

# Maschinelles Lernen

## Übungsblatt 6

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

3. Januar 2012  
Abgabe: 09. Januar bis 14.15

### Optimierungsverfahren (6 Punkte)

Sei  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  durch  $f(x) = \frac{1}{2}x^T \mathbf{A}x - x^T \mathbf{b} + c$  gegeben, wobei  $\mathbf{A}$  eine symmetrische, positiv definite  $n \times n$ -Matrix,  $\mathbf{b} \in \mathbb{R}^n$  und  $c \in \mathbb{R}$  ist.

- Sei  $\mathbf{x}^*$  das Minimum und der Start-Vektor  $\mathbf{x}_0$  so gewählt, daß  $\mathbf{x}_0 - \mathbf{x}^*$  ein Eigenvektor von  $\mathbf{A}$  ist. Was passiert dann, wenn man das *Gradient Descent* Verfahren einsetzt?
- Das *Gradient Descent* Verfahren ist allgemein langsam, aber sicher konvergent. Wann scheidert es? Wann ist es schnell konvergent?
- Zeige: Das Newton-Verfahren konvergiert in einem Schritt für jeden Start-Vektor  $\mathbf{x}_0$ .
- Wenn das Minimum einer Funktion mehrerer Variabler bestimmt werden soll, ist das *Gradient Descent* Verfahren i.A. schneller konvergent als das Newton-Verfahren? Begründen Sie!

### Mahalanobis-Distanz (4 Punkte)

Nach Amthauer (1970) erhalten Ärzte, Juristen und Pädagogen in den Untertests Analogien (AN), Figurenauswahl (FA) und Würfelaufgaben (WÜ) des (Intelligenz-Struktur-Test) die folgenden mittleren Scores:

	Ärzte	Juristen	Pädagogen
Analogien	114	111	105
Figurenauswahl	111	103	101
Würfelaufgaben	110	100	98

mit der Kovarianzmatrix (und ihrer Inversen)

$$\Sigma = \begin{pmatrix} 100 & 30 & 32 \\ 30 & 100 & 44 \\ 32 & 44 & 100 \end{pmatrix}, \quad \Sigma^{-1} = \begin{pmatrix} .0115 & -.0023 & -.0027 \\ -.0023 & .0129 & -.0049 \\ -.0027 & -.0049 & .0130 \end{pmatrix}$$

Ein Abiturient hat in diesen Untertests die Scores 108 (AN), 112 (FA) und 101 (WÜ) erzielt. Welche Berufswahl ist für ihn optimal?