



中国科学技术大学
University of Science and Technology of China

Lab 4

姓名：_____ 高茂航 _____

学号：_____ PB22061161 _____

日期：_____ 2024.4.20 _____

Lab 4

1 Problem Descriptions

1. 对一个随机生成的 $\mathbf{A}_{4 \times 3}$ 矩阵进行奇异值分解;
2. 对鸢尾花数据集进行主成分分析, 并可视化。

2 Analysis and Algorithms

2.1 SVD Decomposition

1. Jacobi 方法求矩阵 $\mathbf{A}\mathbf{A}^T$ 的特征值;
2. 将 $\mathbf{A}\mathbf{A}^T$ 的特征值从大到小排序后, 用高斯消元法求解每个特征值对应的特征向量并归一化, 然后转置为列向量组, 得到 $\mathbf{U}_{4 \times 4}$;
3. 矩阵 $\Sigma_{4 \times 3}$ 的主对角元为 $\mathbf{A}\mathbf{A}^T$ 的非零特征值的算术平方根 (本题中仅考虑了实数情况), 其余位置全为 0;
4. 矩阵 $\mathbf{V}_{3 \times 3}^T$ 每行即 $\mathbf{U}^T\mathbf{A}$ 对应行除以 Σ 矩阵对应的主对角元。

2.2 PCA and Visualization

1. 对 m 个 n 维数据进行中心化处理, 按列排列构成矩阵 $\mathbf{X}_{n \times m}$;
2. 计算协方差矩阵 $\mathbf{Var}_{n \times n} = \frac{1}{m}\mathbf{X}\mathbf{X}^T$ 的特征值;
3. 选取最大两个特征值对应的特征向量构成矩阵 $\mathbf{P}_{2 \times n}$, 则 $\mathbf{Y}_{2 \times m} = \mathbf{P}\mathbf{X}$ 即 PCA 后的结果, 也就是把四维数据压缩为二维, 每个数据对应二维平面上的一个点;
4. 用 python 实现结果的可视化。

Lab 4

3 Results

3.1 SVD Decomposition

```
A:
0.726249 0.915339 0.594324
0.515358 0.975149 0.661561
0.528652 0.788493 0.0741007
0.32985 0.583744 0.0309722

非对角元平方和: 8.24265
非对角元平方和: 3.2226
非对角元平方和: 0.0168247
非对角元平方和: 0.00305683
非对角元平方和: 0.00185419
非对角元平方和: 0.00074638
非对角元平方和: 0.00034016
非对角元平方和: 7.63028e-05
非对角元平方和: 1.51042e-06
非对角元平方和: 5.85301e-08
非对角元平方和: 5.02149e-10
非对角元平方和: 4.96548e-11
非对角元平方和: 1.52447e-11

U:
0.612104 -0.26095 -0.715213 0.213785
0.595669 -0.468124 0.632988 -0.159261
0.426854 0.668536 -0.0597555 -0.606004
0.297157 0.515581 0.290217 0.749432

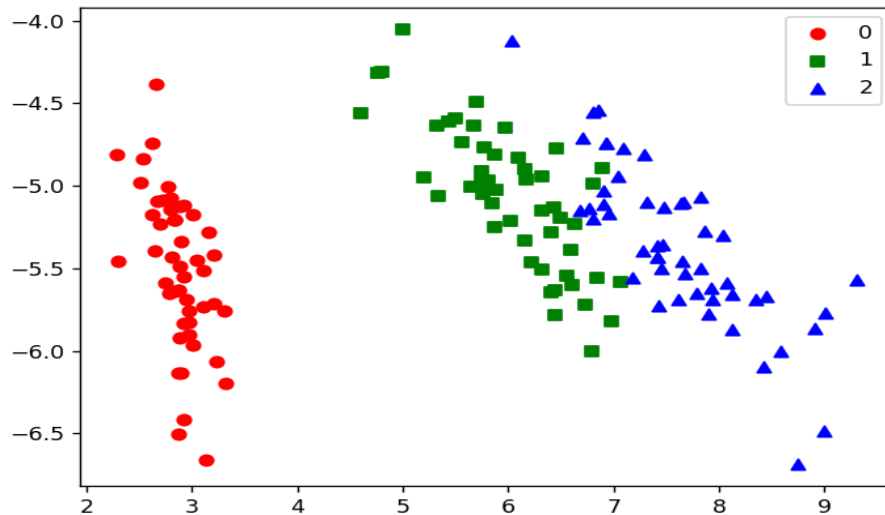
Sigma:
2.12612 0 0
0 0.43086 0
0 0 0.154494
0 0 0

VT:
0.50571 0.77662 0.375658
0.215201 0.308116 -0.926689
-0.835432 0.549478 -0.0113119
```

