

题目：探索搜索、优化策略在学习任务降维中的应用

假设存在一个训练数据集，其有 n 个样本（记为： $\{S_1, S_2, \dots, S_n\}$ ）， d 个特征（记为： $\{f_1, f_2, \dots, f_d\}$ ）。每个样本可表示为一个 d 维向量。另外，每个样本隶属于一个可能标记 L 。若样本与该标记相关，其值为 1。反之，其值为 0。

科学问题：该数据集的 d 个特征中包含大量无关，冗余，甚至噪声特征，如何从这 d 个特征中找到一个对标记 L 预测有效的特征子集，如 k 个特征（ $k \ll d$ ），是一个有挑战性的学习任务。

- （1） 将该科学问题看作成搜索问题。A.描述该搜索问题问题空间的状态集合、初始状态、动作集合、目标检测、路径代价。B.给出一种搜索算法的求解方案。
- （2） 将该科学问题看作成优化问题，给出一种优化算法的求解方案。
- （3） 在一个真实数据集上（如：UCI 数据集），对（1）和（2）给出的二种求解方案进行实验分析，应包含：实验设置，实验结果，结果分析。

UCI ML Repository: <https://archive.ics.uci.edu/ml/datasets.php>

- （4） 降维是机器学习中一个重要的学习任务。对于学习的任务，谈谈你的理解。