TD - Support Vector Machines

Exercise 1 (preliminaries). Set $\mathcal{X} = \mathbb{R}^d$. Define the hyperplane

$$\mathcal{H}_{w,b} = \{x \in \mathcal{X} : \langle w, x\rangle + b = 0\}$$

for some fixed $w \in \mathcal{X}$ ($w \neq 0$) and $b \in \mathbb{R}$. For a fixed $z \in \mathcal{X}$, consider the problem

$$\min_{x \in \mathcal{H}_{w,b}} \frac{1}{2}\|x - z\|^2 .$$

1. Write the Lagragian function $L(x; \nu)$ associated with this problem.
2. Solve the KKT conditions and characterize the solution.
3. Prove that the distance of a point z to $\mathcal{H}$ is equal to

$$d(z, \mathcal{H}_{w,b}) = \frac{|\langle w, z\rangle + b|}{\|w\|} .$$

Exercise 2 (linearly separable case). Consider a training set formed by couples (x_i, y_i) for $i \in \{1, \dots, n\}$ where x_i is a feature vector in $\mathcal{X}$ and $y_i \in \{-1, +1\}$ for all i. The hyperplane $\mathcal{H}_{w,b}$ is called *separating* if

$$\forall i, \quad y_i(\langle w, x_i\rangle + b) > 0 .$$

In the sequel, we assume that a separating hyperplane exists. Among all separating hyperplanes, we seek to find the one which maximizes the minimum distance

$$f(w, b) = \min_{i=1,\dots,n} d(x_i, \mathcal{H}_{w,b}) .$$

1. Show that if (w, b) defines a separating hyperplane, then $f(w, b) = c(w, b)/\|w\|$ where $c(w, b) = \min_i y_i(\langle w, x_i\rangle + b)$.

Thus, we are interested in solving the problem

$$\max_{w,b} \frac{c(w, b)}{\|w\|} \text{ such that } \forall i,\ y_i(\langle w, x_i\rangle + b) \geq 0 .$$

Let $(w^\star, b^\star)$ be a solution and define

$$v^\star = \frac{w^\star}{c(w^\star, b^\star)} \quad \text{and} \quad a^\star = \frac{b^\star}{c(w^\star, b^\star)}$$

2. Justify that $(w^\star, b^\star)$ and $(v^\star, a^\star)$ define the same separating hyperplane.

3. Prove that $(v^\star, a^\star)$ solves the optimization problem

$$\max_{v,a} \frac{1}{\|v\|} \text{ such that } \forall i,\ y_i(\langle v, x_i\rangle + a) \geq 1\,.$$

4. Deduce that $(v^\star, a^\star)$ solves the optimization problem

$$\min_{v,a} \frac{\|v\|^2}{2} \text{ such that } \forall i,\ 1 - y_i(\langle v, x_i\rangle + a) \leq 0\,. \tag{1}$$

5. Write the Lagrangian $L(v, a; \phi)$.
6. Write the KKT conditions.
7. Let $(v, a; \phi)$ be a saddle point of the Lagrangian. Show that ϕ_i is non-zero only if $y_i(\langle v, x_i\rangle + a) = 1$.

The training points (x_i, y_i) satisfying the above property are the closest to the hyperplane $\mathcal{H}_{v,a}$. The corresponding x_i's are often called *support vectors*.

8. If one is given a dual solution $\phi^\star$, how to recover a primal solution $(v^\star, a^\star)$ from $\phi^\star$?

Define the $n \times n$ matrices $K = (\langle x_i, x_j\rangle)_{i,j=1\ldots n}$, $D = \operatorname{diag}(y_1 \ldots y_n)$ and $\mathbf{1}^T = (1, \ldots, 1)$.

9. Prove that the dual problem reduces to

$$\min_{\substack{\phi \geq 0 \\ y^T\phi = 0}} \frac{1}{2}\phi^T DKD\phi - \mathbf{1}^T\phi\,.$$

10. Give one example of an algorithm that can be used to solve the dual problem (provide the main principle, do not write explicitely the algorithm).
11. Assume that this algorithm has identified a dual solution $\phi^\star$. Write explicitely the classifier as a function of $\phi^\star$.
12. What part of the training data do you need in order to implement the above classifier ?

Exercise 3 (non separable case). Consider the case when a separable hyperplane might not exist. The constraints $1 - y_i(\langle v, x_i\rangle + a) \leq 0$ in Problem (1) may not be jointly feasible. For a fixed $c > 0$, we consider the relaxed problem

$$\min_{v,a} \frac{\|v\|^2}{2} + c\sum_i \xi_i \text{ such that } \forall i,\ 1 - y_i(\langle v, x_i\rangle + a) \leq \xi_i \text{ and } \xi_i \geq 0\,. \tag{2}$$

1. How many constraints has this problem ?
2. Write the Lagrangian function.
3. Show that the dual problem reduces to

$$\min_{\substack{c \geq \phi \geq 0 \\ y^T\phi = 0}} \frac{1}{2}\phi^T DKD\phi - \mathbf{1}^T\phi\,.$$