Machine Learning

58122204 Xie xing   *Instructor: Xu-Ying Liu*

SEU — 2024 Spring

Assignment 3

§ Support Vector Machine & Bayesian Classifiers §

Problem 1: Support Vector Machine

(1)

1. (a) Generalized Lagrangian function

$$L(\omega, b, \epsilon, \alpha, \mu) = \frac{1}{2}||\omega||^2 + C\sum_{i=1}^{N}\epsilon_i, -\sum_{i=1}^{N}\alpha_i(y_i(\omega^T x_i + b) - 1 + \epsilon) - \sum_{i=1}^{N}\mu_i\epsilon_i \tag{1.1}$$

, where $\alpha_i >= 0, \mu_i >= 0$. The dual problem of the original function is

$$\max_{\alpha\geq 0,\mu\geq 0}\min_{\omega,b,\epsilon} L(\omega, b, \epsilon, \alpha, \mu) \tag{1.2}$$

Find the partial derivative :

$$\nabla_w L(w, b, \epsilon, \alpha, \mu) = w - \sum_{i=1}^{N}\alpha_i y_i x_i = 0 \tag{1.3}$$

$$\nabla_b L(w, b, \epsilon, \alpha, \mu) = -\sum_{i=1}^{N}\alpha_i y_i = 0 \tag{1.4}$$

$$\nabla_{\epsilon_i} L(w, b, \epsilon, \alpha, \mu) = C - \alpha_i - \mu_i = 0 \tag{1.5}$$

Solutions have to:

$$\begin{cases} w = \sum_{i=1}^{N}\alpha_i y_i x_i \\ \sum_{i=1}^{N}\alpha_i y_i = 0 \\ C - \alpha_i - \mu_i = 0 \end{cases} \tag{1.6}$$

Bringing in $L(w, b, \epsilon, \alpha, \mu)$ gets:

$$\min_{\omega,b,\epsilon} L(w, b, \epsilon, \alpha, \mu) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i \cdot x_j) + \sum_{i=1}^{N}\alpha_i \tag{1.7}$$

calcuate the maximum of $\min_{\omega,b,\epsilon} L(w, b, \epsilon, \alpha, \mu)$ on α, we get the dual problem:

$$\max_{\alpha} -\frac{1}{2}\sum_{i}^{N}\sum_{j}^{N}\alpha_i\alpha_j y_i y_j(x_i \cdot x_j) + \sum_{i=1}^{N}\alpha_i \tag{1.8}$$

$$s.t. \sum_{i}^{N}\alpha_i y_i = 0 \tag{1.9}$$

$$C - \alpha_i - \mu_i = 0 \tag{1.10}$$

$$\alpha_i \geq 0 \tag{1.11}$$

$$\mu_i \geq 0, i = 1, 2, ..., N \tag{1.12}$$

First of all, there is a partial derivative solution $\sum_{i=1}^{N} \alpha_i y_i = 0$,;
Secondly, the Lagrange multiplier is greater than or equal to 0, that is $\alpha, \mu >= 0$, when seeking partial derivatives, we get $C - \alpha_i - \mu_i = 0$;
Finally, comprehensively we get $0 \leq \alpha_i \leq C$.

So the dual problem is:

$$\max_{\alpha} -\frac{1}{2} \sum_{i}^{N} \sum_{j}^{N} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) + \sum_{i=1}^{N} \alpha_i \tag{1.13}$$

$$s.t. \sum_{i}^{N} \alpha_i y_i = 0 \tag{1.14}$$

$$0 \leq \alpha_i \leq C \tag{1.15}$$

2. (b) At the saddle point, the solutions of the primal and dual problems are the same, satisfying the following KKT conditions:

1. Primal feasibility condition:
$$y_i(w \cdot x_i + b) \geq 1 - \epsilon_i, \quad \epsilon_i \geq 0, \quad \forall i \tag{1.16}$$

2. Dual feasibility condition:
$$0 \leq \alpha_i \leq C, \quad \forall i \tag{1.17}$$

3. Complementary slackness condition:
$$\alpha_i [y_i(w \cdot x_i + b) - 1 + \epsilon_i] = 0, \quad \forall i \tag{1.18}$$

4. Slack variable condition:
$$\mu_i \epsilon_i = 0, \quad \forall i \tag{1.19}$$

5. Gradient condition:
$$w = \sum_{i=1}^{n} \alpha_i y_i x_i \tag{1.20}$$

So the primal problem and the dual problem are strongly dual, thus the dual problem is equivalent to the primal problem when at its saddle point.

