Exercises for the course

Machine Learning 2

Summer semester 2018

Abteilung 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 2

A good method to consolidate one's understanding of the LLE algorithm is reading the corresponding scientific publication. However, as in all scientific papers, there are details which are not worked out explicitly, since they are deemed "easy", "widely known", "folklore", "trivial" or "too large to fit on the margin". While this may or may not be true (also depending on the reader's viewpoint), working out these details for oneself is often necessary for understanding the paper's contents in more than a superficial way. This exercise will guide you in doing so for a LLE paper. In the following, we will always refer to the 2001 paper *An Introduction to Locally Linear Embedding* by Lawrence K. Saul and Sam T. Roweis, 2001, which is linked via ISIS.

Exercise 1: Symmetries (30 P)

In the third paragraph of page 3, it is claimed that the optimal weights W_{ij} which minimize the cost function $\mathcal{E}$ are independent with respect to scaling, translation and rotation of the original data $\vec{X}_i$. Prove that. That is, *prove* that the minimum (or minima) of $\mathcal{E}$ is (or are) invariant under the following symmetries:

(i) Replace all $\vec{X}_i$ with $\alpha\vec{X}_i$, for an $\alpha \in \mathbb{R}^+ \setminus \{0\}$,

(ii) Replace all $\vec{X}_i$ with $\vec{X}_i + \vec{v}$, for a vector $\vec{v} \in \mathbb{R}^D$,

(iii) Replace all $\vec{X}_i$ with $U \cdot \vec{X}_i$, where U is an orthogonal $D \times D$ matrix (this additionally includes mirror symmetries).

Briefly explain why you obtain a statement for all possible rotations (e.g. rotating around arbitrary fixed points) by proving statements (i)-(iii), despite the fact that multiplying with U as in (iii) always leaves the origin (i.e., the zero vector) fixed (and only the origin in general).

Exercise 2: Lagrange Multipliers (30 P)

In the second paragraph on page 3, it is stated that finding the optimal W_{ij} is a least-squares problem, which is shown to have an explicit analytic solution in Appendix A. In the following, assume the notation of Appendix A. For abbreviation (and clarity of notation), additionally write $w = (w_1, \dots, w_K)^\top$ for the weight vector which is optimized, $\eta = (\vec{\eta}_1, \dots, \vec{\eta}_K)^\top$ for the $(K \times D)$-matrix of nearest neighbors of $\vec{x}$, $\mathbb{1} = (1, \dots, 1)^\top$ for the K-dimensional vector of ones, and $C = (\mathbb{1}\vec{x}^\top - \eta)(\mathbb{1}\vec{x}^\top - \eta)^\top$ for the local covariance matrix at $\vec{x}$. We would like to work out the following claims from Appendix A:

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

$$\begin{aligned} \min_w \quad & w^\top C w \\ \text{subject to} \quad & w^\top \mathbb{1} = 1. \end{aligned}$$

In particular, prove equation (3) on page 9.

(ii) *Show* by using the Lagrangian method for constrained optimization that the minimum of the optimization problem is explicitly given by

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

(iii) *Show* that the minimum w can be equivalently found by solving the equation

$$Cw = \mathbb{1},$$

and then rescaling w such that $w^\top \mathbb{1} = 1$.

Exercise 3: Programming (40 P)

Download the programming files on ISIS and follow the instructions.