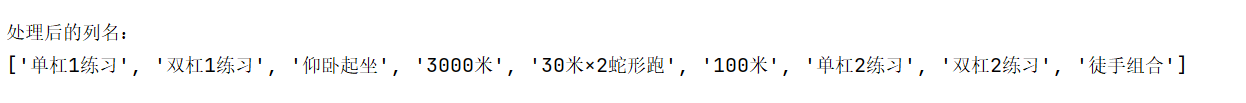
Sy7-学员体能类型KNN分析

1. 数据加载与初步处理

首先，数据通过读取 Excel 文件载入，针对数据表中各项成绩的评定，提取以 .1 结尾的列，这些列代表了评定结果，如优秀、良好、及格、不及格。接着，对于缺失的数据，实验采用了列的众数进行填补，确保数据的完整性。清洗后的数据表头如图所示：



1. 学员体能类型KNN分析

为了分类学员的体能类型，本实验根据体能项目的评分定义了详细的分类规则。首先，将评定转换为分数，其中优秀对应 5 分、良好对应 4 分、及格对应 3 分、不及格对应 2 分。基于不同类型的体能项目计算了学员在力量、耐力、速度、灵活协调和核心力量五个方面的平均得分。分类规则如下：

全能型：五个方面的平均得分达到或超过 4.5 分。

 力量型：力量项目平均得分达到或超过 4.0 分，而耐力项目得分不超过 3.0 分。

 耐力型：耐力项目平均得分达到或超过 4.0 分，而力量项目得分不超过 3.0 分。

