

Maschinelles Lernen: Symbolische Ansätze

Prof. J. Fürnkranz

Technische Universität Darmstadt — Wintersemester 2014/15

Termin: 17. 2. 2015

Name:

Vorname:

Matrikelnummer:

Fachrichtung:

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 Blätter erhalten haben!
- **Zeiteinteilung:** Die Zeit ist mit 90 min knapp bemessen. Wir empfehlen Ihnen, sich zuerst einen kurzen Überblick über die Aufgabenstellungen zu verschaffen, und dann mit den Aufgaben beginnen, die Ihnen am meisten liegen.
- **Papier:** Verwenden Sie nur Papier, das Sie von uns ausgeteilt bekommen. Sie können Ihre Lösungen beliebig auf die Blätter verteilen, solange klar ersichtlich ist, welche Lösung zu welcher Aufgabe gehört. Sollten sich allerdings mehrere Lösungen zu derselben Aufgabe finden, suchen wir uns eine aus.
Insbesondere können Sie auch auf den Rückseiten schreiben!
Brauchen Sie zusätzlich Papier (auch Schmierpapier), 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* sichtbar auf Ihren Platz.
- **Hilfsmittel:** Zur Lösung der Aufgaben ist ein von Ihnen selbst handschriftlich beschriebenes DIN-A4-Blatt erlaubt. Gedruckte Wörterbücher sind für ausländische Studierende 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, Mäntel auf das Pult hinter Ihnen!

Gutes Gelingen!

Aufgabe 1 Verschiedenes (3/3/3/3/3/4 = 22 Punkte)

- 1-a Was ist der Zweck des Einsatzes von kd -Trees bei einem k -NN Klassifizierer?
- 1-b Erklären Sie in Worten (ohne Formeln) die Begriffe *False Positives* und *False Positive Rate*.
- 1-c In welchem Fall können Sie beim Candidate Elimination Algorithmus angeben, daß Sie keine weitere Trainingsbeispiele mehr benötigen?
- 1-d Was haben alle Klassifizierer, die im Versionsraum enthalten sind, in Bezug auf die Trainingsdaten gemeinsam? (genau eine Antwort ist richtig)
- Sie sind alle weder vollständig noch konsistent.
 - Sie sind alle vollständig, aber nicht notwendigerweise konsistent.
 - Sie sind alle konsistent, aber nicht notwendigerweise vollständig.
 - Sie sind alle sowohl konsistent als auch vollständig.
 - Die Klassifizierer im G-Set sind alle vollständig, die Klassifizierer im S-Set sind alle konsistent, bei allen anderen läßt sich das nicht mit Sicherheit sagen.

- 1-e Sie trainieren mit Bagging ein Ensemble von Klassifizierern. Wofür können Sie *Stacking* hier einsetzen?
- 1-f Sowohl *Boosting* als auch *Bagging* können realisiert werden, indem in jeder Iteration aus der ursprünglichen Trainingsmenge der Größe n eine neue Trainingsmenge der Größe n gesampelt wird. Worin unterscheiden sich die dabei verwendeten Sampling-Strategien?
- 1-g Gegeben sei eine Heuristik $h(x) \rightarrow [0, 1]$, die jeder Regel einen Wert zwischen 0 und 1 zuweist, und eine Heuristik $g(x) = h^2(x)$, die die Heuristik $h(\cdot)$ quadriert. Haben h und g dieselben Isometrien? Warum (nicht)?

Aufgabe 2 Decision Trees (3/9/4/4 = 20 Punkte)

Gegeben seien die folgenden 10 Trainingbeispiele zur Vorhersage von T aus den drei nominalen Attributen L , M , und N .

L	M	N	T	L	M	N	T
a	p	x	+	a	p	z	-
a	m	x	+	a	m	z	-
b	m	x	-	b	m	z	-
b	p	x	-	b	m	y	-
a	p	y	+	a	m	y	+

- 2-a Gegeben sei folgendes Maß zur Messung der Unreinheit einer Menge E mit e_+ positiven und e_- negativen Beispielen.

$$M(E) = \min \left(\frac{e_+}{e_+ + e_-}, \frac{e_-}{e_+ + e_-} \right)$$

Definieren Sie basierend auf diesem Maß eine Gain-Heuristik, analog zur Definition des Information Gain Maßes, wo Entropie als Maß für die Unreinheit verwendet wird.

$$MGain(E, A) = \dots$$

- 2-b Verwenden Sie die in a) definierte Gain-Heuristik, um aus diesen Daten einen Entscheidungsbaum zu lernen.

