

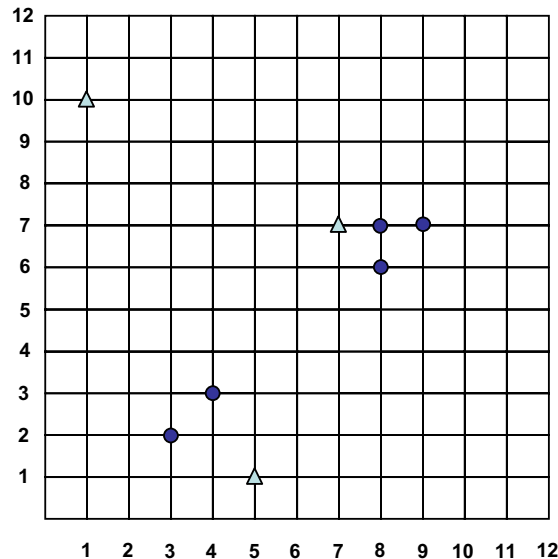
Managing Massive Multiplayer Online Games
 SS 2016

Übungsblatt 6: Knowledge Discovery and Data Mining

Besprechung: 09.06.2016

Aufgabe 6-1 *Supervised Learning: Instanzbasiertes Lernen: Klassifikation mit kNN*

Gegeben sei folgender Datensatz mit 8 Punkten (z.B. zweidimensionale Featurevektoren). Die Menge der Dreiecke bilde eine Klasse, ebenso 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:

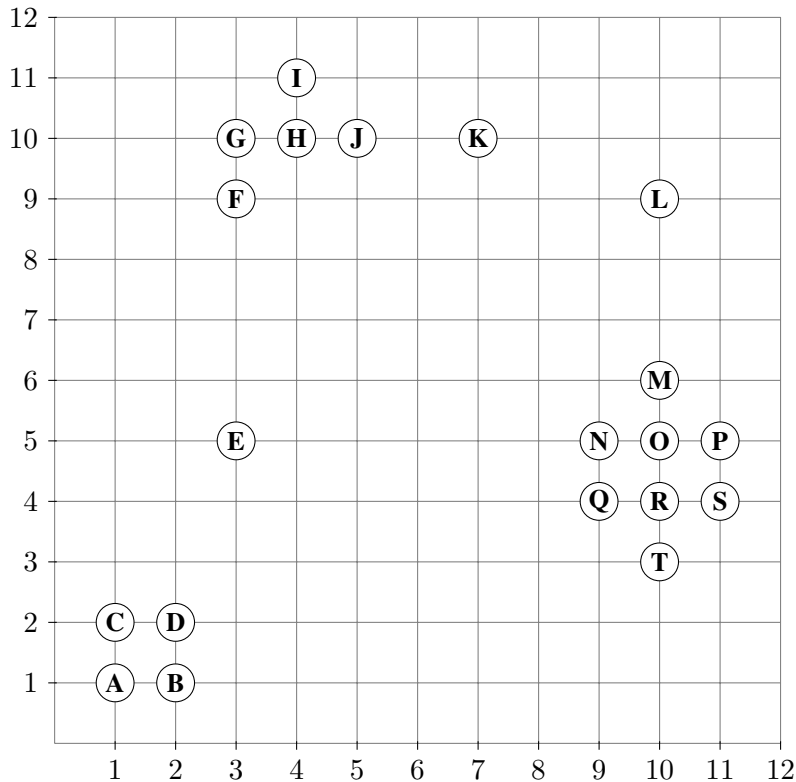
$$L_1(x, y) = \sum_{i=1}^d |x_i - y_i|$$

- (a) Bestimmen Sie die Klassen von Punkt (2,7) für $k = 2$ und Mehrheitsklasse in der Entscheidungsmenge, d.h., der Punkt wird der Klasse zugeordnet, die am häufigsten unter den k -nächsten Nachbarn vorkommt.
- (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.

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

Aufgabe 6-2 *Unsupervised Learning: Clustering mit DBSCAN*

Gegeben sei folgender Datensatz:



Clustern Sie diesen Datensatz mit Hilfe des DBSCAN-Algorithmus. Verwenden Sie als Distanzfunktion die Manhattan Distanz und verwenden Sie dabei die Parameter $\epsilon = 1.1$ und $minPts = 3$.

Aufgabe 6-3 *Bot Detection mit Bayes*

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 Einheiten 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.

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

Aufgabe 6-4 *Probabilistisches 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.