

## 机器学习课程实践作业数据相关说明

- 1、**数据说明：**数据文件“P0\_P49\_train”中，前 29 列为监测数据，第 30 列（最后一列）为数据标签。数据标签共包含 4 种标签，分别为 1,2,3,4。文件中的数据为训练样本，可以用于建模。
- 2、**建模目标：**区分四类数据标签，在要求的综合衡量指标下，其分数最高。
- 3、**衡量指标：**模型效果衡量指标由多个指标加权构成，综合指标的公式如下：

$$CPI = \frac{1}{3} \cdot Macro\ F1 + \frac{1}{3} \cdot Weighted\ F1 + \frac{1}{3} \cdot Macro\ Precision$$

$Macro\ F_1$ ：宏平均  $F1-score$ ，衡量模型在每个类别上的表现，侧重于类别均衡性，公式如下：

$$Macro\ F1 = \frac{1}{n} \sum_{i=1}^n F1_i$$

其中， $i$  表示数据类别， $n$  是类别总数， $F1_i$  是第  $i$  个类别的  $F1$  分数，下同。

$Weighted\ F1$ ：加权  $F1-score$ ，考虑类别不平衡性，对多样性较大的数据集有良好适应性，公式如下：

$$Weighted\ F1 = \frac{\sum_{i=1}^n w_i \cdot F1_i}{\sum_{i=1}^n w_i}$$

其中， $w_i$  是第  $i$  类别的权重，这里认为是该类别样本数量占总样本数量的比例。

$Macro\ Precision$ ：宏平均精确率，关注模型在各类别中的正预测表现，适用于需要高精确性的任务，公式如下：

$$Macro\ Precision = \frac{1}{n} \sum_{i=1}^n Precision_i$$

其中， $Precision_i$  是第  $i$  个类别的精确率。

- 4、**提交作业内容：**提交模型实现的完整代码（不限制语言），提交说明文档。说

明文档需要包含但不限于以下内容：实现思路，模型说明，运行结果，结果分析，可能的改进方向分析。校历第 16 周周一将给大家发测试样本文件“P0\_P49\_test”，文档中的运行结果需要包含训练样本和测试样本的综合指标计算结果。

5、**提交时间：**校历第 16 周(6 月 15 日)周日前,提交至指定邮箱 [hrrr709@163.com](mailto:hrrr709@163.com)。