Machine Learning
Exercise Sheet 6

Prof. Dr. Dr. Lars Schmidt-Thieme, Osman Akcatepe
Information Systems and Machine Learning Lab (ISMLL)
University of Hildesheim

3. Januar 2012
Deadline 09. Januar bis 14.15

Optimization (6 Points)

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ with $f(x) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{b} + c$, so that $\mathbf{A}$ is a symmetric, positive definite $n \times n$ matrix, $\mathbf{b} \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

a)

Suppose that the minimum and the start vector are chosen as $\mathbf{x}^*$, $\mathbf{x}_0$, respectively. That is $\mathbf{x}_0 - \mathbf{x}^*$ one E eigenvector of $\mathbf{A}$. What happens if one uses *gradient descent* method?

b)

Usually *gradient descent* is slow, but definitely convergent. When does it fail? When is it definitely fast convergent?

c)

Show that Newton algorithm converges in one step for every start vector $\mathbf{x}_0$.

d)

Is the gradient descent faster convergent than Newton algorithm, if the minimum of a function with multi-variables should be determined?

Mahalanobis distance (4 points)

According to Amthauer (1970) doctors, jurists and educators achieve the following mean scores in the untertests Analogies (AN), Figure Choice (FC), Cube Tasks (CT), of (Intelligence-Structure-Test):

	doctor	**jurists**	**educators**
analogies	114	111	105
figure choice	111	103	101
cube task	110	100	98

with the covariance matrix (and its inverse)

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

a high-school graduate achieved these untertest scores 108 (AN), 112 (FC) and 101 (CT). Which profession is optimal for him?