# Technische Universität Berlin
# Fakultät IV – Elektrotechnik und Informatik

## Künstliche Intelligenz: Grundlagen und Anwendungen
## Wintersemester 2011 / 2012

Albayrak, Fricke (AOT) – Opper, Ruttor (KI)

## Schriftlicher Test – Teilklausur 2

**21.02.2012**

**Name, Vorname:** ______________________________

**Matrikelnummer:** ______________________________

**Studiengang:** ______________________________

**Hinweise:**

- Überprüfen Sie bitte, ob Sie alle **10** Seiten der Klausur erhalten haben.
- Bitte versehen Sie vor Bearbeitung der Klausur alle **10** Seiten mit Ihrer Matrikelnummer.
- Bitte nicht mit einem roten oder grünen Stift schreiben.
- Bitte keinen Bleistift, keinen Tintenkiller und kein Tipp-Ex benutzen.
- Die Vorder- und Rückseiten der Klausur dürfen verwendet werden. Den Anhang (Seite 10) dürfen Sie abtrennen. Sie müssen ihn nicht abgeben.

Dieser Teil ist zur Auswertung bestimmt und soll von den Teilnehmerinnen und Teilnehmern der Klausur nicht ausgefüllt werden.

| **Aufgabe 1** 28 Punkte | **Aufgabe 2** 28 Punkte | **Aufgabe 3** 24 Punkte | **Aufgabe 4** 20 Punkte | **Summe** 100 Punkte |
|---|---|---|---|---|
| | | | | |

## Aufgabe 1 – Probabilistische Inferenz (28 Punkte)

Die BVG will besser auf Ausfälle im S-Bahn-Verkehr eingestellt sein. Sie geht dabei von folgenden Annahmen aus:

- An 20% der Tage ist es besonders kalt ($K = w$), an den restlichen Tagen ist die Temperatur gemäßigt oder warm ($K = f$).
- An kalten Tagen ($K = w$) ist zu 70% mit einem erhöhtem Krankenstand bei S-Bahn-Fahrern zu rechnen ($F = w$). Bei normalen Temperaturen (K=f) dagegen tritt diese Situation nur zu 10% ein.
- An 70% der Tage wurde die Wartung der Züge ($W = w$) nicht ausreichend durchgeführt.
- Ist es kalt und die Wartung ist nicht ausreichend ($W = w \wedge K = w$), sind zu 90% mehr als 1/4 der Züge nicht funktionstüchtig ($Z = w$). Trifft nur genau eine dieser Voraussetzungen zu ($W = w \oplus K = w$), ist die Wahrscheinlichkeit 0.6 und trifft keine zu ($W = f \wedge K = f$) ist die Wahrscheinlichkeit 0.1.
- Tritt ein erhöhter Krankenstand oder ein Ausfall von mehr als 1/4 der Züge ein ($F = w \vee Z = w$), ist zu 90% mit einem S-Bahn-Chaos zu rechnen (C=w). Andernfalls ist die Wahrscheinlichkeit 25%.

(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 es an einem kalten Tag, an dem alle Züge funktionstüchtig sind, zu einem S-Bahn-Chaos kommt? (6 Punkte)

(c) Wie hoch ist die Wahrscheinlichkeit, dass es ein kalter Tag ist, wenn es einen erhöhten Krankenstand gibt? (6 Punkte)

(d) Mehr als 1/4 der Züge sind ausgefallen. Wie sicher kann man sich sein, dass nicht ausreichend gewartet wurde? (10 Punkte)

## 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 $x_1, x_2, x_3, \ldots$ und der tatsächlich beobachteten Folge von Wörtern $y_1, y_2, y_3, \ldots$ 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 $x_0 = \mathit{Start}$ und endet mit $x_{k+1} = \mathit{Stop}$. 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 „a small village“ in diesem Modell auf? (6 Punkte)

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

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

## 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 8-mal in Folge das Ergebnis Zahl. Als danach wieder 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 einen Fehlschlag, $k$ Erfolgen und eines weiteren Fehlschlags in $k+2$ unabhängigen Versuchen, wenn in jedem Versuch die Erfolgswahrscheinlichkeit $p$ beträgt? (4 Punkte)

(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)

(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 $\alpha$ und $\beta$ wählen, wenn Sie gar kein Vorwissen über $p$ haben, so dass Maximum-a-posteriori- und Maximum-Likelihood-Hypothese übereinstimmen? Begründen Sie! (4 Punkte)

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

(e) Welchen Wert $p$ hat die Maximum-a-posteriori-Hypothese für die Beobachtung $k = 8$, wenn Sie die Hyperparameter auf $\alpha = 3$ und $\beta = 3$ setzen? (4 Punkte)

## 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:

| | | | | |
|---|---|---|---|---|
| $e_1$ | $-2$ | $+1$ | $+3$ | $+4$ |
| $e_2$ | $-1$ | $+3$ | $+4$ | $-1$ |
| $y$ | $-1$ | $+1$ | $-1$ | $+1$ |

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

(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. (6 Punkte)

(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)

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

# Anhang – darf abgetrennt werden

## Aufgabe 1: Annahmen des Modells

- An 20% der Tage ist es besonders kalt ($K = w$), an den restlichen Tagen ist die Temperatur gemäßigt oder warm ($K = f$).
- An kalten Tagen ($K = w$) ist zu 70% mit einem erhöhtem Krankenstand bei S-Bahn-Fahrern zu rechnen ($F = w$). Bei normalen Temperaturen (K=f) dagegen tritt diese Situation nur zu 10% ein.
- An 70% der Tage wurde die Wartung der Züge ($W = w$) nicht ausreichend durchgeführt.
- Ist es kalt und die Wartung ist nicht ausreichend ($W = w \wedge K = w$), sind zu 90% mehr als 1/4 der Züge nicht funktionstüchtig ($Z = w$). Trifft nur genau eine dieser Voraussetzungen zu ($W = w \oplus K = w$), ist die Wahrscheinlichkeit 0.6 und trifft keine zu ($W = f \wedge K = f$) ist die Wahrscheinlichkeit 0.1.
- Tritt ein erhöhter Krankenstand oder ein Ausfall von mehr als 1/4 der Züge ein ($F = w \vee Z = w$), ist zu 90% mit einem S-Bahn-Chaos zu rechnen (C=w). Andernfalls ist die Wahrscheinlichkeit 25%.

## Aufgabe 2: Sprachmodell

| $x_i$ | $x_{i+1}$ | $P(x_{i+1}\|x_i)$ |
|---|---|---|
| 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 |

## Aufgabe 3: Priorverteilung

$$\mathrm{Beta}(p;\alpha,\beta) = B(\alpha,\beta)\, p^{\alpha-1}(1-p)^{\beta-1}$$

## Aufgabe 4: Beispiele

$$\begin{array}{c|cccc} e_1 & -2 & +1 & +3 & +4 \\ e_2 & -1 & +3 & +4 & -1 \\ \hline y & -1 & +1 & -1 & +1 \end{array}$$