Kernel-Based Learning & Multivariate Modeling
MIRI Master

Lluís A. Belanche
belanche@cs.upc.edu

October 22, 2019

Problem 1 Variance of the projection (1)

Prove that, for centered data, the variance of the PCA projection along a direction $\boldsymbol{v}$ is given by $\boldsymbol{v}^T C \boldsymbol{v}$, where C is the sample covariance matrix of the data.

. . . .

Problem 2 Variance of the projection (2)

Prove that, for centered data, the variance of the PCA projection onto the first PC $\boldsymbol{v}_1$ is given by the largest eigenvalue $\lambda_{(1)}$ of the sample covariance matrix of the data.

. . . .

Problem 3 Kernelising PCA

Prove or show that the kernel PCA equation

$$\lambda \sum_{m=1}^{N} \alpha_m k(\boldsymbol{x}_m, \boldsymbol{x}_k) = \frac{1}{N} \sum_{m=1}^{N} \alpha_m \sum_{n=1}^{N} k(\boldsymbol{x}_n, \boldsymbol{x}_k) k(\boldsymbol{x}_n, \boldsymbol{x}_m), \qquad 1 \leq k \leq N$$

is rewritten as:

$$\lambda \mathbf{K} \boldsymbol{\alpha} = \frac{1}{N} \mathbf{K}^2 \boldsymbol{\alpha}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)^T$ and $\mathbf{K} = (k_{nm})$, where $k_{nm} = k(\boldsymbol{x}_n, \boldsymbol{x}_m)$. Hint: recall that the kernel is a symmetric function.

. . . .

Problem 4 Centering in feature space

Some algorithms –like PCA– need a centering procedure in the data space. When we work in input space a data set of d-dimensional vectors $X = \{\boldsymbol{x}_n\}$ for $n = 1, ..., N$ can be centered (around the origin) as $\boldsymbol{x}_n := \boldsymbol{x}_n - \frac{1}{N}\sum_{m=1}^{N} \boldsymbol{x}_m$. In feature space we would need something like

$$\phi(\boldsymbol{x}_n) := \phi(\boldsymbol{x}_n) - \frac{1}{N}\sum_{m=1}^{N} \phi(\boldsymbol{x}_m).$$

Suppose we are given a data set of objects $X = \{\boldsymbol{x}_n\}$ and we choose a kernel function k. We compute the kernel matrix of the data as $\mathbf{K} = (k_{nm})$, where $k_{nm} = k(\boldsymbol{x}_n, \boldsymbol{x}_m)$. Prove that the centered kernel matrix is

$$\mathbf{K} := \mathbf{K} - \frac{1}{N}\mathbf{1K} - \frac{1}{N}\mathbf{K1} + \frac{1}{N^2}\mathbf{1K1}$$

where $\mathbf{1}$ is a $N \times N$ matrix of ones.

. . . .

Problem 5 A simple novelty detector

The novelty problem goes as follows: we have a set of i.i.d. data objects $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$. For a new object $\boldsymbol{x}^*$, we want to classify it as "novel" or "known". A simple algorithm tackles this task by classifying $\boldsymbol{x}^*$ as "novel" if it appears farther to the centroid of the empirical data than any other known data point. More formally, given a feature map ϕ, define:

$$\bar{\phi} = \frac{1}{N}\sum_{n=1}^{N} \phi(\boldsymbol{x}_n)$$

to be centroid of the mapped objects. The (squared) distance of any object to this centroid is $d(\boldsymbol{x}) = \|\phi(\boldsymbol{x}) - \bar{\phi}\|^2$. Therefore our method will classify $\boldsymbol{x}^*$ as "novel" if

$$d(\boldsymbol{x}^*) > \max_{1 \leq n \leq N} d(\boldsymbol{x}_n)$$

1. Derive a kernelized version of the method. Try to simplify it in terms of the number of kernel evaluations
2. Suppose we are now interested in an on-line version that accepts a never-ending sequence of i.i.d. data objects $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$. Explain how to obtain a simple incremental method

. . . .