

Lorum Ipsum

Group 7

Henrik Karlsson `henrik10@kth.se`
Alexios Kotsakis `alexiosk@kth.se`
Markus Videll `mvidell@kth.se`
Wenyi Zhao `wenyizh@kth.se`

January 13, 2017

Begin Written report instructions. The article, the re-implementation, and your results are presented in a written report, to be sent to your supervisor via email in pdf format before the oral presentation on January 16. The report should be at most 7 pages, with 12 pt font and about 2.5 cm margins, including references, images, and tables. In addition to the 7 pages, the report should have a cover page with title, group number, author list and (optional) abstract. The report should be written in (to a reasonable level) grammatically correct English.

In the report you should first describe the article on such a level of detail that your peer students in this course understand the method, and so that it is clear to the reader that you understand the method too.

You should then present your re-implementation of the method, and your reproduction of the results, again on such a level that your peer students understand what you have done, and so that it is clear to the reader what results you got and if, how, and why they deviate from the results presented in the original article.

Finally you should argue for and against the method, possibly suggesting improvements.

All statements made in the report (e.g., "method X is better than method Y") should be supported by either a reference to the original paper or report where the statement was made, or if the statement originates from you, you should explain why this statement is true.

A technically correct, well organized report with good language and a clear line of argument will receive a high grade. Missed hand-in deadline, violations of the length and formatting requirements, as well as statements not supported by references will have a heavy negative effect on the grade.

All group members must participate actively in the writing of the report. By adding a group member to the author list of the report, you certify that this person has written at least one section of the report.

1 Background

Principal component analysis (PCA) is a statistical procedure which aims to convert a set of possibly correlated variables into a set of linearly uncorrelated variables. The PCA cannot detect nonlinear structures in a given dataset, however, the Kernel PCA is well suited to extract nonlinear structures within the data.

The report tries to use kernel PCA to de-noise data that have been exposed to different kinds of noise. This is done by mapping the data onto a high dimensional space \mathbf{F} using a non-linear function Φ .

To de-noise a data point x it is first mapped onto the high dimension space \mathbf{F} using using the function Φ . It is then projected using the high variance eigenvectors in the image of the training data. We then want to find the pre-image z so $\Phi(z)$ is equal to the projected x . Since the eigenvectors with small variance has been removed z should be noise free.

2 Kernel PCA

The Kernel PCA can be seen as a generalization of the PCA that maps the data into some feature space \mathbf{F} via a (usually nonlinear) function

$$\Phi : R^N \rightarrow \mathbf{F}$$

and then performs a PCA on the mapped data.

To perform a PCA in feature space, we need to find eigenvalues $\lambda > 0$ and eigenvectors $\mathbf{V} \in \mathbf{F} \setminus \{\mathbf{0}\}$

$$\lambda \mathbf{V} = C\mathbf{V} \tag{1}$$

where C is the covariance matrix in feature space given by

$$C = \frac{1}{N} \sum_{n=1}^N \Phi(\mathbf{x}_n) \Phi(\mathbf{x}_n)^T \tag{2}$$

assuming the projected data has a zero mean in feature space.

As $C\mathbf{V} = \frac{1}{N} \sum_{n=1}^N (\Phi(\mathbf{x}_n) \cdot \mathbf{V}) \Phi(\mathbf{x}_n)$, all solutions \mathbf{V} must lie in the span of $\Phi(x_1), \dots, \Phi(x_N)$, hence (2) is equivalent to

$$\lambda(\Phi(\mathbf{x}_k) \cdot \mathbf{V}) = (\Phi(\mathbf{x}_k) \cdot C\mathbf{V}) \tag{3}$$

and there exists coefficients α_n ($n = 1, \dots, N$) such that

$$\mathbf{V} = \sum_{n=1}^N \alpha_n \Phi(\mathbf{x}_n) \tag{4}$$

Combining (3) and (4), we get

$$\begin{aligned} \Phi(\mathbf{x}_k) \cdot C\mathbf{V} &= \lambda \Phi(\mathbf{x}_k) \cdot \mathbf{V} \implies \\ [\Phi(\mathbf{x}_k) \cdot \frac{1}{N} \sum_{n=1}^N \Phi(\mathbf{x}_n)] [\Phi(\mathbf{x}_n) \cdot \sum_{m=1}^N \alpha_m \Phi(\mathbf{x}_m)] &= \lambda \Phi(\mathbf{x}_k) \cdot \sum_{n=1}^N \alpha_n \Phi(\mathbf{x}_n) \implies \\ \frac{1}{N} \sum_{n=1}^N \sum_{m=1}^N \alpha_m [\Phi(\mathbf{x}_k) \cdot \Phi(\mathbf{x}_n)] [\Phi(\mathbf{x}_n) \cdot \Phi(\mathbf{x}_m)] &= \lambda \sum_{n=1}^N \alpha_n [\Phi(\mathbf{x}_k) \cdot \Phi(\mathbf{x}_n)] \end{aligned}$$

Now we define the matrix K by $K_{nm} = [\Phi(\mathbf{x}_n) \cdot \Phi(\mathbf{x}_m)]$ to arrive at,

$$K^2\alpha = N\lambda K\alpha \implies K\alpha = \lambda'\alpha \quad \text{with} \quad \lambda' = N\lambda \quad (5)$$

where α is the column vector with elements $\{\alpha_1, \dots, \alpha_N\}$.

3 Results

4 Discussion