

题目：Ensemble classification based on supervised clustering for credit scoring

领域： 信用评分

核心创新点： 利用有监督的聚类方法解决随机取样所导致的类别不平衡

论文结构与实现方法

有监督的聚类与普通聚类的区别在于不同于普通聚类根据整个数据集进行划分,有监督的聚类只利用相同类别进行划分。

重新定义 Kmeans 算法的评估函数（实际上没有过多改动）利用该评估函数决定最后的 K 值通过这种聚类方法确实能适当解决下面的问题

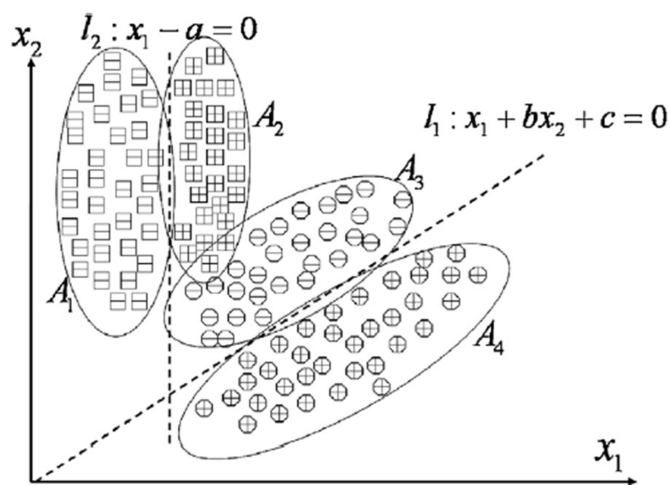


Fig. 2. Illustration of supervised clustering for ensemble classification.

利用两个开源数据集、多种方法结合进行实验，实验很丰富，基本上的检验标准都有提到，并多重对比，提升可信度。最后又利用自己的数据进行实际实验。

优秀之处：

方法简单，操作易行，实验的内容很棒，值得学习，各种方法比较和分析。先用开源数据集显示方法优势，再用自己数据集实际操作，可信度高

可采用的点：

这种定义方法并实现的思路很符合我们的需求。实验的流程、比较的方式都值得学习。针对多重数据集检验来提升可信度这一点也很好。