

实验九：Dimensionality Reduction

一、说明

- 实验采用 jupyter notebook, 请填写完代码后提交完整的 ipynb 文件
- 文件命名 规则: 班级_ 姓名 _ML2019_HW9 .ipynb, 如计科 1701_张三 _ML2019_HW9.ipynb
- 提交方式: 采用在线提交至:
<http://pan.csu.edu.cn:80/invitation/beb398af-5f23-4b2e-8abe-eeeffbbf5bab>
- 实验提交截至日期: 2020.1.12 23:59

二、实验内容

机器学习领域中所谓降维就是指采用某种映射方法, 将原高维空间中的数据点映射到低维度的空间中。降维的本质是学习一个映射函数 $f: x \rightarrow y$, 其中 x 是原始数据点的表达, 目前最多使用向量表达式。 y 是数据点映射后的低维向量表达, 通常 y 的维度小于 x 的维度。 f 可能是显式的或隐式的、线性的或非线性的。

本实验指导用户了解主成分分析 (PCA), 奇异值分解 (SVD) 和线性鉴别分析 (LDA) 算法原理, 并简单实现 PCA 算法、SVD 算法和 LDA 算法。

三、实验目标

- 熟悉并实现 PCA 算法。
- 熟悉并实现 SVD 算法。
- 熟悉并实现 LDA 和随机森林算法。

四、实验操作步骤

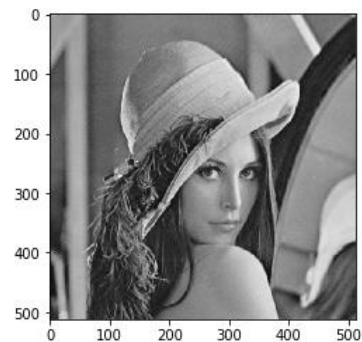
任务 1 实现 PCA 算法。

任务 2 采用 SVD 算法进行图像压缩。

提示：最后压缩后的图像对比原图损失了不到 10%的信息，却减少了 60%的数据量。

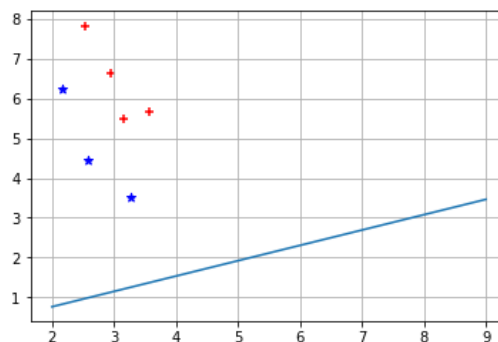


奇异总值216517.25778738962
达到90%时的奇异总值194865.53200865065
前110个奇异值的奇异总值195022.84511463193



任务 3 实现 LDA 算法。

提示：LDA 应得到类似以下输出，实现类内间隔小，类间间隔大的目标



[<matplotlib.lines.Line2D at 0x142ed7a2a58>]

