

Maschinelles Lernen und Data Mining
Sommersemester 2011
Übungsblatt 7

Besprechung des Übungsblattes am 07./08.07.2011

Aufgabe 7-1 Wahrscheinlichkeitsrechnung

Gegeben seien die folgende Daten:

		Y		
		1	2	3
X	1	0,1	0,15	0,25
	2	0,05	0,3	0,15

Berechnen Sie

- die Randwahrscheinlichkeiten $P(X = x_i)$ und $P(Y = y_i)$
- die Erwartungswerte $E(X)$, $E(Y)$
- die Varianzen $Var(X)$, $Var(Y)$, sowie die Kovarianz $Cov(X, Y)$.
- die Korrelation $\rho = \frac{Cov(X, Y)}{\sqrt{Var(X) \cdot Var(Y)}}$
- Sind die Variablen X, Y unabhängig?

Aufgabe 7-2 Gradientenabstieg

Gegeben die Funktion $f(x) = 2 + 1/8 (x - 5)^2 + \sin(x)$.

Ermitteln sie ein lokales Minimum der Funktion im Bereich $x \in [0, 15]$ mit Hilfe des Gradientenverfahrens. Starten sie bei $x = 14$ und passen Sie die Schrittweite mit dem Faktor $\alpha = \nabla f(x)$; $\alpha = 0, 2/1/2$ an.

- Skizzieren sie die Lösung grafisch.
- Beschreiben Sie ihre Beobachtungen.
- Skizzieren Sie mögliche Problemlösungen.

Aufgabe 7-3 Spamfilter

Skizzieren Sie einen einfachen Bayes-Spamfilter. Gehen Sie davon aus, dass die Wörter der zu analysierenden Nachricht voneinander unabhängig sind.

Aufgabe 7-4 PCA

- a) Beschreiben Sie, unter welchen Umständen eine PCA sinnvoll ist bzw. welchen Zweck die PCA verfolgt.
- b) Welche möglichen negativen Auswirkungen nehmen Sie in Kauf, wenn Sie die PCA auf einen Datensatz unbekannter Struktur verwenden?

Aufgabe 7-5 PCA

Gegeben ist die folgende Designmatrix X der 6 Datenpunkte:

x	1	2	3	5	6	7
y	0	0	0	6	6	6

Führen Sie eine PCA auf den gegebenen Daten durch.

Geben Sie dabei die Eigenvektoren, Eigenwerte und die Kovarianzmatrix an und visualisieren Sie die Daten vor und nach der PCA.