题目 Training Classifiers via SGD

姓名：陈东东 学号：MF1633002 邮箱：[chendd@lamda.nju.edu.cn](mailto:chendd@lamda.nju.edu.cn) 联系方式 15062285720

(南京大学 计算机科学与技术系, 南京 210093)

# 实现细节

* **log-likelihood maximum logistic regression**

1. **初始化beta=0的值，初始化步长arfa，初始化lamda=1，设置总共的迭代步数T。**
2. **在特征上扩增一维，作为优化b之用。**

**2.在全体数据集中随机选择一个点,按照公式计算sub-gradient（重复，直接一次性在外面取完）**

**3.根据梯度下降公式更新beta。**

**4.重复2和3直至达到想要测试的迭代次数t（当t=T时，为最后结果），将此时的beta作为模型参数，对训练数据和测试数据进行测试，画图，并计算出此时的目标函数值。**

# 结果

## 实验设置

* 数据来源：

German数据集来源于：Professor Dr. Hans Hofmann Institut f"ur Statistik und "Okonometrie Universit"at Hamburg FB Wirtschaftswissenschaften Von-Melle-Park 5 2000 Hamburg 13.

MNIST数据集来源于：It is a subset of a larger set available from NIST。It was originally selected and experimented with by Chris Burges and Corinna Cortes using bounding-box normalization and centering.

* 数据预处理：

由于题目给的数据都是数值型数据，所以不需要预处理。

* 参数设置：

对于german数据集：在K-Medoids算法中设置聚类数K=2，r=500,T=50.在spectral clustering算法中K=2，k={3,6,9}.

对于MNIST数据集：在K-Medoids算法中设置聚类数K=10，r=500,T=30.在spectral clustering算法中K=2，k={3,6,9}.

## 实验结果

分别对german和MNIST两个数据集进行实验，得到他们的纯度和基尼指数结果如下表：