(2)

1. (a) The corresponding mapping function is:

$$\phi(x) = (x_1^2, \sqrt{2} x_1 x_2, x_2^2) \tag{1.21}$$

2. (b)

3. (c) The decision boundary in the figure (original feature space) is shown in Figure 1.

The decision boundary in the figure (Regenerative Kernel Hilbert Space) is shown in Figure 2.

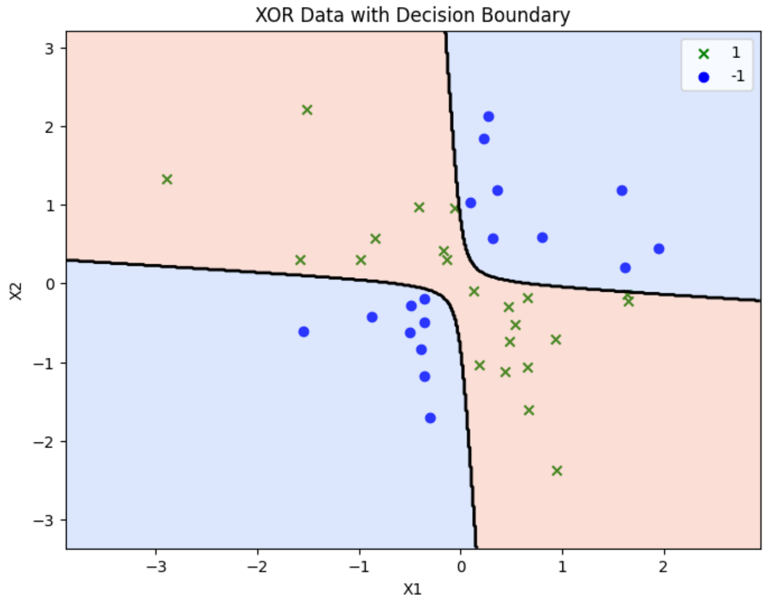

Figure 1: XOR Data with Decision Boundary

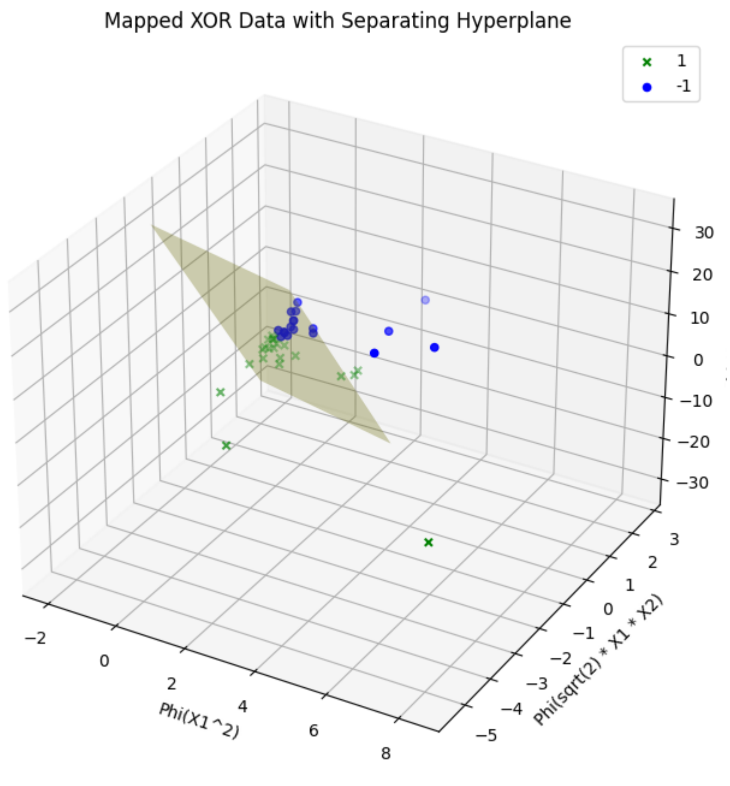

Figure 2: Mapped XOR Data with Separating Hyperplane