Anm: Wenn Sie a) nicht lösen konnten, können Sie auch eins der in der Vorlesung kennen gelernten Maße verwenden (z.B. Gini-Index).

2-c Konvertieren Sie den entstandenen Baum in eine Regelmenge und geben Sie diese an. Welche Bedingungen der Regelmenge können geprunt werden, ohne daß sich der Fehler auf den Trainingsdaten erhöht?

2-d Welche Methode würden Sie auf dieser Datenmenge zur Evaluierung Ihres Entscheidungsbaumlerners verwenden (genau eine Antwort ist richtig):

- 10-fold Cross-Validation
- Leave-One-Out Cross-Validation
- egal

Begründen Sie Ihre Antwort.

Aufgabe 3 Multi-Klassen- und Meta-Klassifikation (4/7/3/3 = 17 Punkte)

Gegeben sei folgender Multi-Klassen-Datensatz:

ID	A1	A2	A3	Klasse
#1	0	1	0	b
#2	0	1	1	b
#3	0	0	0	a
#4	0	0	1	b
#5	1	1	0	c
#6	1	0	0	c

3-a Sie wollen diesen Datensatz mit einem Konzept-Lerner bearbeiten und zu diesem Zweck *paarweise Klassifikation* einsetzen. Geben Sie die Trainingsmengen an, die Sie für den Konzept-Lerner generieren müssen.

3-b Als Alternative werden Ihnen Error-Correcting Output Codes vorgeschlagen, wobei Sie folgende Coding Matrix verwenden sollen, bei der die Klassen in der Reihenfolge (a, b, c) kodiert werden:

$$C = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \end{pmatrix}$$

Beantworten Sie folgende Fragen:

1. Wie viele binäre Klassifizierer müssen Sie trainieren?
2. Wie viele Trainingsbeispiele werden für diese binären Klassifizierer *insgesamt* (über alle Trainingssets) als positiv, wie viele als negativ gelabelt?
3. Angenommen die Klassifikatoren liefern folgenden Klassen-Vektor zurück:

$$c = (0 \ 0 \ 0 \ 0 \ 1)$$

Welche der Klassen (a, b, c) würden Sie vorhersagen? Warum?

3-c Wie sähe die Coding Matrix aus, wenn Sie mit Hilfe von ECOCs eine One-Against-All Klassifikation realisieren wollten?

3-d Sie verwenden die Beispiele $\{\#3, \#4, \#5\}$ in einer ersten Iteration von Windowing. Dabei wird folgende *Decision List* gelernt:

```
IF       $A2 = 0 \wedge A3 = 0$  THEN   $a$   
ELSIF   $A3 = 1$           THEN   $b$   
ELSE                                      $c$ 
```

Welche Beispiele werden für die nächste Iteration von Windowing als Trainingsmenge verwendet und warum?

Aufgabe 4 Evaluierung (5/3/3/5/4 = 20 Punkte)

Für die folgenden 10 Beispiele x ist jeweils ihre wahre Klasse $h(x)$ angegeben, sowie die Wahrscheinlichkeit $P(+|x)$, mit der ein Lerner das Beispiel als positiv klassifizieren würde.

x	1	2	3	4	5	6	7	8	9	10
$h(x)$	+	+	-	+	-	+	-	-	+	-
$P(+ x)$	0.95	0.25	0.15	0.55	0.35	0.75	0.85	0.45	0.65	0.05

4-a Zeichnen Sie die ROC-Kurve für die oben gegebenen 10 Beispiele.



Hinweis: Der oben stehende karierte Teil des Papiers ist als Zeichenhilfe gedacht. Sie müssen diese aber nicht verwenden, bzw. können Sie diese in einem beliebigen Maßstab verwenden. Vergessen Sie nicht, die Begrenzungen des ROC-Raums einzuzeichnen.

4-b Bestimmen Sie die Fläche unter ROC-Kurve (AUC).