 速度型：速度项目平均得分达到或超过 4.0 分，其他项目得分相对较低。

 灵活协调型：灵活协调和核心力量项目平均得分均达到或超过 4.0 分。

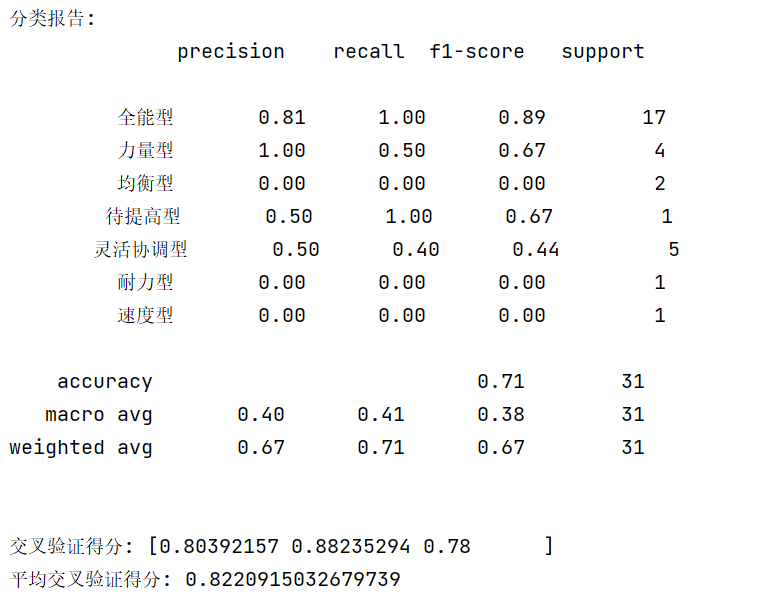
 均衡型：所有方面的平均得分在 3.5 到 4.5 之间。

 待提高型：所有方面的平均得分低于 3.5 分。

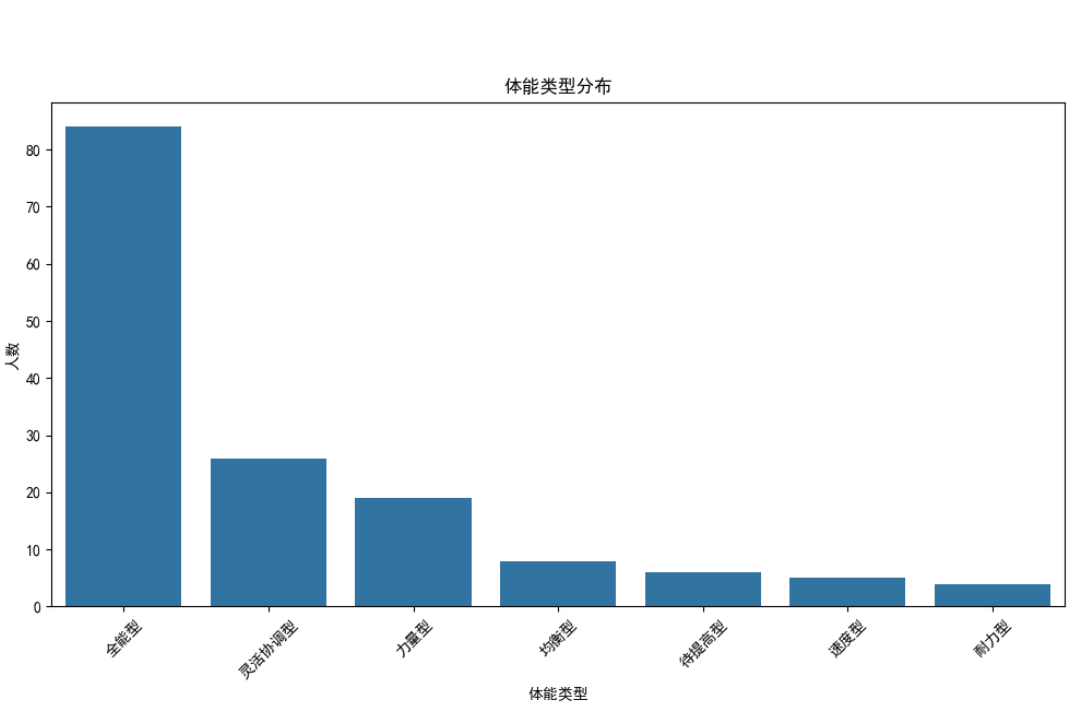
为了消除量纲的影响并提高分类的准确性，我们使用了 StandardScaler 对数据进行标准化处理。随后，我们采用 K 最近邻算法（KNN）作为分类模型，通过网格搜索（GridSearchCV）优化了模型的超参数，包括邻居数、距离度量和权重方式。网格搜索的参数设置为：n\_neighbors = [1, 3, 5]，metric = ['euclidean', 'manhattan']，weights = ['uniform', 'distance']。

