ML-As-3

Problem 1: Hard-margin SVM. **(18 pts)**

You are given the following two sets of data points, each belonging to one of the two classes
(class 1 and class -1):

- Class 1 (labeled as +1):

$$(1,2),(2,3) \tag{1}$$

- Class -1 (labeled as -1):

$$(2,1),(3,2) \tag{2}$$

Please find the optimal separating hyperplane using a linear SVM and derive the equation of the hyperplane. Assume the hard-margin SVM.

1. Write down the formulation of SVM, including the separation hyperplane, the constraints and the final optimization problem with parameters. ***(4 pts)***

The **hyperplane** is defined through w and b as a set of points such that

$$H = \{x | w^T x + b = 0\} \tag{3}$$

- $w = (w_1, w_2, \ldots, w_n)$: weight vector
- b: scalar

Subject to the **constraint**

$$y_i(w^T x_i + b) \geq 1, \quad \forall i \tag{4}$$

- y_i is the class label of x_i

Final optimization problem:

$$\min_{w,b} \frac{1}{2}||w||^2 \tag{5}$$

2. Write down the Lagrangian form for this problem using the parameters and Lagrange multipliers. Please also write out its dual form. ***(10 pts)***

The Lagrangian form:

$$L(w,b,\alpha) = \frac{1}{2}||w||^2 - \sum_{i=1}^{n} \alpha_i [y_i(w^T x_i + b) - 1] \tag{6}$$

where $\alpha_i \geq 0$ are the Lagrange multipliers associated with each constraint.

The dual form of the optimization problem

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j \tag{7}$$

subject to

$$\sum_{i=1}^{n} \alpha_i y_i = 0 \quad \text{and} \quad \alpha_i \geq 0 \, \forall i \tag{8}$$

3. Assume that the Lagrangian multipliers α's are all 0.5 and that the point $(1, 2)$ is a support vector for ease of calculation. Please calculate the values of weight vector w and bias b. Write out the explicit form of the hyperplane. ***(4 pts)***

$$w^* = \sum_{i=1}^{m} \alpha_i^* y_i x_i \tag{9}$$

If

$$w^* = 0.5 \times 1 \times (1, 2) + 0.5 \times 1 \times (2, 3) + 0.5 \times -1 \times (2, 1) + 0.5 \times -1 \times (3, 2) = (-1, 1) \tag{10}$$

$$b^* = y_i - \sum_{i=1}^{m} \alpha_i^* y_i x_i^T x_j \tag{11}$$

Since the support vector is $(1, 2)$, we have: $y_j = 1$

$$b^* = 1 - (0.5 \times 1 \times (1, 2)^T \times (1, 2) + 0.5 \times 1 \times (2, 3)^T \times (1, 2) + 0.5 \times -1 \times (2, 1)^T \times (1, 2) + 0.5 \times -1 \times (3, 2)^T \times (1, 2)) = 0 \tag{12}$$

The explicit form of the hyperplane.

$$H = \{x | w^T x + b = 0\} \tag{13}$$

$$H = \{x | (1, 1)^T x = 0\} \tag{14}$$

Problem 2: Soft-margin SVM. **(20 pts)**

Suppose we have the data points $x \in \mathbb{R}^{n \times d}$ with corresponding labels $y \in \mathbb{R}^n$. We want to use a soft-margin SVM to classify these data points with a regularization parameter $C = 1$.

1. Write down the formulation of the soft-margin SVM. for this problem using w, x, y, b and ξ. Write out explicitly their dimensions. ***(3 pts)***

For a soft-margin SVM, the optimization problem can be formulated as follows:

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} \xi_i \tag{15}$$

subject to:

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad \forall i \tag{16}$$

where:

- $w \in \mathbb{R}^d$ is the weight vector,
- $b \in \mathbb{R}$ is the bias,
- $\xi \in \mathbb{R}^n$ is the vector of slack variables, and
- $C = 1$ is the regularization parameter that controls the trade-off between maximizing the margin and minimizing the classification error.

Dimensions:

- w has dimension $d \times 1$,
- x has dimension $n \times d$,
- y has dimension $n \times 1$,

- b is a scalar,
- ξ has dimension $n \times 1$.

