Exercises for the course

Machine Learning 2

Summer semester 2024

Fachgebiet Maschinelles Lernen
Institut für Softwaretechnik und theoretische Informatik
Fakultät IV, Technische Universität Berlin
Prof. Dr. Klaus-Robert Müller
Email: klaus-robert.mueller@tu-berlin.de

Exercise Sheet 1

Exercise 1: Symmetries in LLE (30 P)

The Locally Linear Embedding (LLE) method takes as input a collection of data points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_N \in \mathbb{R}^d$ and embeds them in some low-dimensional space. LLE operates in two steps, with the first step consisting of minimizing the objective

$$\mathcal{E}(W) = \sum_{i=1}^{N} \Big\| \boldsymbol{x}_i - \sum_j W_{ij} \boldsymbol{x}_j \Big\|^2$$

where W is a collection of reconstruction weights subject to the constraint $\forall i : \sum_j W_{ij} = 1$, and where $\sum_j$ sums over the K nearest neighbors of the data point $\boldsymbol{x}_i$. The solution that minimizes the LLE objective can be shown to be invariant to various transformations of the data.

Show that invariance holds in particular for the following transformations:

(a) Replacement of all $\boldsymbol{x}_i$ with $\alpha \boldsymbol{x}_i$, for an $\alpha \in \mathbb{R}^+ \setminus \{0\}$,

(b) Replacement of all $\boldsymbol{x}_i$ with $\boldsymbol{x}_i + \boldsymbol{v}$, for a vector $\boldsymbol{v} \in \mathbb{R}^d$,

(c) Replacement of all $\boldsymbol{x}_i$ with $U\boldsymbol{x}_i$, where U is an orthogonal $d \times d$ matrix.

Exercise 2: Closed form for LLE (30 P)

In the following, we would like to show that the optimal weights W have an explicit analytic solution. For this, we first observe that the objective function can be decomposed as a sum of as many subobjectives as there are data points:

$$\mathcal{E}(W) = \sum_{i=1}^{N} \mathcal{E}_i(W) \qquad \text{with} \quad \mathcal{E}_i(W) = \Big\| \boldsymbol{x}_i - \sum_j W_{ij} \boldsymbol{x}_j \Big\|^2$$

Furthermore, because each subobjective depends on different parameters, they can be optimized independently. We consider one such subobjective and for simplicity of notation, we rewrite it as:

$$\mathcal{E}_i(\boldsymbol{w}) = \Big\| \boldsymbol{x} - \sum_{j=1}^{K} w_j \boldsymbol{\eta}_j \Big\|^2$$

where $\boldsymbol{x}$ is the current data point (we have dropped the index i), where $\eta = (\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_K)$ is a matrix of size $K \times d$ containing the K nearest neighbors of $\boldsymbol{x}$, and $\boldsymbol{w}$ is the vector of size K containing the weights to optimize and subject to the constraint $\sum_{j=1}^{K} w_j = 1$.

(a) *Prove* that the optimal weights for $\boldsymbol{x}$ are found by solving the following optimization problem:

$$\min_{\boldsymbol{w}} \quad \boldsymbol{w}^\top C \boldsymbol{w} \qquad \text{subject to} \quad \boldsymbol{w}^\top \mathbb{1} = 1.$$

where $C = (\mathbb{1}\boldsymbol{x}^\top - \eta)(\mathbb{1}\boldsymbol{x}^\top - \eta)^\top$ is the covariance matrix associated to the data point $\boldsymbol{x}$ and $\mathbb{1}$ is a vector of ones of size K.

(b) *Show* using the method of Lagrange multipliers that the minimum of the optimization problem found in (a) is given analytically as:

$$w = \frac{C^{-1}\mathbb{1}}{\mathbb{1}^\top C^{-1} \mathbb{1}}.$$

(c) *Show* that the optimal $\boldsymbol{w}$ can be equivalently found by solving the equation $C\boldsymbol{w} = \mathbb{1}$ and then rescaling $\boldsymbol{w}$ such that $\boldsymbol{w}^\top \mathbb{1} = 1$.

Exercise 3: Programming (40 P)

Download the programming files on ISIS and follow the instructions.