Exercises for the course

Machine Learning 1

Winter semester 2024/25

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 6

Exercise 1: Dual formulation of the Soft-Margin SVM $(5+20+10+5$ P)

The primal program for the linear soft-margin SVM is

$$\min_{\boldsymbol{w},b,\boldsymbol{\xi}} \ \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N}\xi_i$$

subject to

$$\forall_{i=1}^{N}: \ y_i \cdot (\boldsymbol{w}^\top \phi(\boldsymbol{x}_i) + b) \geq 1 - \xi_i \quad \text{and} \quad \xi_i \geq 0$$

where $\|.\|$ denotes the Euclidean norm, ϕ is a feature map, $\boldsymbol{w} \in \mathbb{R}^d, b \in \mathbb{R}$ are the parameter to optimize, and $\boldsymbol{x}_i \in \mathbb{R}^d, y_i \in \{-1, 1\}$ are the labeled data points regarded as fixed constants. Once the hard-margin SVM has been learned, prediction for any data point $\boldsymbol{x} \in \mathbb{R}^d$ is given by the function

$$f(\boldsymbol{x}) = \text{sign}(\boldsymbol{w}^\top \phi(\boldsymbol{x}) + b).$$

(a) *State* the conditions on the data under which a solution to this program can be found from the Lagrange dual formulation *(Hint: verify the Slater's conditions)*.

(b) *Derive* the Lagrange dual and show that it reduces to a constrained quadratic optimization problem. State both the objective function and the constraints of this optimization problem.

(c) *Describe* how the solution $(\boldsymbol{w}, b)$ of the primal program can be obtained from a solution of the dual program.

(d) *Write* a kernelized version of the dual program and of the learned decision function.

Exercise 2: SVMs and Quadratic Programming (10 P)

We consider the `CVXOPT` Python software for convex optimization. The method `cvxopt.solvers.qp` solves quadratic optimization problems given in the matrix form:

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \frac{1}{2}\boldsymbol{x}^\top P\boldsymbol{x} + \boldsymbol{q}^\top \boldsymbol{x} \\ \text{subject to} \quad & G\boldsymbol{x} \preceq \boldsymbol{h} \\ \text{and} \quad & A\boldsymbol{x} = \boldsymbol{b}. \end{aligned}$$

Here, $\preceq$ denotes the element-wise inequality: $(\boldsymbol{h} \preceq \boldsymbol{h}') \Leftrightarrow (\forall_i : h_i \leq h_i')$. Note that the meaning of the variables $\boldsymbol{x}$ and $\boldsymbol{b}$ is different from that of the same variables in the previous exercise.

(a) *Express* the matrices and vectors $P, \boldsymbol{q}, G, \boldsymbol{h}, A, \boldsymbol{b}$ in terms of the variables of Exercise 1, such that this quadratic minimization problem corresponds to the kernel dual SVM derived above.

Exercise 3: Programming (50 P)

Download the programming files on ISIS and follow the instructions.

Exercise 1: Dual formulation of the Soft-Margin SVM (5 + 20 + 10 + 5 P)

The primal program for the linear soft-margin SVM is

$$\min_{\boldsymbol{w}, b, \boldsymbol{\xi}} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N} \xi_i$$

subject to

$$\forall_{i=1}^{N}: \; y_i \cdot (\boldsymbol{w}^\top \phi(\boldsymbol{x}_i) + b) \geq 1 - \xi_i \quad \text{and} \quad \xi_i \geq 0$$

where $\|.\|$ denotes the Euclidean norm, ϕ is a feature map, $\boldsymbol{w} \in \mathbb{R}^d, b \in \mathbb{R}$ are the parameter to optimize, and $\boldsymbol{x}_i \in \mathbb{R}^d, y_i \in \{-1, 1\}$ are the labeled data points regarded as fixed constants. Once the hard-margin SVM has been learned, prediction for any data point $\boldsymbol{x} \in \mathbb{R}^d$ is given by the function

$$f(\boldsymbol{x}) = \text{sign}(\boldsymbol{w}^\top \phi(\boldsymbol{x}) + b).$$

(a) *State* the conditions on the data under which a solution to this program can be found from the Lagrange dual formulation *(Hint: verify the Slater's conditions)*.

Solution: ???

To verify the Slater's conditions, for any w and ρ, we can always increase ξ_i until constraints become strict inequalities.

$$\forall_{i=1}^{N}: y_i(w^T\phi(x_i)+b) > 1 - \xi_i \quad \text{and} \quad \xi_i > 0$$

So we can always find a feasible solution for the primal problem.

always increase ξ_i until constraints are satifisfied with strict inequalities

Then the dual problem can be written as:

$$\min_{\alpha} \frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j y_i y_j k(x_i, x_j) - \sum_i \alpha_i$$

$$\text{s.t.} \quad 0 \leq \alpha_i \leq C, \; \forall i$$

$$\sum_i \alpha_i y_i = 0$$

so if we set $\alpha_i = \frac{1}{N}$, then we have.

$$0 < \alpha_i < C, \quad \sum_{i=1}^{N} \alpha_i y_i = 1$$

$\therefore$ there is also a feasible solution for the dual problem

$\therefore$ Then the strong duality holds.

(b) *Derive* the Lagrange dual and show that it reduces to a constrained quadratic optimization problem. State both the objective function and the constraints of this optimization problem.

