

9. Übung zur Vorlesung Künstliche Intelligenz

Institut für Informatik, FU Berlin, SoSe 2005
Prof. Dr. Raúl Rojas, Marco Block, Ernesto Tapia

Neben der schriftlichen Abgabe, sind die Programmieraufgaben **zusätzlich** per e-mail an den Tutor zu schicken. Eine e-mail ersetzt nicht die schriftliche Abgabe! Die folgenden Programmieraufgaben können mit Prolog, C, C++, C#, Delphi oder Java bearbeitet werden. Die beschriebenen Daten sind verfügbar unter

"<http://www.inf.fu-berlin.de/lehre/SS05/KI/Daten/>".

1. Aufgabe (15 Punkte)

- (5 Punkte) In den Tutorien wurde das Format der Datei *"pendigits-orig.test.txt"* vorgestellt. Schreiben Sie ein Programm, das diese Datei ausliest und die entsprechenden Zahlendarstellungen geeignet speichert. Sie sollten in Ihrem Programm eine Möglichkeit bieten die Daten als Bilder anzuzeigen. Z.B. mit `showDigit(k)` sollte der k-te Eintrag angezeigt werden.
- (10 Punkte) Schlagen Sie typische Merkmale und Charakteristika (wie in der Vorlesung angedeutet) der Zahlen 0 - 9 vor. Sie sollten berücksichtigen, dass Ihnen als Datenbasis nur die Datei *"pendigits-orig.test.txt"* zur Verfügung steht. Dementsprechend müssen Sie für jedes Merkmal ein Verfahren beschreiben, das es Ihnen ermöglicht aus der Datei *"pendigits-orig.test.txt"* dieses Merkmal zu extrahieren.

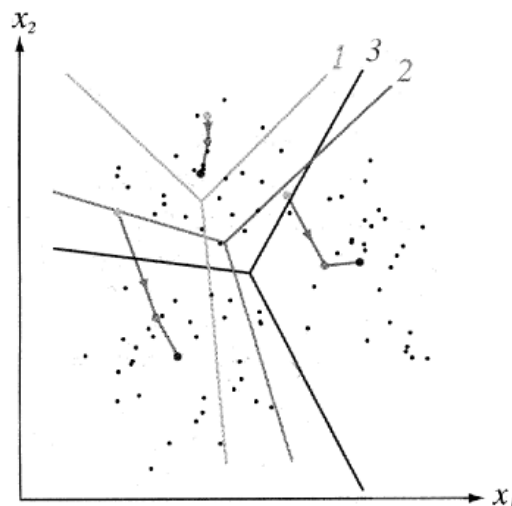


Figure 1: Arbeitsweise des LBG-Algorithmus

2. Aufgabe (10 Punkte)

In der Vorlesung wurde der Linde-Buzzo-Gray-Algorithmus (LBG) vorgestellt.

- a) (4 Punkte) Implementieren Sie den LBG-Algorithmus.
- b) (3 Punkte) Verwenden Sie als Trainingsmenge die Datei "*pendigits.tra.txt*". Es sollen 10 Cluster gelernt werden. Verwenden Sie anschliessend sowohl die Trainingsmenge als Testmenge, als auch folgende Datei "*pendigits.tes.txt*" und geben Sie die beiden Erkennungsraten prozentual an.
- c) (3 Punkte) Diese Aufgabe soll die Qualität Ihres Algorithmus in Abhängigkeit zur Clusteranzahl testen. Geben Sie die Fehlerraten-Funktion (in diesem Fall soll nur die Testmenge "*pendigits.tes.txt*" verwendet werden) für die Clusteranzahlen 5 - 30 (tabellarisch oder grafisch) an.

Abgabe: (Montag) 27.06.2005, 12:00 Uhr (s.t.)
(verspätete Abgaben werden nicht mehr entgegen genommen!)