ und $p = 0.7$ für die Erfolgswahrscheinlichkeit. Wie wahrscheinlich sind diese nach der Beobachtung der 8 Erfolge in 10 Versuchen, wenn Sie beide vorher als gleichwahrscheinlich angesehen haben und andere Möglichkeiten ausschließen? (6 Punkte)

$$\begin{aligned}
P(k=8|p=0.5) &= \binom{10}{8} 0.5^8 \, (1-0.5)^{10-8} \\
&= 45 \cdot 0.5^8 \cdot 0.5^2 \\
&\approx 0.0439 \\
P(k=8|p=0.7) &\approx 0.2335 \\
P(p=0.7|k=8) &= \frac{P(k=8|p=0.7)P(p=0.7)}{P(k=8|p=0.7)P(p=0.7) + P(k=8|p=0.5)P(p=0.5)} \\
&\approx \frac{0.2335 \cdot 0.5}{0.2335 \cdot 0.5 + 0.0439 \cdot 0.5} \\
&\approx 0.8416 \\
P(p=0.5|k=8) &= 1 - P(p=0.7|k=8) \\
&\approx 1 - 0.8416 \\
&\approx 0.1584
\end{aligned}$$

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

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

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

Nur Beispiel 3 wird ohne Training falsch klassifiziert.

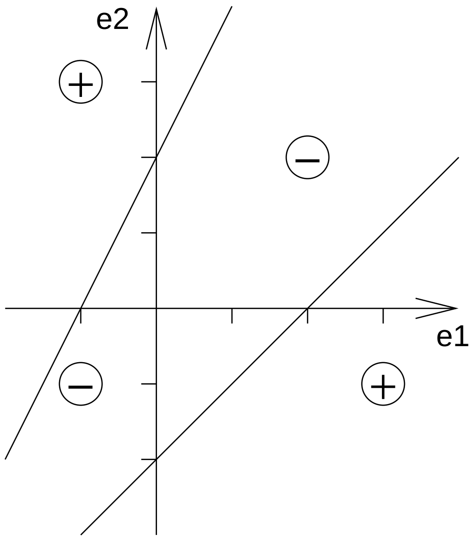
(b) Wie ändern sich die Gewichte, wenn das neuronale Netz gemäß der Perzeptron-Lernregel mit jedem Beispiel einmal trainiert wird? Verwenden Sie $\lambda = 0.1$ als Lernrate und passen Sie auch den Bias w_0 an. (4 Punkte)

Bei der Perzeptron-Lernregel müssen nur für das fehlklassifizierte Beispiel 3 die Gewichte angepasst werden:

$$\begin{aligned}
w_0^+ &= w_0 + \lambda\, y_3\, e_0 = 1 + 0.1 \cdot (-1) \cdot (-1) = 1.1 \\
w_1^+ &= w_1 + \lambda\, y_3\, e_1 = 1 + 0.1 \cdot (-1) \cdot (+2) = 0.8 \\
w_2^+ &= w_2 + \lambda\, y_3\, e_2 = 1 + 0.1 \cdot (-1) \cdot (+2) = 0.8
\end{aligned}$$

Die Klassifikation der anderen Beispiele ändert sich dadurch nicht.

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



(d) Kann ein Perzeptron die Beispiele bei genügend langem Training exakt lernen? 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.

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

Wenn eines der Beispiele weggelassen wird, kann das Perzeptron die restliche Trainingsmenge lernen. Bereits der Startwert der Gewichte führt zu einer korrekten Klassifikation, wenn speziell Beispiel 3 entfernt wird.