Algorithm for High Dimensional Principal Component Analysis

Alexander Ilin and Tapani Raiko

June 3, 2010

This document provides an algorithm that is efficient when the dimensionality of the data d is high compared to the number of principal components c needed, that is $c \ll d$.

The basic model equation in PCA is $\mathbf{y}_j \approx \mathbf{W}\mathbf{x}_j + \mathbf{m}$, where column vectors $\mathbf{y}_j$ are the data cases, $\mathbf{W}$ is the $d \times c$ matrix that maps the principal components $\mathbf{x}_j$ to the data, and $\mathbf{m}$ is the bias vector. We also use the matrix notation $\mathbf{Y} = [\mathbf{y}_1\ \mathbf{y}_2\ \dots\ \mathbf{y}_n]$ and $\mathbf{X} = [\mathbf{x}_1\ \mathbf{x}_2\ \dots\ \mathbf{x}_n]$.

The inputs of the algorithm are the data matrix $\mathbf{Y}$ and the number of components c. The outputs are the matrix $\mathbf{W}$, the principal components $\mathbf{X}$ and the bias vector $\mathbf{m}$,.

Step 1: Find and remove bias by:

$$\mathbf{m} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{y}_j$$

$$\mathbf{y}_j \leftarrow \mathbf{y}_j - \mathbf{m} \quad \forall j$$

Step 2: Initialize $\mathbf{W}$ to a random $d \times c$ matrix.

Step 3: Alternate between the updates until convergence:

$$\mathbf{X} \leftarrow (\mathbf{W}^{\mathrm{T}}\mathbf{W})^{-1}\mathbf{W}^{\mathrm{T}}\mathbf{Y}$$

$$\mathbf{W} \leftarrow \mathbf{Y}\mathbf{X}^{\mathrm{T}}(\mathbf{X}\mathbf{X}^{\mathrm{T}})^{-1}$$

Step 4: Compute eigen-decompositions of the left sides:

$$\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathrm{T}} = \mathbf{U}\mathbf{D}_x\mathbf{U}^{\mathrm{T}}$$

$$\mathbf{D}_x^{1/2}\mathbf{U}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{U}\mathbf{D}_x^{1/2} = \mathbf{V}\mathbf{D}_w\mathbf{V}^{\mathrm{T}}$$

Step 5: Postprocessing of the solution:

$$\mathbf{W} \leftarrow \mathbf{W}\mathbf{U}\mathbf{D}_x^{1/2}\mathbf{V}$$
$$\mathbf{X} \leftarrow \mathbf{V}^{\mathrm{T}}\mathbf{D}_x^{-1/2}\mathbf{U}^{\mathrm{T}}\mathbf{X}$$

This algorithm was presented in [1], please give a citation if you find this useful. The paper and provided Matlab package also includes extensions such as variational Bayesian treatment of missing values.

References

[1] A. Ilin and T. Raiko. Practical approaches to principal component analysis in the presence of missing values. *Journal of Machine Learning Research*, 2010. To appear.