2. *Write down the Lagrangian form and derive the dual for the problem. Write down the detailed derivation steps.* ***(12 pts)***

The primal objective function is:

$$L(w, b, \xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i[y_i(w \cdot x_i + b) - 1 + \xi_i] - \sum_{i=1}^{n}\mu_i\xi_i \tag{17}$$

where $\alpha_i \geq 0$ and $\mu_i \geq 0$ are Lagrange multipliers for the constraints $y_i(w \cdot x_i + b) \geq 1 - \xi_i$ and $\xi_i \geq 0$, respectively.

To derive the dual problem, we take the partial derivatives of L with respect to w, b, and ξ and set them to zero:

Partial derivative with respect to w:

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^{n}\alpha_i y_i x_i = 0 \Rightarrow w = \sum_{i=1}^{n}\alpha_i y_i x_i \tag{18}$$

Partial derivative with respect to b:

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^{n}\alpha_i y_i = 0 \Rightarrow \sum_{i=1}^{n}\alpha_i y_i = 0 \tag{19}$$

Partial derivative with respect to ξ_i:

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \mu_i = 0 \Rightarrow \alpha_i \leq C \tag{20}$$

By substituting w back into the objective function, we obtain the dual problem:

$$\max_{\alpha} \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j(x_i \cdot x_j) \tag{21}$$

subject to:

$$\sum_{i=1}^{n}\alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C \tag{22}$$

3. *Obtain the decision boundary.* ***(3 pts)***

The decision boundary is given by:

$$w^T \cdot x + b = 0 \tag{23}$$

where $w = \sum_{i=1}^{n}\alpha_i y_i x_i$ from the dual problem. To classify a new point x, we use the decision function:

$$w^T x + b = (\sum_{i=1}^{n}\alpha_i y_i x_i)^T x + b = 0 \tag{24}$$

4. *Explain why ξ disappears in the dual.* ***(2 pts)***

In the dual formulation, ξ disappears because it only appears in the primal objective function as part of the constraints and is not involved in the dual variables.
By taking the partial derivatives with respect to ξ and setting them to zero, we express ξ_i in terms of α_i and C.
Consequently, ξ_i is not explicitly present in the dual formulation, as it is fully captured by the constraints on α_i (specifically, $0 \leq \alpha_i \leq C$).

Problem 3: Kernel SVM. **(17 pts)**

Consider the following $2D$ dataset with four training points:

$$x_1 = (1, 2), y_1 = 1 \tag{25}$$

$$x_2 = (2, 3), y_2 = 1 \tag{26}$$

$$x_3 = (3, 1), y_3 = -1 \tag{27}$$

$$x_4 = (4, 3), y_4 = -1 \tag{28}$$

We want to use the **polynomial kernel** $k(x_i, x_j) = (x_i^T x_j + 1)^2$ to classify these data points with a soft-margin SVM. The regularization parameters $C = 1$.

To solve Problem 3 on Kernel SVM, let's go through each part step-by-step.

*1. Compute the Kernel Matrix K **(6 pts)***

$$k(x_i, x_j) = (x_i^T x_j + 1)^2 \tag{29}$$

The kernel matrix K is a 4×4 matrix where each element $K_{ij} = k(x_i, x_j)$. Let's compute each entry using the given kernel function.

$K_{11} = k(x_1, x_1) = ((1 \cdot 1 + 2 \cdot 2) + 1)^2 = 36$
$K_{12} = k(x_1, x_2) = ((1 \cdot 2 + 2 \cdot 3) + 1)^2 = 81$
$K_{13} = k(x_1, x_3) = ((1 \cdot 3 + 2 \cdot 1) + 1)^2 = 36$
$K_{14} = k(x_1, x_4) = ((1 \cdot 4 + 2 \cdot 3) + 1)^2 = 121$
$K_{22} = k(x_2, x_2) = ((2 \cdot 2 + 3 \cdot 3) + 1)^2 = 196$
$K_{23} = k(x_2, x_3) = ((2 \cdot 3 + 3 \cdot 1) + 1)^2 = 100$
$K_{24} = k(x_2, x_4) = ((2 \cdot 4 + 3 \cdot 3) + 1)^2 = 324$
$K_{33} = k(x_3, x_3) = ((3 \cdot 3 + 1 \cdot 1) + 1)^2 = 121$
$K_{34} = k(x_3, x_4) = ((3 \cdot 4 + 1 \cdot 3) + 1)^2 = 256$
$K_{44} = k(x_4, x_4) = ((4 \cdot 4 + 3 \cdot 3) + 1)^2 = 676$

