Machine Learning 2 – Group ESHG
Assignment 06

Willi Gierke, Arik Elimelech, Mehmed Halilovic, Leon Sixt

June 6, 2017

Exercise 1: Kernel Eigenvectors (30 P)

Consider a dataset $x_1, \dots, x_N \in \mathcal{X}$ composed of N data points drawn i.i.d. from some probability distribution p. Let $\mathrm{k} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ define a kernel, to which we can associate a feature map $\phi : \mathcal{X} \to \mathbb{R}^d$ such that:

$$\forall x_i, x_j \in \mathcal{X} : \; k(x_i, x_j) = \phi(x_i)^\top \phi(x_j).$$

Let K be the Gram matrix of size $N \times N$ containing the kernel scores for all pairs of data points, and Φ be a matrix of size $N \times d$ containing feature representations of all N data points. In the feature space, we consider an analysis that consists of finding the direction $\boldsymbol{v} \in \mathbb{R}^d$ solution to the optimization problem:

$$\max_{\boldsymbol{v}} \boldsymbol{v}^\top C \boldsymbol{v} \qquad \text{such that} \quad \|\boldsymbol{v}\|^2 = 1$$

where

$$C = \sum_{i=1}^{N} \phi(x_i)\phi(x_i)^\top$$

is a noncentered scatter matrix of size $d \times d$.

(a) Show that this maximization problem is equivalent to solving an eigenvalue equation

$$C\boldsymbol{v} = \lambda \boldsymbol{v}$$

for the largest eigenvalue λ.

(b) Show that when representing the direction $\boldsymbol{v}$ in terms of feature vectors as $\boldsymbol{v} = \Phi^\top \boldsymbol{\alpha}$, the optimization problem can be rewritten as an eigenvalue equation

$$K\boldsymbol{\alpha} = \lambda \boldsymbol{\alpha},$$

where $\boldsymbol{\alpha} \in \mathbb{R}^N$ becomes the new unknown.

Exercise 2: RDE Coefficients (20 P)

Given a kernel Gram matrix K of size $N \times N$ built from a kernel function $k(x, x')$ and a dataset $x_1, \dots, x_N \in \mathcal{X}$, we can compute the kernel eigenvectors from the multiple solutions $(\lambda_i, u_i)_{i=1}^N$ of the eigenvalue equation:

$$K u_i = \lambda_i u_i$$

where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N \geq 0$ are the sorted eigenvalues, and u_i is the eigenvector associated to λ_i. Considering for each data point x_i a label $y_i \in \mathbb{R}$, and building a vector of labels $\boldsymbol{y} = (y_i)_{i=1}^N \in \mathbb{R}^N$, we can define the RDE coefficients

$$\forall_{i=1}^N : \; s_i = u_i^\top \boldsymbol{y}$$

that are used as a basis for RDE analysis.

(a) Show that $\sum_{i=1}^N s_i^2 = \|\boldsymbol{y}\|^2$. (This technical result is useful to verify in practice that the computation of RDE coefficients has been correctly implemented.)

Exercise 3: Programming (50 P)

Download the programming files on ISIS and follow the instructions.

1

a

As in slides from last semester. We just have to replace C with S and v with w.

$$\max_{w} \; w^{\mathrm{T}} S w$$
$$\text{s.t.} \; ||w|| = 1$$

1. Define Lagrangian: $\mathcal{L}(w, \lambda) = w^{\mathrm{T}} S w + \lambda(1 - w^{\mathrm{T}} w)$

2. Compute gradient: $\dfrac{\partial \mathcal{L}(w, \lambda)}{\partial w} = 2Sw - 2\lambda w$

3. Set to zero: $Sw = \lambda w$

This is an Eigenvalue problem!

b

$$C = \sum_{i=1}^{N} \phi(x_i)\phi(x_i)^T$$
$$K_{ij} = \phi(x_i)^T \cdot \phi(x_j)$$
$$PCA : Cv = \lambda v$$

Show: $K\alpha = \lambda\alpha$, where $v = \phi^T \alpha$, $\Phi = [\phi(x_i) ... \phi(x_n)]$

$$Cv = \lambda v$$
$$\left(\sum_{i=1}^{N} \phi(x_i)\phi(x_i)^T\right) \cdot \Phi^T \alpha = \lambda \cdot \Phi^T \alpha$$
$$(\Phi\Phi^T)^{(-1)}\left(\sum_{i=1}^{N} \phi(x_i)\phi(x_i)^T\right) \cdot \Phi^T \alpha = \lambda\alpha$$

2

$\forall_{i=1}^{N} : s_i = u_i^T y$

Show: $\sum_{i=1}^{N} s_i^2 = ||y||^2$
Let $U = [u_1 \dots u_N]$, then we can rewrite the above condition with

$$s = U^T y \tag{1}$$

where $s = [s_i \dots s_N]^T$.
Now we can rearrange $\sum_{i=1}^{N} s_i^2$ esaily to give $||y||^2$.

$$\begin{aligned}\sum_{i=1}^{N} s_i^2 &= s^T s = (U^T y)^T (U^T y) \\ &= y^T U U^T y = y^T y = ||y||^2\end{aligned}$$

As we know that U is an orthogonal matrix (because it columns are eigenvectors of K), $UU^T = I$ where I is the identity matrix.