-
- 4-c Die Fläche unter der ROC-Kurve kann auch als eine Wahrscheinlichkeit betrachtet werden. Beschreiben Sie in Worten, was diese Wahrscheinlichkeit misst.
- 4-d Sie wollen nun den Ranker in einen Klassifikator K_t verwandeln, indem Sie einen Threshold t für die Wahrscheinlichkeit $P(+|x)$ einführen, sodaß alle Beispiele mit $P(+|x) > t$ als positiv, alle anderen als negativ klassifiziert werden. Bestimmen Sie aus der ROC-Kurve einen möglichen Threshold, der die Genauigkeit (*Accuracy*) des Klassifikators maximiert, und berechnen Sie diese maximale Genauigkeit.
- 4-e Vergleichen Sie den Klassifizierer $K_{0.5}$, der alle Beispiele mit $P(+|x) > 0.5$ als positiv klassifiziert, mit dem Klassifizierer $K_{0.2}$, der alle Beispiele mit $P(+|x) > 0.2$ als positiv klassifiziert. Ab welchem Kostenverhältnis $\frac{c_-}{c_+}$ ist es besser, $K_{0.2}$ zu verwenden als $K_{0.5}$?

Aufgabe 5 Assoziationsregeln (3/3/3/2/3/3/4 = 21 Punkte)

Ein Obstladen macht folgende Beobachtungen über das Kaufverhalten seiner Kunden:

Kunde	Einkaufskorb
K1	<i>Apfel, Birne, Clementine, Dattel, Erdbeere</i>
K2	<i>Apfel, Clementine, Erdbeere</i>
K3	<i>Apfel, Dattel, Erdbeere</i>
K4	<i>Birne, Dattel</i>
K5	<i>Apfel, Clementine, Dattel, Erdbeere</i>
K6	<i>Apfel, Birne, Dattel, Erdbeere</i>

Hier ist eine Liste aller Paare von Produkten, die von zumindest 40% der Kunden gekauft wurden:

$$S_2 = \{ \{ \text{Apfel, Clementine} \}, \{ \text{Apfel, Dattel} \}, \{ \text{Apfel, Erdbeere} \}, \\ \{ \text{Birne, Dattel} \}, \{ \text{Clementine, Erdbeere} \}, \{ \text{Dattel, Erdbeere} \} \}$$

Hinweis: Sie können die Obstsorten mit ihren Anfangsbuchstaben abkürzen.

5-a Welche Kandidaten für Itemsets der Länge 3 werden mittels der in der Vorlesung besprochenen Methode von APRIORI generiert?

Hinweis: Die Items in einem Warenkorb werden alphabetisch sortiert!

5-b Welche der unter a) generierten Itemsets können aus C_3 entfernt werden *bevor* deren Häufigkeit gezählt werden muß?

5-c Welche Itemsets bilden letztendlich die Menge S_3 der häufigen Itemsets der Länge 3? Geben Sie für jedes Mitglied von S_3 den Support an.

5-d Berechnen Sie den Lift für die Assoziationsregel *Apfel* → *Erdbeere*.

Hier nochmals die Einkaufskörbe:

Kunde	Einkaufskorb
K1	<i>Apfel, Birne, Clementine, Dattel, Erdbeere</i>
K2	<i>Apfel, Clementine, Erdbeere</i>
K3	<i>Apfel, Dattel, Erdbeere</i>
K4	<i>Birne, Dattel</i>
K5	<i>Apfel, Clementine, Dattel, Erdbeere</i>
K6	<i>Apfel, Birne, Dattel, Erdbeere</i>

$$S_2 = \{ \{Apfel, Clementine\}, \{Apfel, Dattel\}, \{Apfel, Erdbeere\}, \\ \{Birne, Dattel\}, \{Clementine, Erdbeere\}, \{Dattel, Erdbeere\} \}$$

5-e Welches der sechs Elemente aus S_2 gehört der positiven Border an?

5-f Geben Sie ein Beispiel für ein 2-elementiges Itemset der negativen Border.

5-g Geben Sie alle Assoziationsregeln mit $c_{min} = 0.8$ an, die sich aus dem Itemset $\{Apfel, Erdbeere, Clementine\}$ bilden lassen.