

文章题目：

Credit scoring using the clustered support vector machine

Terry Harris Credit Research Unit, Department of Management Studies, The University of the West Indies, Cave Hill Campus, P.O. Box 64, Barbados

这篇论文没有什么太多实质性的内容，唯一的改进是先将 data 分成 cluster 分别训练模型。

主要内容：

将 SVM 运用到 Credit scoring 中可以取得较好的效果，但是训练时间较长，需要与数据规模成指数级别的复杂度，为了在保证正确率的情况下减少模型训练的复杂度，本文自称采用了一种 CSVM 的方法，实际上就是先用 K-MEANS 算法（根据实际需求设置 K 的值）将数据集分成 K 个 cluster，然后在每个 cluster 中分别训练 svm，这样总的训练复杂度会比较小，并且从本文中提供的实验结果来看，这种做法确实减少了训练时间并且最后的结果会优于一般的逻辑回归和 svm，本文中采用 AUC 曲线进行测评。