Klausur zur Vorlesung

Maschinelles Lernen: Symbolische Ansätze

Prof. J. Fürnkranz / Dr. G. Grieser
Technische Universität Darmstadt — Wintersemester 2006/07
Termin: 14. 2. 2007

Name: **Vorname:** **Matrikelnummer:**

Fachrichtung: **Wiederholer:** ☐ ja ☐ nein

Punkte: (1) .... (2) .... (3) .... (4) .... (5) .... **Summe:**

- **Aufgaben:** Diese Klausur enthält auf den folgenden Seiten 5 Aufgaben zu insgesamt 100 Punkten. Jede Aufgabe steht auf einem eigenen Blatt. Kontrollieren Sie *sofort*, ob Sie alle sechs Blätter erhalten haben!
- **Zeiteinteilung:** Die Zeit ist knapp bemessen. Wir empfehlen Ihnen, sich zuerst einen kurzen Überblick über die Aufgabenstellungen zu verschaffen, und dann mit den Aufgaben zu beginnen, die Ihnen am besten liegen.
- **Papier:** Verwenden Sie nur Papier, das Sie von uns ausgeteilt bekommen. Bitte lösen Sie die Aufgaben auf den dafür vorgesehenen Seiten. Falls der Platz nicht ausreicht, vermerken sie dies bitte und setzen die Lösung auf der letzten Seite fort. Brauchen Sie zusätzlich Papier (auch Schmierpaper), bitte melden.
- **Fragen:** Sollten Sie Teile der Aufgabenstellung nicht verstehen, bitte fragen Sie!
- **Abschreiben:** Sollte es sich (wie in den letzten Jahren leider immer wieder) herausstellen, daß Ihre Lösung und die eines Kommilitonen über das zu erwartende Maß hinaus übereinstimmen, werden beide Arbeiten negativ beurteilt (ganz egal wer von wem in welchem Umfang abgeschrieben hat).
- **Ausweis:** Legen Sie Ihren *Studentenausweis* und *Lichtbildausweis* sichtbar auf Ihren Platz. Füllen Sie das Deckblatt sofort aus!
- **Hilfsmittel:** Zur Lösung der Aufgaben ist ein von Ihnen selbst handschriftlich beschriebenes DIN-A4-Blatt erlaubt. Gedruckte Wörterbücher sind für ausländische Studenten erlaubt, elektronische Hilfsmittel (Taschenrechner, elektronische Wörterbücher, Handy, etc.) sind verboten! Sollten Sie etwas verwenden wollen, was nicht in diese Kategorien fällt, bitte klären Sie das *bevor* Sie zu arbeiten beginnen.
- **Aufräumen:** Sonst darf außer Schreibgerät, Essbarem, von uns ausgeteiltem Papier und eventuell Wörterbüchern nichts auf Ihrem Platz liegen. Taschen bitte unter den Tisch!

Gutes Gelingen!

Aufgabe 1 Verschiedenes (16 Punkte)

1–a Zur Beantwortung der folgenden Fragen sollten Sie keine Berechnungen durchführen müssen.

- Wie groß ist die Entropie einer Datenmenge, die nur Beispiele einer Klasse enthält?

- Was passiert, wenn nun Beispiele einer anderen Klasse hinzukommen, wird die Entropie steigen oder sinken?

- Bei welcher Klassen-Verteilung der n Trainingsbeispiele wird die Entropie maximal?

1–b Sie wollen die Genauigkeit eines k-NN Klassifizierers abschätzen. Sie haben die Optionen

- Leave-One-Out Cross-Validation auf Ihrer Trainingsmenge mit n Beispielen
- Test an einer neuen, von den Trainingsdaten unabhängigen Stichprobe von $n - 1$ Beispielen

Diskutieren Sie die zu erwartende Laufzeit beider Verfahren.