将数据集按照 80:20 的比例划分为训练集和测试集，并在训练集上进行了 3 折交叉验证，以评估模型的稳定性。

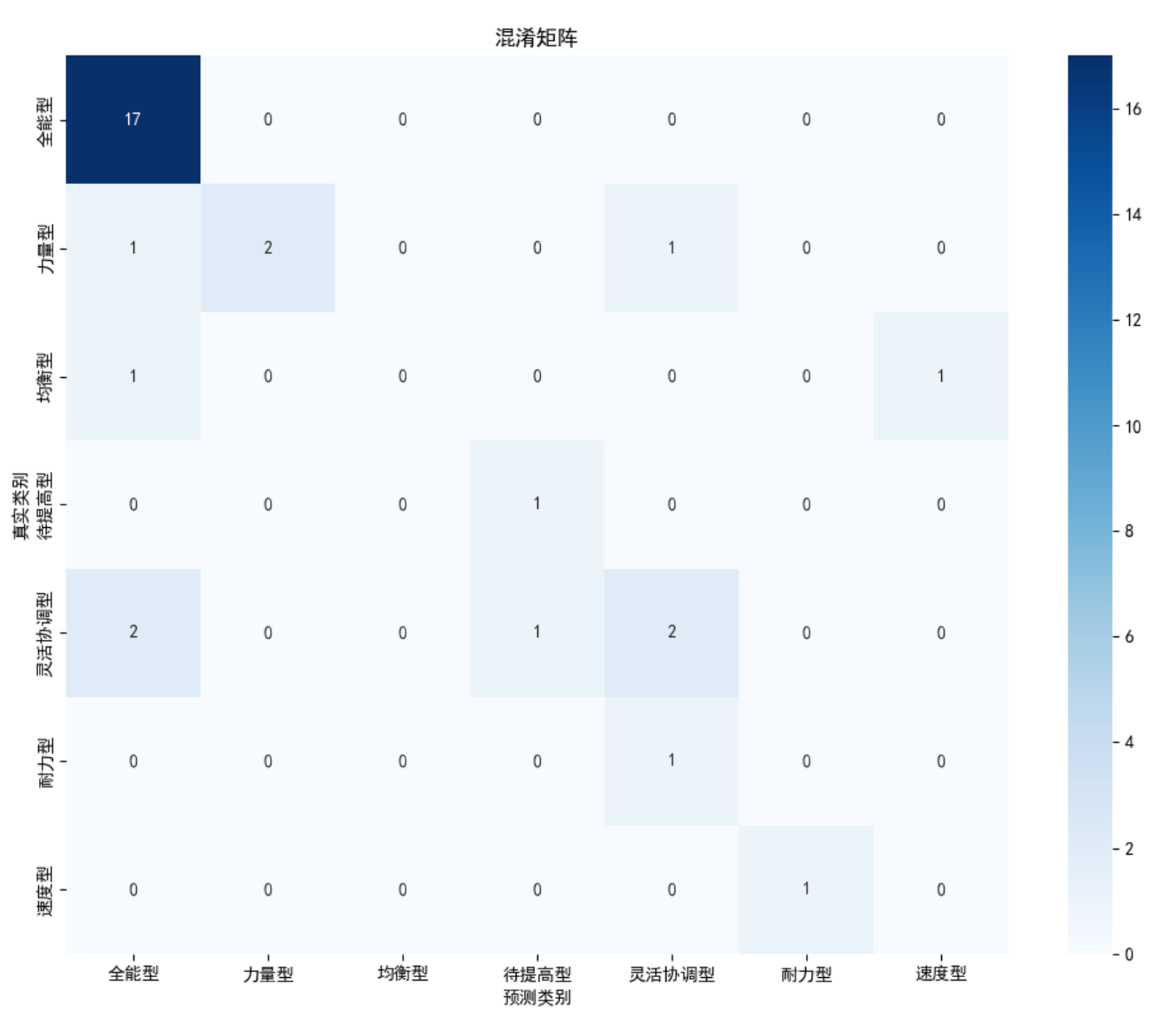
最后使用优化后的 KNN 模型对测试集进行预测，生成了详细的分类报告，包括精确率、召回率和 F1 分数。此外，绘制了混淆矩阵，直观地展示了模型预测结果与实际类别的对应关系。我们还进行了交叉验证，平均准确率达到了 82.2%，显示模型在训练集上的表现较为稳定。



1. 实验结论



经过分类，学员的体能类型分布如图所示，全能型为84 名，灵活协调型：26 名，力量型：19 名，均衡型：8 名，待提高型：6 名，速度型：5 名，耐力型：4 名。



实验通过网格搜索得出的最优参数为：{'metric': 'euclidean', 'n\_neighbors': 3, 'weights': 'distance'}。测试集上的整体准确率为 71%，由混淆矩阵可以看出，由于样本不均衡，小样本类别的预测表现较差。

本实验通过数据处理、分类规则定义、模型建立和优化，有效地对学员的体能类型进行了分类。模型在大样本类别上表现较好，但对于小样本类别需要进一步优化。未来可以通过增加数据量、采用不同的算法或进行特征工程等方法，提高模型在所有类别上的预测精度。