Exercises for the course

Machine Learning 2

Summer semester 2021

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 8

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 $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ 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.

Exercise Sheet 8

1 One-Class SVM

a) $- \xi_i \geq 0$ if we set $\xi_i \geq 0$

$- \langle \phi(x_i), w \rangle \geq \rho - \xi_i$ can be achieved by setting ξ_i very large

b) $\mathcal{L}(w, \rho, \xi, \lambda_1, \lambda_2) = \frac{1}{2}\|w\|^2 - \rho + \frac{1}{N\nu}\sum_{i=1}^{N} \xi_i$

$+ \sum_i \lambda_{1_i} \underbrace{(\rho - \xi_i - \langle \phi(x_i), w \rangle)}_{\leq 0} + \sum_i \lambda_{2_i} \underbrace{(-\xi_i)}_{\leq 0}$

c) $\max_{\alpha \geq 0, \beta \geq 0} \min_{w, \rho, \xi} \mathcal{L}(\cdot)$ $\quad \lambda_1 \to \alpha,\ \lambda_2 \to \beta$

$\frac{\partial \mathcal{L}(\cdot)}{\partial w} = w - \sum_i \alpha_i \phi(x_i) \overset{!}{=} 0 \iff w = \sum_i \alpha_i \phi(x_i)$

$\frac{\partial \mathcal{L}(\cdot)}{\partial \rho} = -1 + \sum_i \alpha_i \phi(x_i) \overset{!}{=} 0 \iff \sum_i \alpha_i \overset{!}{=} 1$

$\frac{\partial \mathcal{L}(\cdot)}{\partial \xi_i} = \frac{1}{N\nu} - \alpha_i - \beta_i = 0$

$\max_{\alpha \geq 0, \cancel{\beta \geq 0}} \frac{1}{2} \|\sum_i \alpha_i \phi(x_i)\|^2 - \sum_i \alpha_i \langle \phi(x_i), \sum_j \alpha_j \phi(x_j) \rangle$

$= \max_{\alpha \geq 0} -\frac{1}{2} \sum_{ij} \alpha_i \alpha_j \underbrace{\phi(x_i)^\top \phi(x_j)}_{k(x_i, x_j)}$

$\sum_i \alpha_i = 1$, $\alpha_i \geq 0$, $\frac{1}{N\nu} - \alpha_i - \beta_i = 0 \iff \frac{1}{N\nu} \geq \alpha_i$

$\rightarrow 0 \leq \alpha_i \leq \frac{1}{N\nu}$

d) $\max_{\alpha} -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j k(x_i, x_j) \rightarrow \min_{\alpha} \underbrace{\frac{1}{2}\sum_{ij} \alpha_i \alpha_j \underbrace{k(x_i, x_j)}_{K_{ij}}}_{\alpha^T K \alpha}$

$\sum_{i=1}^{N} \alpha_n = 1 \rightarrow 1^T \alpha = 1$

$0 \leq \alpha_i \leq \frac{1}{N\nu} \rightarrow \begin{pmatrix} -I \\ I \end{pmatrix} \alpha \leq \begin{pmatrix} 0 \\ \frac{1}{N\nu} \end{pmatrix}$

e) $\langle \phi(x), w \rangle < \langle \phi(x_{sv}), w \rangle$

$w = \sum_i \alpha_i \phi(x_i)$

$\langle \phi(x), \sum_i \alpha_i \phi(x_i) \rangle < \langle \phi(x_{sv}), \sum_i \alpha_i \phi(x_i) \rangle$

$= \sum_i \alpha \underbrace{\langle \phi(x), \phi(x_i) \rangle}_{k(x, x_i)} < \sum_i \alpha_i \underbrace{\langle \phi(x_{sv}), \phi(x_i) \rangle}_{k(x_{sv}, x_i)}$

$= \sum_i \alpha_i k(x, x_i) < \sum_i \alpha_i k(x_{sv}, x_i)$