

Exercise Sheet 10

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 $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ define a kernel, to which we can associate a feature map $\phi : \mathcal{X} \rightarrow \mathbb{R}^d$ 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 $\mathbf{v} \in \mathbb{R}^d$ solution to the optimization problem:

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

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

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

for the largest eigenvalue λ .

- (b) Show that when representing the direction \mathbf{v} in terms of feature vectors as $\mathbf{v} = \Phi^\top \boldsymbol{\alpha}$, 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 $\mathbf{y} = (y_i)_{i=1}^N \in \mathbb{R}^N$, we can define the RDE coefficients

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

that are used as a basis for RDE analysis.

- (a) Show that $\sum_{i=1}^N s_i^2 = \|\mathbf{y}\|^2$. (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.