

Maschinelles Lernen

Übungsblatt 7

Prof. Dr. Dr. Lars Schmidt-Thieme, Osman Akcatepe
Wirtschaftsinformatik und Maschinelles Lernen (ISMLL)
Universität Hildesheim

10. Januar 2012
Abgabe: 16. Januar bis 14.15

Nächste-Nachbar-Verfahren (5 Punkte)

Gegeben seien 12 Städte mit den folgenden Koordinaten:

Stadt i	x_i	y_i
1	11	5
2	6	4
3	4	10
4	4	2
5	2	4
6	7	7
7	8	8
8	9	2
9	5	7
10	7	1
11	1	6
12	11	11

Die Distanz zwischen der Stadt a mit Koordinaten (a_1, a_2) und der Stadt b mit Koordinaten (b_1, b_2) ist durch folgende Formel definiert:

$$d(a, b) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}$$

- Legen Sie ein Koordinatensystem an, in welches Sie die 12 Städte einzeichnen.
- Bestimmen Sie zu den gegebenen 12 Städten die zugehörige Distanzmatrix $\mathbf{D} = (d(a, b))$. Schicken Sie die Lösung dieser Teilaufgabe per E-Mail.
- Lösen Sie das TSP (*travelling salesman problem*, das Finden der optimalen Reiseroute) mit der 1-Nearest-Neighbor-Heuristik, wobei Sie bei der Stadt 1 beginnen. Zeichnen Sie die Wegstrecke in das Koordinatensystem ein. Wie lang ist die zurückgelegte Strecke? Ist der gefundene Weg optimal, d.h. hat er die kleinstmögliche Wegstrecke?
Wählen Sie eine beliebige andere Stadt als Startpunkt und bestimmen Sie von dort aus eine Lösung mit der 1-Nearest-Neighbor-Heuristik. Zeichnen Sie die Wegstrecke in das Koordinatensystem ein.

Programmieren in R (3 Punkte)

Lesen Sie Kapitel 9 und 10 von „An Introduction to R“.

- a) Welche Schleifen-Konstrukte gibt es in R? Warum wird `for()` in R weniger verwendet als in Java oder C?
- b) Woran erkennt man benutzerdefinierte binäre Operatoren in R? Wann ist es bei einem Funktionsaufruf nicht notwendig, Funktionsargumente zu benennen?
- c) Schreiben Sie zwei R-Funktionen, die die Fakultät einer natürlichen Zahl berechnen. Eine der Funktionen soll dabei iterativ vorgehen, die andere soll rekursiv definiert sein. Was passiert, wenn man den Namen einer Funktion in der R-Konsole eingibt?

kNN in R (2 Punkte)

Starten Sie R und laden Sie die Bibliothek `class`. Wenden Sie `kNN` mit unterschiedlichen Parametern k auf den Iris-Datensatz an und vergleichen Sie das Ergebnis mit den LDA-/QDA-Ergebnissen vom fünften Arbeitsblatt. Verwenden Sie zur Auswertung die Funktion `knn.cv()` und lesen Sie in der Dokumentation der Funktion nach, wie Sie den Datensatz passend umwandeln.