Solution:

Use the Lagrange method to derive the Lagrange dual.

the problem can be reformulated as

$$\max_{\alpha,\beta} \min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_i \xi_i + \sum_i \alpha_i \left[1-\xi_i - y_i(w^T\phi(x_i)+b)\right] + \sum_i \beta_i(-\xi_i)$$

$$\text{s.t. } \alpha_i \geq 0, \quad \beta_i \geq 0 \quad \forall i$$

Then we compute the partial derivatives w.r.t (w, b, ξ_i), and set them to 0

$$\frac{\partial L}{\partial w} = w - \sum_i \alpha_i y_i \phi(x_i) = 0$$

$$\frac{\partial L}{\partial b} = \sum_i \alpha_i y_i = 0$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \beta_i = 0$$

$\rightarrow$

$$w^* = \sum_i \alpha_i y_i \phi(x_i)$$

$$\sum_i \alpha_i y_i = 0$$

$$\beta_i = C - \alpha_i \Rightarrow \alpha_i \leq C$$

Then the Lagrange dual can be written as:

$$\max_{\alpha} \frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j y_i y_j \phi^T(x_i)\phi(x_j) + \sum_i \alpha_i - \sum_i \alpha_i y_i \Big(\sum_j \alpha_j y_j \phi^T(x_j)\Big)\phi(x_i)$$

$$= \sum_i \alpha_i - \frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j y_i y_j \phi^T(x_i)\phi(x_j)$$

$$\text{s.t. } 0 \leq \alpha_i \leq C, \ \forall i, \quad \sum_i \alpha_i y_i = 0$$

(c) *Describe* how the solution $(\boldsymbol{w}, b)$ of the primal program can be obtained from a solution of the dual program.

Solution.

We need to derive (w, ξ, b)

① w

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

② ξ_i

if $\alpha_i < C$, then $\beta_i > 0$

$$\beta_i(-\xi_i) = 0$$

$$\xi_i = 0$$

③ b

$$\alpha_i(1 - \xi_i - y_i(w^T\phi(x_i) + b)) = 0$$

if $\alpha_i > 0$, then

$$1 - \xi_i - y_i(w^T\phi(x_i) + b) = 0$$

$$y_i(w^T\phi(x_i) + b) = 1 - \xi_i$$

since $y_i = \pm 1 \rightarrow \frac{1}{y_i} = y_i$

$\therefore$ $b = y_i(1 - \xi_i) - w^T\phi(x_i)$

since $\xi_i = 0$

$\therefore$ $b = y_i - w^T\phi(x_i)$ $\quad 0 < \alpha < C$

b is the soft margin factor. Therefore, we can conclude that the points in the margin have penalty factor $\alpha_i = C$, the support vectors have penalty factor $0 < \alpha_i < C$

日期: /

then the other points has 0 penalty factor.

(d) *Write* a kernelized version of the dual program and of the learned decision function.

Solution:

feature map

define a kernel function as $k(x_i, x_j) = \phi^T(x_i)\phi(x_j)$

$$\boxed{\begin{aligned} \max_{\alpha} \ & \sum_i \alpha_i - \frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j y_i y_j k(x_i, x_j) \\ \text{s.t.} \ & 0 \le \alpha_i \le C, \ \forall i \\ & \sum_i \alpha_i y_i = 0 \end{aligned}}$$

$$\boxed{\begin{aligned} \min_{\alpha} \ & \frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j y_i y_j k(x_i, x_j) - \sum_i \alpha_i \\ \text{s.t.} \ & 0 \le \alpha_i \le C, \ \forall i \\ & \sum_i \alpha_i y_i = 0 \end{aligned}}$$

The learned decision function can be written as

$$\boxed{\begin{aligned} f(x) &= \operatorname{sign}(w^T x + b) \\ &= \operatorname{sign}\Big(\sum_i \alpha_i y_i \phi^T(x_i)\phi(x) + b\Big) \\ &= \operatorname{sign}\Big(\sum_i \alpha_i y_i k(x_i, x) + b\Big) \end{aligned}}$$

Exercise 2: SVMs and Quadratic Programming (10 P)

We consider the `CVXOPT` Python software for convex optimization. The method `cvxopt.solvers.qp` solves quadratic optimization problems given in the matrix form:

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \frac{1}{2}\boldsymbol{x}^\top P\boldsymbol{x} + \boldsymbol{q}^\top \boldsymbol{x} \\ \text{subject to} \quad & G\boldsymbol{x} \preceq \boldsymbol{h} \\ \text{and} \quad & A\boldsymbol{x} = \boldsymbol{b}. \end{aligned}$$

Here, $\preceq$ denotes the element-wise inequality: $(\boldsymbol{h} \preceq \boldsymbol{h}') \Leftrightarrow (\forall_i : h_i \leq h_i')$. Note that the meaning of the variables $\boldsymbol{x}$ and $\boldsymbol{b}$ is different from that of the same variables in the previous exercise.

(a) *Express* the matrices and vectors $P, \boldsymbol{q}, G, \boldsymbol{h}, A, \boldsymbol{b}$ in terms of the variables of Exercise 1, such that this quadratic minimization problem corresponds to the kernel dual SVM derived above.

Solution:

$$\boxed{\begin{aligned} \min_{x} \quad & \tfrac{1}{2} x^T P x + q^T x \\ \text{s.t.} \quad & Gx \preceq h \\ & Ax = b \end{aligned}} \Longrightarrow \boxed{\begin{aligned} \min_{\alpha} \quad & \tfrac{1}{2} \alpha^T y^T K y \alpha - \mathbb{1}^T \alpha \\ \text{s.t.} \quad & \begin{bmatrix} -\alpha \\ \alpha \end{bmatrix} \preceq \begin{bmatrix} 0 \\ C \end{bmatrix} \end{aligned}}$$