1–c Gegeben seien die beiden Assoziationsregeln

R_1: Brot, Käse → Milch

R_2: Brot → Käse, Milch

- Welche der folgenden Möglichkeiten können zutreffen:
  - ☐ support(R_1) < support(R_2)
  - ☐ support(R_1) = support(R_2)
  - ☐ support(R_1) > support(R_2)
- Welche der folgenden Möglichkeiten können zutreffen:
  - ☐ confidence(R_1) < confidence(R_2)
  - ☐ confidence(R_1) = confidence(R_2)
  - ☐ confidence(R_1) > confidence(R_2)

1–d Erklären Sie kurz, warum Windowing mit Noise in den Daten Probleme hat.
Ist es in dieser Hinsicht ähnlicher zu Bagging oder zu Boosting?

1–e Sie verwenden den Candidate Elimination-Algorithmus und stellen fest, daß $S = G$ wird.
Was folgern Sie daraus?

Aufgabe 2 Naive Bayes (20 Punkte)

Gegeben seien die folgenden Daten, die angeben, welche Sportart eine Person abhängig von den meteorologischen Daten des jeweiligen Tages ausübt.

Aussicht	Temperatur	Luftfeuchtigkeit	Windstärke	Sportart
Sonnig	Kalt	Hoch	Schwach	Golf
Bewölkt	Kalt	Niedrig	Stark	Golf
Sonnig	Warm	Niedrig	Schwach	Golf
Regen	Kalt	Hoch	Schwach	Squash
Sonnig	Kalt	Hoch	Schwach	Squash
Regen	Warm	Hoch	Stark	Squash
Bewölkt	Kalt	Hoch	Schwach	Squash
Regen	Warm	Hoch	Schwach	Squash
Bewölkt	Warm	Hoch	Schwach	Tennis
Bewölkt	Kalt	Niedrig	Stark	Tennis
Sonnig	Kalt	Niedrig	Stark	Tennis
Bewölkt	Kalt	Hoch	Schwach	Tennis

2–a Nennen Sie alle Wahrscheinlichkeiten, die Sie aus den Daten schätzen müssen, um einen vollständigen Naive Bayes Klassifizierer zur Vorhersage der ausgeübten Sportart zu erhalten.

Hinweis: Sie müssen die Wahrscheinlichkeiten (noch) nicht ausrechnen! Sie können die Bezeichnungen der Attribute und Attributwerte auch (eindeutig) abkürzen.

2–b Berechnen Sie aus den Daten Schätzwerte für diejenigen Wahrscheinlichkeiten, die der Naive Bayes Klassifizierer benötigt, um das Beispiel

Aussicht	Temperatur	Luftfeuchtigkeit	Windstärke	Sportart
?	Kalt	Hoch	Schwach	?

zu klassifizieren.

Welche Klasse wird dann für dieses Beispiel vorhergesagt?

Hinweis: Die Wahrscheinlichkeiten sollen einfach durch die relative Häufigkeit des Auftretens des fraglichen Ereignisses geschätzt werden, und können selbstverständlich als Bruchzahlen angegeben werden.

2–c Für das folgende Beispiel ist es (nach obiger Methode) nicht möglich, eine begründete Klassifikation vorzunehmen.

Aussicht	Temperatur	Luftfeuchtigkeit	Windstärke	Sportart
Regen	Warm	Niedrig	Schwach	?

Auf welches Problem ist das zurückzuführen? Wie können Sie dieses Problem mit einer in der Vorlesung besprochenen Methode beheben?

Aufgabe 3 Regellernen (22 Punkte)

Gegeben folgende Beispielmenge:

