

Übungsblatt 4

26. November 2008

Hinweis: Soweit nicht anders angegeben, gibt es für jede korrekt bearbeitete Teilaufgabe einen Punkt. Die Abgabe der Hausübungen ist bis spätestens zum Beginn der nächsten Vorlesung möglich – entweder persönlich direkt vor der Vorlesung oder per Einwurf in den Briefkasten des Instituts (Informatikzentrum, zweiter Stock, vor Raum 238).

Aufgabe 8 (Latent Semantic Analysis)

In dieser Aufgabe betrachten wir die folgende Dokumentenkollektion:

Dokumentnummer	Inhalt
1	hello
2	open house
3	mi casa
4	hola Profesor
5	hola y bienvenido
6	hello and welcome

Hinweis: In Ihrem eigenen Interesse sollten Sie die in dieser Aufgabe verlangten Berechnungen nicht per Hand ausführen. Verwenden Sie geeignete Hilfsmittel Ihrer Wahl (z. B. MATLAB oder Octave). Geben Sie an, welche Hilfsmittel Sie verwendet haben und wie Sie dabei vorgegangen sind.

- Bestimmen Sie die Term-Dokument-Matrix A zu dieser Kollektion. Ordnen Sie dabei die Terme nach der Position ihres ersten Auftretens in der obigen Tabelle (also: hello = 1, open = 2, house = 3, ...).
- Berechnen Sie die Singulärwertzerlegung der Matrix A in die Matrizen U , S und V .
- Berechnen Sie die Rang-2-Approximation $A_2 = U_2 S_2 V_2$ der Matrix A .
- Wie groß ist der Frobeniusabstand zwischen A und A_2 ?
- Bestimmen Sie Koordinaten der obigen Terme und Dokumente bezüglich des neuen (zweidimensionalen) Koordinatensystems, das die Rang-2-Approximation von A liefert.
- Zeichnen Sie ein Diagramm, das die Positionen aller Terme und Dokumente bezüglich der neuen Koordinatenachsen zeigt; verwenden Sie das in der Vorlesung gezeigte Beispiel aus (Berry u. a., 1995) als Vorlage.
- Berechnen Sie die Koordinaten der Anfrage „welcome“ und zeichnen Sie diese in das Diagramm aus dem vorherigen Aufgabenteil ein.
- Beschreiben Sie anschaulich, was die einzelnen Einträge der Matrizen AA^T und $A_2 A_2^T$ darstellen. (2 Punkte)