** **

**数学建模竞赛论文**

**论文主题：混合物中待定成分的检测**

**作者1：李聪聪 院系：物理科学与技术学院**

**作者2：贾海龙 院系：物理科学与技术学院**

**作者3：魏弘量 院系：物理科学与技术学院**

摘要

本文研究了如何检测混合物中是否含有特定成分，并为其设计最优模型的问题。已知一个混合物中明确是否含有特定成分的训练数据，包含20000个混合物样本，已知一个未明确是否含有特定成分的测试数据，包含5000个混合物样本，我们通过分析训练数据，给出判定特定成分存在的主要指标，并指出是否存在模糊区域，其次建立模型，判断测试数据中每一个混合物是否含有特定成分。针对这一问题我们给出自己的方法以及优化方案。

在任务一中，针对给定的训练数据，我们对特征维度为7的训练集数据进行主成分分析（PCA），得到主成分向量，由主成分向量得到了判定特定化合物是否存在的主要指标。同时，我们通过PCA的方法对训练集进行了数据降维，保留两个主特征，将七维特征的分类问题化简为二维特征的分类问题。

针对任务二，要求我们判断是否存在模糊区域。我们决定将七维的特征空间降维至二维的主特征空间。在二维主特征空间中，定义一个“明显度”函数，采用明显度函数去量化“模糊”，从而得到二维特征空间的明显度函数的分布图，将明显度低于某一阈值的区域定义为模糊区域。由于两个主特征向量是原来7个特征的线性叠加，因此求出二维主特征空间的模糊区域后，我们由线性关系反推出七维特征空间中的模糊区域。

针对任务三，我们沿用前两问主成分分析的思路，将测试集降至二维，在二维特征空间中采用支持向量机（SVM）的数学模型完成分类任务。我们分别采取了基于线性核、高斯径向基（RBF）函数核、多项式核的支持向量机作为分类器完成分类任务，成功将测试集中前10个数据分类；并且对各个分类器模型进行比较探究。

关键词：支持向量机（SVM）、主成分分析法（PCA）、明显度函数