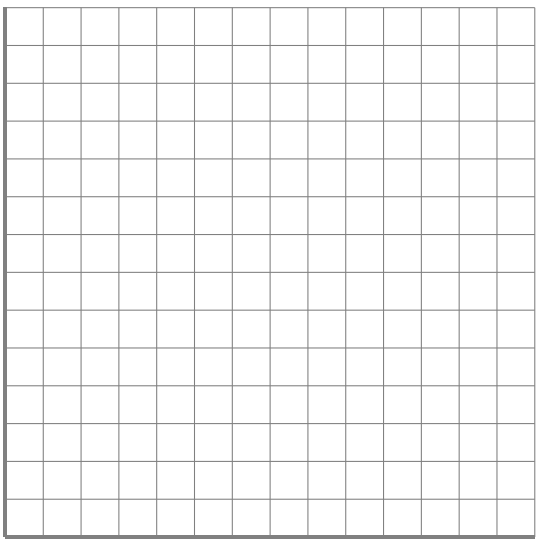
Beispiel	humidity	outlook	wind	temperature	Play?
1	high	sunny	true	mild	+
2	high	overcast	false	cool	+
3	high	sunny	false	hot	+
4	normal	rainy	true	hot	+
5	high	rainy	false	mild	+
6	high	overcast	false	cool	+
7	high	sunny	true	mild	−
8	high	overcast	false	hot	−
9	normal	sunny	false	mild	−
10	normal	rainy	false	cool	−

Aus diesen Daten hat ein Separate-and-Conquer Algorithmus bisher folgende Regel gelernt, wobei die beiden Bedingungen in der angegebenen Reihenfolge (von links nach rechts) gelernt wurden.

R_1: **if** humidity = high **and** wind = false **then** +

3–a Was ist die beste Verfeinerung von R_1, die der Regellerner finden kann (unabhängig von einer konkreten Heuristik) ?

3–b Zeichnen Sie das Lernen der Regel R_1 im Coverage Space. Beschriften Sie die Dimensionen des Raums, sowie alle gezeichneten Punkte mit den jeweiligen Koordinaten (konkrete Zahlen). Zeichnen Sie auch die im vorigen Punkt gefundene Verfeinerung ein.

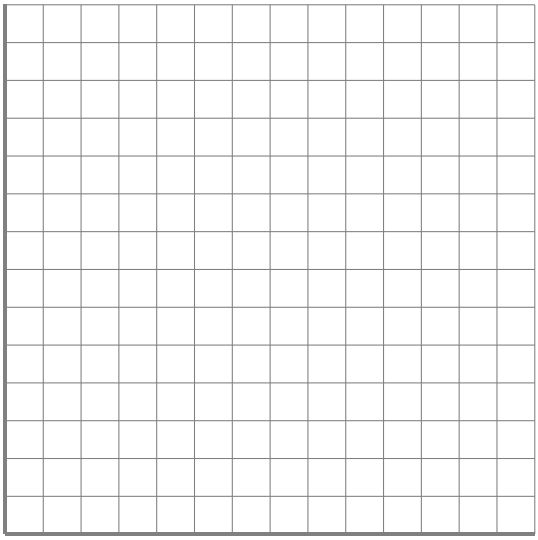


Hinweis: Der karierte Teil des Papiers ist als Zeichenhilfe gedacht. Er gibt keinerlei Aufschluss über die tatsächlichen Dimensionen des Raumes.

3–c Zeichnen Sie für jede der folgenden Heuristiken eine Linie im Coverage Raum, die alle Regeln zeigt, die von der Heuristik gleich gut bewertet werden wie R_1.

Sie müssen natürlich auch wieder den Coverage Space und die Regeln aus der vorigen Aufgabe eintragen. Alternativ können Sie die Lösung dieser Aufgabe auch in der vorherigen Zeichnung eintragen.

