Machine Learning

58122204 Xie xing *Instructor: Xu-Ying Liu*

SEU — 2024 Spring

Assignment 3

§ Model Selection and Evaluation & Neural Networks §

Problem 1: Support Vector Machine

(1)

1. (a) Generalized Lagrangian function

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

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 $L(w, b, \epsilon_i, \alpha, \mu)$ gets:

$$L(w, b, \epsilon_i, \alpha_i, \mu_i) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j u_i y_j\left(x_i\cdot x_j\right) + \sum_{i=1}^{N}\alpha_i \tag{1.7}$$

Next, we originally found the maximum value of the above equation. If we add a negative sign to the entire equation, we can transform it into finding its minimum value, that is,

$$\min_{\alpha}\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j u_i y_j\left(x_i\cdot x_j\right) - \sum_{i=1}^{N}\alpha \tag{1.8}$$

After getting the objective function, sort out the constraints. 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:

2. (b)

(2)

1. (a) The corresponding mapping function is:

$$\phi(x) = (x_1^2, \sqrt{2}x_1x_2, x_2^2) \tag{1.9}$$

2. (b)

3. (c)