TECHNISCHE UNIVERSITÄT MÜNCHEN
Fakultät für Elektrotechnik und Informationstechnik
Lehrstuhl für Datenverarbeitung
PD Martin Kleinsteuber

Information Retrieval in High Dimensional Data

Tutorial #7, 27.6.2018

Kernel PCA

Task 1. Let $\phi : \mathbb{R}^p \to \mathbb{R}^q$ be a function which maps a vector from the observation space to a feature space. When applied to matrices, let us assume it operates column-wise. The respective kernel shall be defined as

$$\begin{aligned} \kappa : \mathbb{R}^p \times \mathbb{R}^p &\to \mathbb{R} \\ (\mathbf{x}, \mathbf{y}) &\mapsto \phi(\mathbf{x})^\top \phi(\mathbf{y}). \end{aligned}$$

a) Let $\mathbf{X} \in \mathbb{R}^{p \times N}$ be a training data matrix. Give an expression for centering $\phi(\mathbf{X})$. Show that it can be written as $\phi(\mathbf{X})\mathbf{H}$, where $\mathbf{H}$ is a square matrix.

Solution:

$$\phi(\mathbf{X}) - \frac{1}{N}\phi(\mathbf{X})\mathbf{1}_N\mathbf{1}_N^\top = \phi(\mathbf{X})(\mathbf{I} - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^\top).$$

Thus,

$$\mathbf{H} = \mathbf{I} - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^\top.$$

b) Define $\mathbf{K} = \phi(\mathbf{X})^\top \phi(\mathbf{X})$ and $\tilde{\mathbf{K}} = \mathbf{HKH} = (\phi(\mathbf{X})\mathbf{H})^\top (\phi(\mathbf{X})\mathbf{H})$ with the sorted EVD $\tilde{\mathbf{K}} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top$. Let $\mathbf{V}_k \in \mathbb{R}^{N,k}$ denote the first k columns of $\mathbf{V}$. Express the leading k left singular vectors $\mathbf{U}_k \in \mathbb{R}^{p,k}$ of $\phi(\mathbf{X})\mathbf{H}$ in terms of the previously defined matrices.

Solution:

$$\begin{aligned} \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{V}^\top &= \phi(\mathbf{X})\mathbf{H} \\ \Leftrightarrow \mathbf{U}\mathbf{\Lambda}_k^{1/2} &= \phi(\mathbf{X})\mathbf{H}\mathbf{V}_k \\ \Leftrightarrow \mathbf{U}_k &= \phi(\mathbf{X})\mathbf{H}\mathbf{V}_k\mathbf{\Lambda}_k^{-1/2}. \end{aligned}$$

c) Let $\mathbf{Y} \in \mathbb{R}^{p \times N_{\text{test}}}$ be a test data matrix. Express its Kernel PCA scores $\mathbf{S}_k$ by using the result from b) and describe how you can compute them, given that κ is known, but ϕ is not.

Solution: The feature matrix $\phi(\mathbf{Y})$ needs to be centered:

$$\tilde{\phi}(\mathbf{Y}) = \phi(\mathbf{Y}) - \frac{1}{N}\phi(\mathbf{X})\mathbf{1}_N\mathbf{1}_{N_{\text{test}}}^\top.$$

The scores are given by

$$\begin{aligned}
\mathbf{U}_k^\top \phi(\mathbf{Y}) &= \mathbf{\Lambda}_k^{-1/2}\mathbf{V}_k^\top \mathbf{H}\phi(\mathbf{X})^\top \tilde{\phi}(\mathbf{Y}) \\
&= \mathbf{\Lambda}_k^{-1/2}\mathbf{V}_k^\top \mathbf{H}\phi(\mathbf{X})^\top (\phi(\mathbf{Y}) - \frac{1}{N}\phi(\mathbf{X})\mathbf{1}_N\mathbf{1}_{N_{\text{test}}}^\top) \\
&= \mathbf{\Lambda}_k^{-1/2}\mathbf{V}_k^\top \mathbf{H}(\kappa(\mathbf{X},\mathbf{Y}) - \frac{1}{N}\mathbf{K}\mathbf{1}_N\mathbf{1}_{N_{\text{test}}}^\top).
\end{aligned}$$

If we want to compute the score of our original training data, i.e. $\mathbf{X} = \mathbf{Y}$, then the expression simplifies to

$$\begin{aligned}
\mathbf{\Lambda}_k^{-1/2}\mathbf{V}_k^\top \mathbf{H}(\kappa(\mathbf{X},\mathbf{X}) - \frac{1}{N}\mathbf{K}\mathbf{1}_N\mathbf{1}_N^\top.) &= \mathbf{\Lambda}_k^{-1/2}\mathbf{V}_k^\top \mathbf{HKH} \\
&= \mathbf{\Lambda}_k^{1/2}\mathbf{V}_k^\top.
\end{aligned}$$