Machine Learning 1

Homework 9: Support Vector Machine (SVM)

Student: Nguyễn Thị Minh Ngọc - ID: 11219280

1 Problem 1

Given the problem of classifying 2 classes using the SVM algorithm with the assumption that these two classes can be divided by a linear classifier. Accordingly, a new data point x0 will be classified based on the sign of the following expression:

$$y(\mathbf{x}) = \mathbf{w}^T\phi(\mathbf{x}) + b$$

knowing that:

- Target of the problem is $t = 1$ or $t = -1$, corresponding to 2 classes
- If $y(\mathbf{x}_0) = 0, \mathbf{x}_0$ is classified into class $t = 1$, and vice versa
- $\phi(\mathbf{x})$ is a designed matrix of $\mathbf{x}$
- $\mathbf{w}$ and b are weight and bias of the linear classifier respectively

Compute $\mathbf{w}$ and b in this case.

Solution. The support vector machine (SVM) approaches this problem through the concept of the margin, which is given by the perpendicular distance to the closest point $\mathbf{x}_n$ from the data set. In SVM, the decision boundary is chosen to be the one for which the margin is maximized.

The distance of a point $\mathbf{x}_n$ to the decision surface defined by $y(x) = 0$ is given by:

$$\frac{t_n y(\mathbf{x}_n)}{\|\mathbf{w}\|} = \frac{t_n(\mathbf{w}^T\phi(\mathbf{x}_n) + b)}{\|\mathbf{w}\|}$$

Thus, we need to optimize the parameters $\mathbf{w}$ and b in order to maximize this distance.

$$\arg\max_{\mathbf{w},b} \left\{ \frac{1}{\|\mathbf{w}\|} \min_n [t_n(\mathbf{w}^T\phi(\mathbf{x}_n) + b)] \right\}$$

where we have taken the factor $\frac{1}{\|\mathbf{w}\|}$ outside the optimization over n because $\|\mathbf{w}\|$ does not depend on n.

If we make the re-scaling $\mathbf{w} \to \kappa\mathbf{w}$ and $b \to \kappa b$, then the distance from any point $\mathbf{x}_n$ to the decision surface, given by $t_n y(\mathbf{x}_n)/\|\mathbf{w}\|$ is unchanged. We can use this freedom to set:

$$t_n(\mathbf{w}^T\phi(\mathbf{x}_n) + b) = 1$$

for the point that is closest to the surface. In this case, all data points will satisfy the constraints:

$$t_n(\mathbf{w}^T\phi(\mathbf{x}_n) + b) \geq 1, \quad n = 1, 2...N$$

or in other words:

$$\min_n [t_n(\mathbf{w}^T\phi(\mathbf{x}_n) + b)] = 1$$

and our problem became:

$$\arg\max_{\mathbf{w},b} \frac{1}{\|\mathbf{w}\|}$$

which is equivalent to:

$$\arg\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2$$

with constraint: $t_n(\mathbf{w}^T\phi(\mathbf{x}_n) + b) \geq 1$. (the factor of $\frac{1}{2}$ is included for later convenience)

In order to solve this constrained optimization problem, we introduce Lagrange multipliers $a_n \geq 0$, with one multiplier an for each of the constraints, giving the Lagrangian function:

$$L(\mathbf{w}, b, \mathbf{a}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{n=1}^{N} a_n\{t_n(\mathbf{w}^T\phi(\mathbf{x}_n) + b) - 1\}$$

Setting the derivatives of L(w, b, a) with respect to w and b equal to 0:

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{n=1}^{N} a_n t_n \phi(\mathbf{x}_n) = 0$$

$$\frac{\partial L}{\partial b} = \sum_{n=1}^{N} a_n t_n = 0$$

then:

$$\mathbf{w} = \sum_{n=1}^{N} a_n t_n \phi(\mathbf{x}_n)$$

$$0 = \sum_{n=1}^{N} a_n t_n$$

Substitute (3) and (4) into the Lagrangian Function (note that $\|\mathbf{w}\|^2 = \mathbf{w}^T\mathbf{w}$), we got:

$$L(a) = \frac{1}{2}\sum_{n=1}^{N} a_n t_n \phi(\mathbf{x}_n)^T \sum_{m=1}^{N} a_m t_m \phi(\mathbf{x}_m) - \sum_{n=1}^{N} a_n t_n \phi(\mathbf{x}_n) - \sum_{n=1}^{N} a_n t_n b + \sum_{n=1}^{N} a_n \qquad = \sum_{n=1}^{N} a_n$$