
题目：PCA,SVD 和 ISOMAP 的应用

姓名：叶康 学号：MG1633093 邮箱：604379334@qq.com 联系方式：18851822212

(南京大学 计算机科学与技术系, 南京 210093)

1 实现细节

1.1 PCA

PCA 实现的伪代码如下所示：

输入：训练集，k 的大小

输出：投影矩阵

```
function Topca(data,numberofk)
    meandata<--mean(data)
    dataminusmean<--data-meandata
    covdata<--cov(dataminusmean,rowvar=0)//对每列求协方差
    evals,evects<--linalg.eig(covdata)//求相应特征值特征向量
    evalsex<--argsort(evals)    //对特征值从小到大排序
    evalsex<--evalsex[-1:-(numberofk+1):-1]    //取最大的前 k 个特征值
    evectsex<--evects[:,evalsex]    //取与特征值相应的前 k 个特征向量
    return evectsex    //返回训练集的投影矩阵
end function
```

1.2 SVD

SVD 实现的伪代码如下所示：

输入：训练集，k 的大小

输出：投影矩阵

```
function ToSVD(data,numberofk):
    dataU,datasigma,datav<--linalg.svd(data) //应用 linalg 包进行 svd 处理
    for x in dataU do    //选择前 k 列
        temu<--x[:numberofk]
    selectu<--temu
    selectsigma<--datasigma[:numberofk]    //选择前 k 个 sigma
    selectv<--datav[:numberofk]    //选择前 k 行
    finalv<--selectv.T
    return finalv    //返回训练集的投影矩阵
end function
```

1.3 ISOMAP&&MDS

ISOMAP&MDS 实现的伪代码如下所示：

输入：各节点距离的邻接矩阵，k 的大小

输出：降维数据矩阵

```
function ISOMAPMDS(dataarray,numberofk): #对数据进行 MDS 处理
    sqrdata<--dataarray*dataarray.T //各节点距离平方的矩阵
    H<--eye(nofdata)-1.0/nofdata //计算 S 当中的 I - U/n
    T<---0.5 * dot(dot(H,sqrdata),H) //计算 S
    evals,evects<--linalg.eig(T) //对 S 进行特征分解
    evalsex<--argsort(evals)
    evalsex<--evalsex[-1:-(numberofk+1):-1] //选取最大的前 k 个
    evectsex<--evects[:,evalsex]
    diagvec<--diag(sqrt(evalsex)) //对其开方求出 sigma
    returndata<--dot(evectsex,diagvec) //进行点乘得到 Dk
    return returndata
end function
```

2 结果

2.1 实验设置

代码是在 Windows 10 系统下 Python3.5 环境中运行的，其中包用到了numpy，数据来源于 <https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary.html>

2.2 实验结果

实验结果如下图所示：

K	测试集 1			测试集 2		
	10	20	30	10	20	30
PCA	56.31%	55.33%	51.22%	75.72%	74.11%	72.41%
SVD	59.22%	58.25%	56.31%	75.86%	76.41%	74.80%
ISOMAP (k=10)	52.42%	49.51%	53.39%	71.21%	70.35%	68.54%

实验结果中准确率不高有很大一部分原因是最后用了 1-NN 的方法，若使用更好的分类方法如将 1-NN 变为 k-NN 或者使用 svm 那准确率会得到很大的提升。