

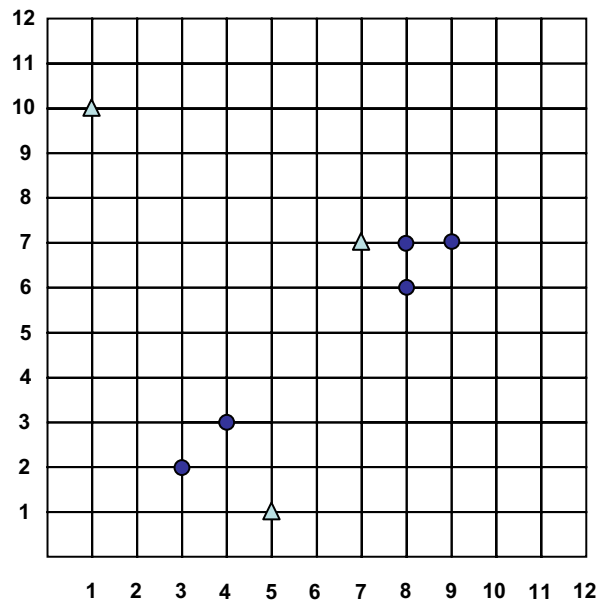
Managing Massive Multiplayer Online Games  
SS 2013

Übungsblatt 7: Klassifikation / Clustering / Signifikanztests

Besprechung: 27./28.06.2013

**Aufgabe 7-1** *Clustering mit Varianzminimierung*

Gegeben sei erneut folgender Datensatz mit 8 Punkten (2-dimensionalen Featurevektoren).



Im folgenden sollen vollständige Partitionierungen des Datensatzes in  $k = 2$  Cluster berechnet werden. Als Distanzfunktion zwischen den Punkten soll dabei die Manhattan-Distanz ( $L_1$ -Norm) verwendet werden.

- (a) Erzeugen Sie eine Partitionierung in  $k = 2$  Cluster mit dem einfachen Verfahren "Clustering durch Varianz Minimierung". Die initiale Partitionierung der Daten ist durch die Dreiecke und Punkte gegeben (die Dreiecke bilden einen initialen Cluster, genauso die Punkte). Beschreiben Sie jede Aktion des Algorithmus. Zeichnen Sie nach jedem Schritt die Zentroiden ein und markieren Sie die Punkte anhand ihrer Clusterzugehörigkeit. Denken Sie daran, bei der Zuordnung zu den Zentroiden die  $L_1$ -Norm zu verwenden.

Tipp: Hierzu können Sie die Vorlage auf der letzten Seite benutzen, die Sie am besten mehrmals kopieren.

- (b) Zeigen Sie, dass dieses Verfahren von der Wahl des initialen Clusterings abhängt.

**Aufgabe 7-2** *Klassifikation*

Betrachten Sie im Folgenden ein abstraktes Spiel, bei dem ein Spieler regelmäßig eine Auswahl aus mehreren Entscheidungsmöglichkeiten treffen muss. Beispiele sind:

- In welcher Reihenfolge sollen Gegenstände aufgesammelt werden.
- In welche Richtung soll der Spieler an einer Kreuzung in einem Labyrinth gehen.
- In welche Richtung schickt der Spieler seine Einheit zum Erkunden.

Wir nehmen an, dass es immer vier Alternativen  $\{a_1, \dots, a_4\}$  gibt. Es sei bekannt, dass ein BOT-programm bei dieser Entscheidung jede Alternative mit gleicher Wahrscheinlichkeit auswählt. Aus Log-Dateien sei außerdem empirisch geschätzt worden, dass reale Spieler ihre Entscheidung folgendermaßen auf die Alternativen verteilen:

	$a_1$	$a_2$	$a_3$	$a_4$
Wahrscheinlichkeit	10%	20%	30%	40%

Bei einem Spieler  $S_1$  sei folgende Sequenz von Entscheidungen beobachtet worden:

$$O = [a_3, a_2, a_1, a_4, a_1, a_2, a_2, a_3, a_1]$$

Im Folgenden sei  $B$  das Zufallsereignis, dass es sich bei  $S_1$  um einen BOT handelt, und  $\bar{B}$  sei das Zufallsereignis, dass es sich bei  $S_1$  um einen realen Spieler handelt.

- Berechnen Sie die Wahrscheinlichkeit  $P(O|B)$ , dass der BOT die obige Sequenz erzeugt.
- Berechnen Sie die Wahrscheinlichkeit  $P(O|\bar{B})$ , dass ein realer Spieler die obige Sequenz erzeugt.
- Angenommen wir wissen, dass es sich bei 1% aller Spieler um BOTs handelt. Berechnen Sie die Wahrscheinlichkeit  $P(B|O)$ , dass es sich bei  $S_1$  um einen BOT handelt.

### Aufgabe 7-3 *Balancing*

Bei einem Spiel können Spieler zu Beginn zwischen mehreren verschiedenen Voreinstellungen (z.B. Rassen, Klassen, Fraktionen) auswählen. Seien  $v_1, \dots, v_n$  solche Voreinstellungen.

Nehmen Sie an, dass 1,000 Spielen zwischen jeweils einem Spieler mit Voreinstellung  $v_1$  und einem Spieler mit Voreinstellung  $v_2$  (kurz:  $v_1$  vs  $v_2$ ) beobachtet wurden, wobei diese Spiele in 400 Fällen zugunsten von dem Spieler mit Voreinstellung  $v_1$  verliefen.

- Ist das Spiel fair bezüglich Voreinstellungen  $v_1$  und  $v_2$ ? Berechnen Sie dazu die Wahrscheinlichkeit der Beobachtung, unter der Annahme dass das Spiel fair ist, dass die Gewinnwahrscheinlichkeit also für beide Spieler immer 50% beträgt.
- Eine Begegnung zwischen  $v_i$  und  $v_j$  soll fair sein, falls die entsprechende Wahrscheinlichkeit aus Teilaufgabe a) mindestens 5% beträgt. Nehmen Sie für  $n = 10$  an, dass alle mögliche Begegnungen fair sind. Wie hoch ist die Wahrscheinlichkeit dass mindestens eine Begegnung fälschlicherweise als unfair eingestuft wird?

