

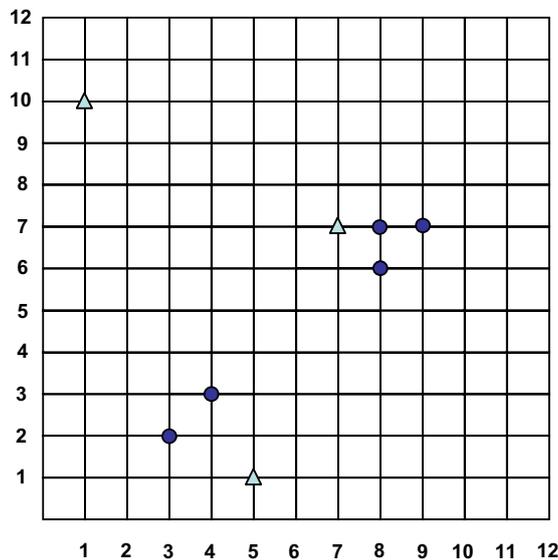
**Managing Massive Multiplayer Online Games**  
 SS 2012

**Übungsblatt 6: Klassifikation und Clustering**

Besprechung: 21.06.2012

**Aufgabe 6-1** *Instanzbasierte Klassifikation*

Gegeben sei folgender Datensatz mit 8 Punkten (2-dimensionalen Featurevektoren). Die Klassen dieser Punkte seien durch Dreiecke und Punkte gegeben (die Menge der Dreiecke bilden eine Klasse, genauso die Menge der Punkte).



Im folgenden sollen die Klassen von Punkten mittels k-Nächste-Nachbarn Klassifikation bestimmt werden. Als Distanzfunktion zwischen den Punkten soll dabei die Manhattan-Distanz ( $L_1$ -Norm) verwendet werden, die für zwei Punkte  $x, y$  wie folgt definiert ist:

$$L_1(x, y) = \sum_{i=1}^d |x_i - y_i|$$

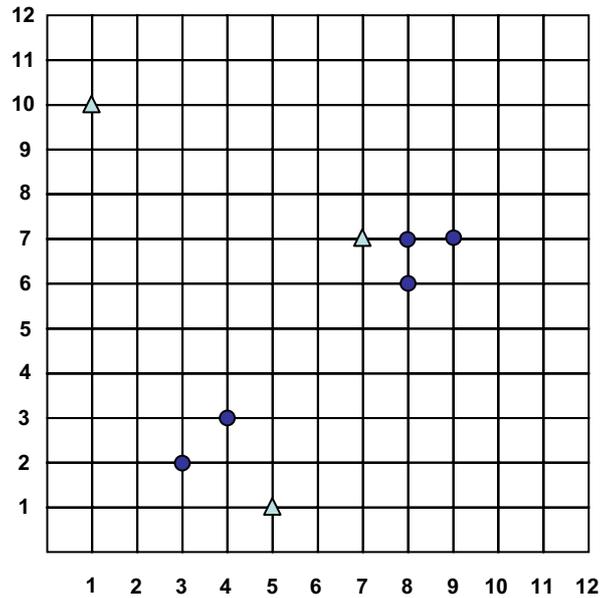
- (a) Bestimmen Sie die Klassen von Punkt (2,7) für  $k = 3$  und Mehrheitsklasse in der Entscheidungsmenge.
- (b) Bestimmen Sie die Klassen von Punkt (2,7) für  $k = 5$  und Mehrheitsklasse in der Entscheidungsmenge.
- (c) Bestimmen Sie die Klassen von Punkt (6,1) für  $k = 3$  und Mehrheitsklasse in der Entscheidungsmenge.
- (d) Bestimmen Sie die Klassen von Punkt (6,1) für  $k = 3$  und Mehrheitsklasse und Gewichtung nach inverser Manhattan-Distanz.

$$L_1(x, y)^{-1} = \frac{1}{\sum_{i=1}^d |x_i - y_i|}$$

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

**Aufgabe 6-2**    *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) Begründen Sie kurz, warum dieses Verfahren reihenfolgeabhängig ist.