- Precision
- Accuracy
- Weighted Relative Accuracy

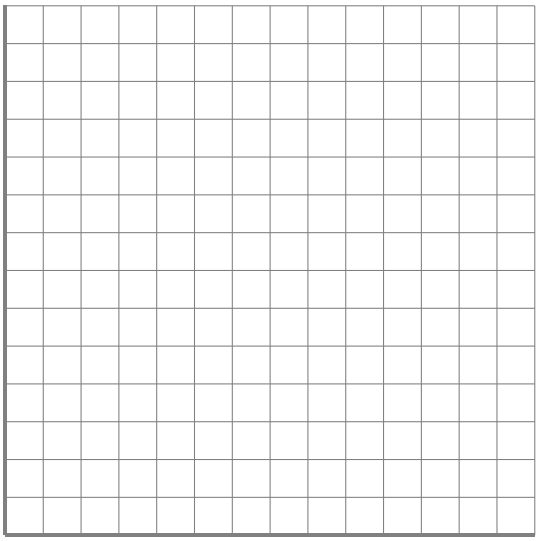


3–d Angenommen der Lerner würde die Regel R_1 in die Theorie aufnehmen. Welche der drei oben genannten Heuristiken würde(n) diese Auswahl ergeben? Geben Sie eine auf den Isometrien der Heuristiken beruhende Begründung an (verwenden Sie die Ergebnisse aus der vorherigen Aufgabe).

3–e Der Lerner findet als nächste Regel die folgende:

R_2: **if** temperature = hot **then** +

Zeichnen und beschriften Sie wiederum wie unter 3-b den Coverage Space, in dem diese Regel gelernt wird, und markieren Sie die Lage der Regel in diesem Raum.

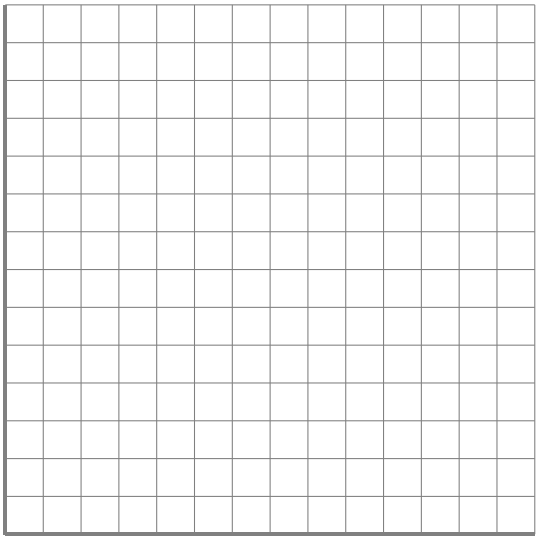


Aufgabe 4 ROC-Raum und kosten-sensitives Lernen (22 Punkte)

Ein Separate-And-Conquer Regellerner findet in einer Trainingsmenge von 200 positiven und 200 negativen Beispielen sukzessive drei Regeln R_1, R_2, R_3 (in dieser Reihenfolge). Die Abdeckung der Beispiele für jede der Theorien ist wie folgt:

Theorie	**abgedeckte Beispiele**	
	positiv	negativ
$\{R_1\}$	100	40
$\{R_1, R_2\}$	120	80
$\{R_1, R_2, R_3\}$	180	120

4–a Skizzieren Sie den Weg des Separate-and-Conquer Regellerners durch den ROC-Raum, der bei der leeren Theorie beginnt, eine Regel nach der anderen hinzufügt, und bei der universellen Theorie endet.



4–b Berechnen Sie die Fläche unter der ROC-Kurve.

4–c Erklären Sie anhand dieses konkreten Falls die Rolle der konvexen Hülle der ROC-Kurve. Was können Sie in diesem Beispiel daraus ablesen?

4–d Ihr Auftraggeber nennt Ihnen drei verschiedene Kosten für jeweils ein misklassifiziertes positives oder negatives Beispiel. Geben Sie für jedes Szenario die geeignetste der vom Regellerner untersuchten Theorien an, und geben Sie die bei der Klassifikation der gegebenen 400 Beispiele entstehenden Kosten an.

c_+	c_-	optimale Theorie(n)	entstandene Kosten
1	3		
2	2		
3	1		

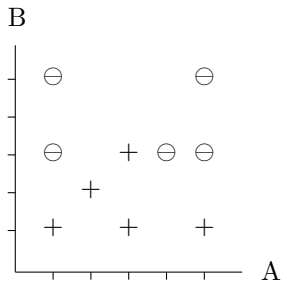
Hinweis: Man kann die jeweils optimale Theorie am schnellsten aus der ROC-Kurve ablesen.

