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 10

Exercise 1: One-Class SVM ($\mathbf{5+5+20+10+10}$ P)

The one-class SVM is given by the minimization problem:

$$\min_{\boldsymbol{w},\rho,\xi} \quad \frac{1}{2}\|\boldsymbol{w}\|^2 - \rho + \frac{1}{N\nu}\sum_{i=1}^{N}\xi_i$$

$$\text{s.t.} \quad \forall_{i=1}^{N} : \langle \phi(\boldsymbol{x}_i), \boldsymbol{w}\rangle \geq \rho - \xi_i \quad \text{and} \quad \xi_i \geq 0$$

where $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are the training data and $\phi(\boldsymbol{x}_i) \in \mathbb{R}^d$ is a feature space representation.

(a) *Show* that strong duality holds (i.e. verify the Slater's conditions).

(b) *Write* the Lagrange function associated to this optimization problem.

(c) *Show* the dual program for the one-class SVM is given by:

$$\max_{\boldsymbol{\alpha}} \quad -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j k(\boldsymbol{x}_i, \boldsymbol{x}_j)$$

$$\text{s.t.} \quad \sum_{i=1}^{N}\alpha_i = 1 \quad \text{and} \quad \forall_{i=1}^{N} : \; 0 \leq \alpha_i \leq \frac{1}{N\nu}$$

(d) *Show* that the problem can be equivalently rewritten in canonical matrix form as:

$$\min_{\boldsymbol{\alpha}} \quad \frac{1}{2}\boldsymbol{\alpha}^\top K \boldsymbol{\alpha}$$

$$\text{s.t.} \quad \mathbf{1}^\top \boldsymbol{\alpha} = 1 \quad \text{and} \quad \begin{pmatrix} -I \\ I \end{pmatrix} \boldsymbol{\alpha} \preceq \begin{pmatrix} \mathbf{0} \\ \mathbf{1}/N\nu \end{pmatrix}$$

where K is the Gram matrix whose elements are defined as $K_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$.

(e) The decision rule in the primal for classifying a point as an outlier is given by:

$$\langle \phi(\boldsymbol{x}), \boldsymbol{w}\rangle < \rho$$

Also, one can verify that for any data point $\boldsymbol{x}_i$ whose associated dual variable satisfies the strict inequalities $0 < \alpha_i < \frac{1}{N\nu}$, and calling one such point a support vector $\boldsymbol{x}_{\mathrm{SV}}$, the following equality holds:

$$\langle \phi(\boldsymbol{x}_{\mathrm{SV}}), \boldsymbol{w}\rangle = \rho$$

Show that the outlier detection rule can be expressed as:

$$\sum_{i=1}^{N}\alpha_i k(\boldsymbol{x}, \boldsymbol{x}_i) < \sum_{i=1}^{N}\alpha_i k(\boldsymbol{x}_{\mathrm{SV}}, \boldsymbol{x}_i)$$

Exercise 2: Programming (50 P)

Download the programming files on ISIS and follow the instructions.