Since K is symmetric, we can fill in the remaining entries by symmetry:

$$K = \begin{pmatrix} 36 & 81 & 36 & 121 \\ 81 & 196 & 100 & 324 \\ 36 & 100 & 121 & 256 \\ 121 & 324 & 256 & 676 \end{pmatrix} \tag{30}$$

*2. Set up the Dual Optimization Problem. You can use the results from Problem 2. **(4 pts)***

Using the results from Problem 2, the dual problem for a soft-margin SVM with a kernel function becomes:

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K_{ij} \tag{31}$$

subject to:

$$\sum_{i=1}^{n} \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C \tag{32}$$

where $C = 1$ in this problem.

3. Suppose the Lagrange multipliers α's are

$\alpha_1 = 0.0182$, $\alpha_2 = 0.0068$, $\alpha_3 = 0.0250$, and $\alpha_4 = 0$.

and x_3 is a support vector. *(2 pts)*

The bias b is calculated as:

$$b = y_j - \sum_{i=1}^{n} \alpha_i y_i K_{ij} \tag{33}$$

where we can use x_3 (with $y_3 = -1$) as the support vector.

Substitute $j = 3$:

$$b = y_3 - \sum_{i=1}^{4} \alpha_i y_i K_{i3} \tag{34}$$

Calculating each term in the summation:

$\alpha_1 y_1 K_{13} = 0.0182 \times 1 \times 36 = 0.6552$
$\alpha_2 y_2 K_{23} = 0.0068 \times 1 \times 100 = 0.68$
$\alpha_3 y_3 K_{33} = 0.0250 \times (-1) \times 121 = -3.025$
$\alpha_4 y_4 K_{43} = 0$ (since $\alpha_4 = 0$)

Summing these values:

$$\sum_{i=1}^{4} \alpha_i y_i K_{i3} = 0.6552 + 0.68 - 3.025 + 0 = -1.6898 \tag{35}$$

$$b = -1 - (-1.6898) = -1 + 1.6898 = 0.6898 \tag{36}$$

4. Classify a New Point $x_5 = (2, 1)$ using the learned kernel SVM model. (5 pts)

To classify the point $x_5 = (2, 1)$, we use the decision function:

$$f(x_5) = \sum_{i=1}^{n} \alpha_i y_i k(x_i, x_5) + b \tag{37}$$

Let's compute each $k(x_i, x_5)$:

1. $k(x_1, x_5) = ((1 \cdot 2 + 2 \cdot 1) + 1)^2 = 25$
2. $k(x_2, x_5) = ((2 \cdot 2 + 3 \cdot 1) + 1)^2 = 64$
3. $k(x_3, x_5) = ((3 \cdot 2 + 1 \cdot 1) + 1)^2 = 64$
4. $k(x_4, x_5) = ((4 \cdot 2 + 3 \cdot 1) + 1)^2 = 144$

Now, calculate $f(x_5)$:

$$f(x_5) = \alpha_1 y_1 k(x_1, x_5) + \alpha_2 y_2 k(x_2, x_5) + \alpha_3 y_3 k(x_3, x_5) + \alpha_4 y_4 k(x_4, x_5) + b \tag{38}$$

Substitute the values:

$$f(x_5) = (0.0182 \times 1 \times 25) + (0.0068 \times 1 \times 64) + (0.0250 \times -1 \times 64) + (0 \times -1 \times 144) = 0.6898 \tag{39}$$

Calculate each term:

$0.0182 \times 25 = 0.455$
$0.0068 \times 64 = 0.4352$
$0.0250 \times -64 = -1.6$

Adding them up with b:

$$f(x_5) = 0.455 + 0.4352 - 1.6 + 0.6898 = -0.02 \tag{40}$$

Since $f(x_5) < 0$, we classify x_5 as belonging to class -1.