Machine Learning
Exercise Sheet 11

Prof. Dr. Dr. Lars Schmidt-Thieme, Martin Wistuba
Information Systems and Machine Learning Lab
University of Hildesheim

January 26th, 2016
Submission until February 2nd, 13.00 to wistuba@ismll.de

Exercise 21: Learning PCA with Gradient Descent (5 Points)

Principal Component Analysis (PCA) is a dimensionality reduction technique which aims at projecting a data set $X \in \mathbb{R}^{N \times M}$ via latent principal components $V \in \mathbb{R}^{K \times M}$ for $K << M$. The procedure aims at learning both the latent components and a linear combinations of the components via weights $Z \in \mathbb{R}^{N \times K}$, such that the original data is approximated via the following loss L:

$$\underset{Z,V}{\operatorname{argmin}} L = ||X - Z \cdot V||^2 = \sum_{i=1}^{N} \sum_{j=1}^{M} \left(X_{i,j} - \sum_{k=1}^{K} Z_{i,k} V_{k,j} \right)^2 \tag{1}$$

Another method to compute the PCA of a data set is through gradient descent, where the latent data Z and the principal components Z are updated via computing the full gradient over L, as shown in Algorithm 1.

Algorithm 1 Compute PCA through Gradient Descent

Require: Original Data $X \in \mathbb{R}^{N \times M}$, Number of latent dimensions K, Learning Rate η, Number of epochs E
Ensure: Low-rank data $Z \in \mathbb{R}^{N \times K}$, Principal components $V \in \mathbb{R}^{K \times M}$

for $1, \ldots, E$ **do**
  for $i = 1, \ldots, N\ j = 1, \ldots, M, k = 1, \ldots, K$ **do**
    $V_{k,j} \leftarrow V_{k,j} - \eta \frac{\partial L}{\partial V_{k,j}}$
    $Z_{i,k} \leftarrow Z_{i,k} - \eta \frac{\partial L}{\partial Z_{i,k}}$
  end for
end for
return Z, V

Derive the update rule gradients $\frac{\partial L}{\partial V_{k,j}}$, $\frac{\partial L}{\partial Z_{i,k}}$.

Exercise 22: Dimensionality Reduction with PCA (5 Points)

a) R contains the very famous Iris data set. You can access it using the variable *iris*. Create a scatter plot for two arbitrary dimensions that shows the distribution of the three different species.

b) Reduce the predictor matrix X (these are all columns but *Species*) to two dimensions. Create a scatter plot as in a) for the reduced dimension.