3.2 PCA and Visualization

```
特征值:
4.19668 0.240629 0.0780004 0.0235251
P:
0.36159 -0.0822689 0.856572 0.358844
-0.65654 -0.729712 0.175767 0.0747065
Y:
2.82714 2.79595 2.62152 2.76491 2.78275 3.23145 2.69045 2.88486 2.62338 2.8375 3.00482 2.8982 2.72391 2.28614 2.8678 3.12747 2.88882 2.86302 3.31227
2.924 3.20081 2.96811 2.29549 3.20821 3.15517 3.00343 3.04229 2.94895 2.87152 2.8785 2.92288 3.10127 2.86371 2.91418 2.8375 2.64434 2.88611 2.8375
2.5295 2.92102 2.7412 2.65913 2.51305 3.10583 3.30251 2.79568 2.97377 2.67102 2.96866 2.80743 6.79614 6.44375 6.9754 5.69231 6.59848 6.15178 6.60657
4.75988 6.55464 5.50115 5.00026 6.02244 5.77368 6.49539 5.33648 6.43892 6.17094 5.74588 6.4537 5.55459 6.62758 5.86813 6.80781 6.43185 6.22535 6.41
098 6.84238 7.06874 6.3238 5.20401 5.441 5.31946 5.64634 6.89008 6.09862 6.31855 6.73177 6.32421 5.75654 5.67586 5.97437 6.4015 5.74022 4.80426 5.86
688 5.84247 5.88658 6.15303 4.60288 5.80915 8.04307 6.92542 8.12783 7.48216 7.8611 8.90822 6.03073 8.44335 7.83102 8.42948 7.17328 7.31368 7.67672 6
.85594 7.09661 7.41609 7.46059 9.00011 9.30603 6.80967 7.93951 6.70944 9.01061 6.89901 7.7872 8.12554 6.76897 6.80201 7.6342 7.89891 8.3523 8.74369
7.67008 6.95445 7.29098 8.58786 7.65633 7.41621 6.6802 7.619 7.82565 7.43379 6.92542 8.07467 7.93073 7.45536 7.03701 7.27539 7.41297 6.90101
-5.64133 -5.14517 -5.17738 -5.0036 -5.64865 -6.06251 -5.23262 -5.48513 -4.74393 -5.20803 -5.96666 -5.33624 -5.08698 -4.81144 -6.50092 -6.65948 -6.13
281 -5.63386 -6.19397 -5.8352 -5.71259 -5.75476 -5.45634 -5.42025 -5.28351 -5.17567 -5.45261 -5.68941 -5.63401 -5.12465 -5.11733 -5.7328 -6.13471 -6
.41475 -5.20803 -5.39192 -5.92152 -5.20803 -4.83447 -5.55078 -5.58578 -4.38186 -4.98042 -5.51064 -5.75742 -5.07204 -5.82509 -5.09415 -5.901 -5.42973
-6.00016 -5.63392 -5.81891 -4.48912 -5.39011 -4.8974 -5.59861 -4.31362 -5.54368 -4.59415 -4.05223 -5.21244 -4.76683 -5.19036 -5.06291 -5.78296 -4.9
6275 -4.9828 -4.7729 -4.73323 -5.23051 -5.2479 -4.98716 -5.13233 -5.4651 -5.64433 -5.55939 -5.58212 -5.15239 -4.94964 -4.61219 -4.63723 -5.00301 -4.
89352 -4.83144 -5.50978 -5.72276 -4.94404 -5.048 -4.63506 -4.6452 -5.28091 -4.91247 -4.3063 -4.81151 -5.10354 -5.0231 -5.33379 -4.56316 -4.96771 -5.
30288 -4.7398 -5.65666 -5.1336 -5.27284 -5.86189 -4.12337 -5.6671 -5.06918 -6.0951 -5.55676 -5.09857 -5.53 -4.53831 -4.77542 -5.43354 -5.35545 -6.48
627 -5.56799 -4.55371 -5.69151 -4.70914 -5.7715 -5.11069 -5.64811 -5.87309 -5.13559 -5.1983 -5.10387 -5.77724 -5.68747 -6.68525 -5.0964 -5.17092 -4.
81326 -6.00049 -5.45363 -5.36277 -5.15022 -5.68621 -5.49733 -5.72399 -4.7398 -5.5907 -5.61823 -5.50214 -4.9397 -5.39324 -5.4386 -5.03184
```

Lab 4



4 Conclusion

(a) Jacobi 方法迭代过程中矩阵非对角元素的平方和呈下降趋势, 且求得特征值是为对称矩阵特征值的近似值, 但在分解过程中遇到了一个问题: 理论上 $\mathbf{A}\mathbf{A}^T$ 和 $\mathbf{A}^T\mathbf{A}$ 的非零特征值相同, 但是程序计算出的结果会有一些差异, 因此 的主对角元不能取 $\mathbf{A}^T\mathbf{A}$ 的特征值的算术平方根, \mathbf{U}^T 的每行元素也不能直接取 $\mathbf{A}^T\mathbf{A}$ 的特征向量, 而要按照算法描述里的第 4 步计算。

(b) 可视化结果中, 明显可见标签为 0, 1, 2 的数据分别集中分布在二维平面的不同位置, 说明 PCA 对数据进行了有效的降维处理, 提取出了数据的有效信息。