4–e Weiters gibt Ihr Auftraggeber an, daß er maximal 10% *false positives* akzeptieren kann. Wie konstruieren Sie aus den gelernten Regeln einen für dieses Szenario passenden Klassifizierer?

Aufgabe 5 Ensemble Methods (20 Punkte)

Gegeben seien folgende Beispiele einer numerischen Domäne:

A	B	Class	A	B	Class
1	1	+	1	3	⊖
2	2	+	1	5	⊖
3	1	+	4	3	⊖
3	3	+	5	3	⊖
5	1	+	5	5	⊖

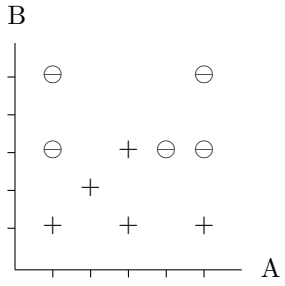


Die Hypothesen-Sprache besteht aus konjunktiven Regeln, die überprüfen können, ob die Attribut-Werte für einzelne Beispiele $<$ oder $>$ als Konstanten sind. Der Head einer Regel ist immer +. Eine gültige Regel wäre also z.B.

if $A > 2$ **and** $A < 4$ **and** $B < 2$ **then** +

Alle von der Regel nicht abgedeckten Beispiele würden als negativ klassifiziert werden.

5–a Welche einzelne Regel ist die beste (i.e., hat den geringsten Fehler), die auf dieser Datenmenge gefunden werden kann?

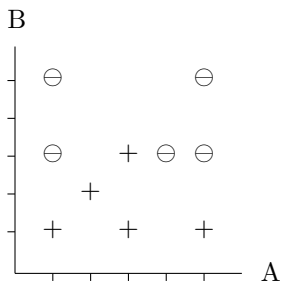


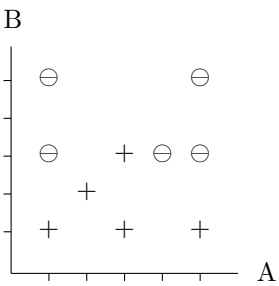
5–b Mit Bagging werden folgende Regeln gelernt:

R_1: **if** $B < 4$ **then** +
R_2: **if** $A < 4$ **and** $B < 3$ **then** +
R_3: **if** $B < 3$ **then** +

Skizzieren Sie die Regionen, die als + bzw. − klassifiziert werden, wenn die einzelnen Klassifikatoren

- mit ungewichtetem Voting kombiniert werden
- mit gewichtetem Voting kombiniert werden, wobei jede Regel R_i mit $\alpha_i = \frac{1}{fp_i + fn_i}$ gewichtet wird ($fp_i + fn_i$ ist die Anzahl der von R_i falsch klassifizierten Beispiele).





5–c Nehmen Sie an, daß die Beispiele zu Beginn gleich gewichtet werden, i.e., $w_{i,0} = 1/10$ für alle Beispiele i. Führen Sie zwei Iterationen des Boosting-Algorithmus durch:

1. finden Sie eine Theorie mit minimalem gewichteten Fehler
   (falls es mehr als eine gibt, können Sie eine beliebige auswählen)
2. geben Sie den gewichteten Fehler dieser Theorie an
3. bestimmen Sie die Änderung in den Gewichten der Beispiele
4. goto 1.

Verwenden Sie dabei folgende Update-Regel:

$$w_{i,j+1} \leftarrow \begin{cases} 3 \cdot w_{i,j} & \text{wenn Beispiel } i \text{ fehlerhaft klassifiziert wurde} \\ w_{i,j} & \text{wenn Beispiel } i \text{ korrekt klassifiziert wurde} \end{cases}$$

Hinweis: Vergessen Sie nicht auf die Normalisierung der Gewichte. Sie können die untenstehenden Diagramme für Skizzen verwenden.