

B 样条基—主成分分析

摘要: 为了对非高斯情况下的数据进行主成分分析, 本文先对数据以基函数的形式展开, 再利用传统的主成分分析方法进行分析。

一、改进措施

主成分分析具有使得数据集更易使用, 完全无参数限制, 去除噪声等优点。但是当主成分分析在非高斯分布情况下时, 主成分方法得出的主元可能并不是最优的。因而有学者提出了核主成分方法, 借鉴核主成分方法的思想, 本文对原始的主成分分析通过基函数展开的形式进行改进。在这里, 本文选择了 B 样条基进行展开。

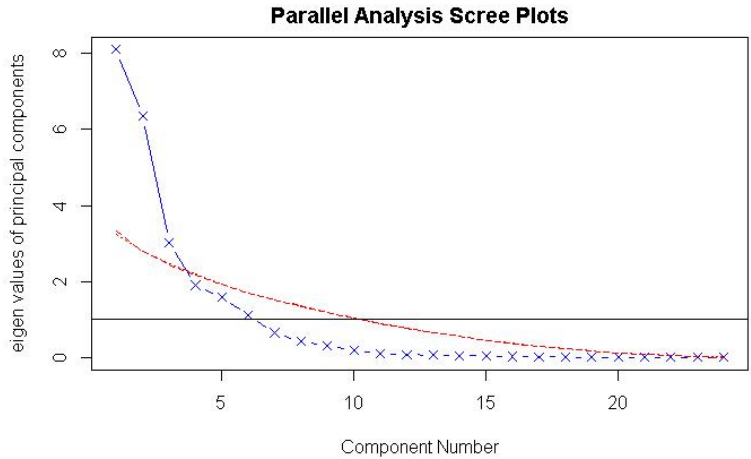
二、数据来源与采集时间

本文数据来自西南金点子财富终端的 2021 年 4 月 16 日 27 个多元金融板块和 35 个石油板块的股票数据的共计 62 个数据, 在进行分类时保留后 5 个数据作为测试集, 其余作为训练集。

三、R 程序分析结果

首先, 按 3 次 B 样条基展开得到 24 个变量。

然后, 根据碎石图选择主成分的个数为 6 个, 如图(1):



图(1) 碎石图

进而得到因子载荷矩阵, 并计算主成得分对各个样本进行排序, 表(1)是不同主成分方法得分排序的结果 (仅展示前 10 个)。

表(1) 不同主成分方法得分排序的结果

	PCA	PCA 旋转	SPCA	BPCA
1	仁东控股	仁东控股	海德股份	泛海控股
2	鲁信创投	鲁信创投	中油资本	南华期货
3	泛海控股	泛海控股	越秀金控	鲁信创投
4	绿庭投资	绿庭投资	中航资本	绿庭投资
5	*ST 安信	*ST 安信	九鼎投资	*ST 安信
6	南华期货	南华期货	五矿资本	仁东控股
7	渤海租赁	瑞达期货	江苏租赁	瑞达期货
8	瑞达期货	渤海租赁	华金资本	香溢融通
9	吉艾科技	吉艾科技	中粮资本	渤海租赁
10	民生控股	民生控股	经纬纺机	民生控股

由于没有建立一套对于主成分降维效果的评价指标体系, 所以无法对本文提出的方法以

指标的方式进行量化对比。

而主成分通常可以与不同分类或聚类方法结合达到更好的分类效果，本文尝试将 PCA、BPCA 与 SVM 结合进行分类，对比其分类效果，如表(2)不同方法的分类效果。

表(2) 不同方法的分类正确率

	LR	PCA+LR	BPAC+LR
内预测正确率	0.577	0.557	0.596
外预测正确率	0.7	0.6	0.9

四、结论与展望

采用 BPCA 与 LR 集合的方式分类效果是最好的，但是本次仅仅使用了一组数据，不具有代表性。