题目解析与答案

(E) 以下是软间隔分类器的优化问题

$$
\begin{aligned}
&\underset{\beta_0,\ldots,\beta_p,\epsilon_1,\ldots,\epsilon_n,M}{\text{maximize}} && M \\
&\text{subject to} && \textstyle\sum_{j=1}^{p}\beta_j^2 = 1 \\
& && y_i(\beta_0+\beta_1 x_{i1}+\ldots+\beta_p x_{ip}) \geq M(1-\epsilon_i) \\
& && \epsilon_i \geq 0, \textstyle\sum_{i=1}^{n}\epsilon_i \leq C
\end{aligned}
$$

其中，C 是一个非负调节参数。

(i) 解释变量 M 的作用

变量 M 在优化问题中的作用是确定分类间隔的大小。在软间隔支持向量机 (SVM) 中，我们试图最大化这个间隔 M。在没有任何约束 (即 $\epsilon_i = 0$) 的理想情况下，M 表示最小间隔。然而，由于现实数据集可能无法完全线性可分，我们引入了松弛变量 ϵ_i 来允许一些误差，从而找到一个平衡，使得大多数数据点可以被正确分类，同时最大化分类间隔 M。

(ii) 解释 C 和变量 ϵ_i 的关系，其中 $i = 1,\ldots,n$，n 是训练数据集中的观测数量。

参数 C 和松弛变量 ϵ_i 之间的关系如下：

- C 是一个超参数，用来控制模型对误分类的容忍度。- 松弛变量 ϵ_i 表示第 i 个样本的误差程度。如果 $\epsilon_i = 0$，则表示第 i 个样本被正确分类；如果 $\epsilon_i > 0$，则表示第 i 个样本被误分类或者在间隔边界以内。- 约束条件 $\sum_{i=1}^{n}\epsilon_i \leq C$ 表示所有误差的总和不能超过 C。这意味着 C 控制了误分类点的数量和程度。

较小的 C 值表示我们对误分类的容忍度较低，即我们希望间隔更大，但可能会有更多的误分类点。较大的 C 值表示我们对误分类的容忍度较高，即我们希望大多数点被正确分类，但允许间隔较小。

English Explanation and Answer

(E) Below is the optimization problem for a soft margin classifier

$$\begin{array}{ll} \underset{\beta_0,\ldots,\beta_p,\epsilon_1,\ldots,\epsilon_n,M}{\text{maximize}} & M \\ \text{subject to} & \sum_{j=1}^{p} \beta_j^2 = 1 \\ & y_i(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}) \geq M(1 - \epsilon_i) \\ & \epsilon_i \geq 0, \sum_{i=1}^{n} \epsilon_i \leq C \end{array}$$

where C is a non-negative tuning parameter.

(i) Explain the role of variable M

The role of the variable M in the optimization problem is to determine the size of the margin in the soft margin Support Vector Machine (SVM). We aim to maximize this margin M. Ideally, without any constraints (i.e., $\epsilon_i = 0$), M represents the minimum margin. However, since real-world datasets may not be perfectly linearly separable, we introduce slack variables ϵ_i to allow for some errors, thereby finding a balance that maximizes the margin M while correctly classifying most data points.

(ii) Explain the relationships between C and the variables ϵ_i, where $i = 1, \ldots, n$, and n is the number of observations in the training dataset.

The relationship between the parameter C and the slack variables ϵ_i is as follows:

- C is a hyperparameter that controls the model's tolerance for misclassification. - The slack variable ϵ_i represents the error term for the i-th sample. If $\epsilon_i = 0$, the i-th sample is correctly classified; if $\epsilon_i > 0$, the i-th sample is misclassified or within the margin. - The constraint $\sum_{i=1}^{n} \epsilon_i \leq C$ indicates that the total amount of slack (error) cannot exceed C. Thus, C controls the number and extent of misclassified points.

A smaller C value means lower tolerance for misclassification, resulting in a larger margin but potentially more misclassified points. A larger C value means higher tolerance for misclassification, allowing most points to be correctly classified but with a smaller margin.