Exercise Sheet 8

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)*.

no conditions i.e. always possible to obtain the dual

(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.

$$\max_{\substack{\alpha \geq 0 \\ \beta \geq 0}} \min_{w,b,\xi} \underbrace{\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i + \sum_{i=1}^{N}\alpha_i\left(1-\xi_i - y_i(\vec{w}^T\phi(x_i)+b)\right) + \sum_{i=1}^{N}\beta_i(-\xi_i)}_{J(w,b,\xi)}$$

$$\max_{\boldsymbol{\alpha}} \left\{ \min_{\theta} \left\{ f(\theta) + \sum_{i=1}^{m} \alpha_i g_i(\theta) \right\} \right\} \quad \forall_{i=1}^{m} \; \alpha_i \geq 0,$$

$$\frac{\partial J}{\partial w} = w - \sum_{i=1}^{N}\alpha_i y_i \phi(x_i) = 0 \quad \Rightarrow \quad w = \sum_{i=1}^{N}\alpha_i y_i \phi(x_i)$$

$$\frac{\partial J}{\partial b} = -\sum_{i=1}^{N}\alpha_i y_i = 0$$

$$\frac{\partial J}{\partial \xi_i} = C - \alpha_i - \beta_i = 0 \qquad 0 \leq \alpha_i \leq C$$

box-constraint

$$\max_{\alpha} -\frac{1}{2}\sum_{ij}\alpha_i\alpha_j y_i y_j \phi(x_i)^T\phi(x_j) + 0 + 0 + \sum_{i=1}^{N}\alpha_i$$

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

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

$$w = \sum_{i=1}^{N} \alpha_i y_i \phi(x_i) \qquad b\,?$$

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

$$(C - \alpha_i) \cdot (-\xi_i) = 0$$

$y_i = 1$: $\alpha_i \neq C \Rightarrow \xi_i = 0$; $\alpha_i \neq 0 \Rightarrow 1 - w^T\phi(x_i) - b = 0$

KKT conditions

$$\nabla f(\theta) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\theta) = 0 \qquad \text{(stationarity)}$$

$$\forall_{i=1}^{m} : g_i(\theta) \leq 0 \qquad \text{(primal feasibility)}$$

$$\forall_{i=1}^{m} : \lambda_i \geq 0 \qquad \text{(dual feasibility)}$$

$$\boxed{\lambda_i g_i(\theta) = 0} \qquad \text{(complementary slackness)}$$

if $0 < \alpha_i < C$, $y_i = 1$

$$\Rightarrow b = 1 - w^T \phi(x_i)$$

SV

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

$$f(x) = \operatorname{sign}\left(w^T \phi(x) + b\right)$$

$$= \operatorname{sign}\left(\sum_{i=1}^{N} \alpha_i y_i k(x_i, x) + 1 - \sum_{i=1}^{N} \alpha_i y_i k(x_i, x_{SV})\right)$$

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.

$P = yy^T \odot K$

$q = \mathbb{1}$

$G = \begin{bmatrix} -I \\ I \end{bmatrix}$ $\quad h = \begin{bmatrix} 0 \cdot \mathbb{1} \\ C \cdot \mathbb{1} \end{bmatrix}$

$A = y^T$ $\quad b = 0$

$$\min_{\alpha} \; \frac{1}{2} \sum_{ij} \alpha_i \alpha_j \overbrace{y_i y_j K_{ij}}^{P_{ij}} - \sum_{i=1}^{N} \alpha_i$$

$-\alpha_i \leq 0$

$\alpha_i \leq C$

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