Asmt 7: DimReduce

Vai Suliafu, u0742607
Wednesday, April 07

1 Singular Value Decomposition (70 points)

First we will compute the SVD of the matrix `A` we have loaded

```
import numpy as np
from scipy import linalg as LA
U, s, Vt = LA.svd(A, full_matrices=False)
```

Then take the top k components of `A` for values of $k = 1$ through $k = 10$ using

```
Uk = U[:,:k)
Sk = S[:k,:k]
Vtk = Vt[:k,:]
Ak = Uk @ Sk @ Vtk
```

A (40 points): Compute and report the L_2 norm of the difference between `A` and `Ak` for each value of k using `LA.norm(A-Ak,2)`

These results can be seen below in Figure 1.

B (10 points): Find the smallest value k so that the L_2 norm of `A-Ak` is less than 10% that of `A`; k might or might not be larger than 10.

Result: K = 19. More information on these results can be seen in Figure 2 below.

C (20 points): Treat the matrix as 5000 points in 40 dimensions. Plot the points in 2 dimensions in the way that minimizes the sum of residuals squared, and describe briefly how you did it.

In order to plot the points of A in two dimensions such that the sum of residuals squared was minimized, I computed the right singular matrix of A

using Singular Value Decomposition. I then took the top two right singular vectors, v1 and v2 and computed the dot products (ai, v1) and (ai, v2) for each ai in A. The results are shown below in Figure 3.

2 Frequent Directions and Random Projections (30 points)

Use the stub file `FD.py` to create a function for the Frequent Directions algorithm (**Algorithm 16.2.1**). Consider running this code on matrix `A`.

A (30 points): Measure the error $\max_{\|x\|=1} |\|Ax\|^2 - \|Bx\|^2|$ as `LA.norm(A.T @ A - B.T @ B)`

- How large does `l` need to be for the above error to be at most $\|A\|_F^2/10$?
- How does this compare to the theoretical bound (e.g. for $k = 0$).
- How large does `l` need to be for the above error to be at most $\|A - A_k\|_F^2/10$ (for $k = 2$)?

- Given that $\|A\|_F^2/10 = 6462.898$, L = 10 satisfies the desired property. More details can be seen in Figure 4 at the bottom.
- Given $k = 0$, the theoretical bound $\|A - A_k\|_F^2/L = \|A\|_F^2/10$. Thus, we observe that the theoretical upper bound is upheld by our results.
- Given that $\|A - A_{k=2}\|_F^2/10 = 3788.277$, L = 12 satisfies the desired property. More details can be seen in figure 5 at the bottom.

```
K=1 Error: 106.80400116745678
K=2 Error: 98.93131911624847
K=3 Error: 93.82335917328975
K=4 Error: 75.57170361248677
K=5 Error: 62.986805432603646
K=6 Error: 61.56671899951177
K=7 Error: 27.67987618113234
K=8 Error: 26.445160689201842
K=9 Error: 26.26866531890766
K=10 Error: 24.60274852950896
```

Figure 1: Norms of Differences For K = 1:10

```
K=0 Error: 123.85117487853258
K=1 Error: 106.80400116745678
K=2 Error: 98.93131911624847
K=3 Error: 93.82335917328975
K=4 Error: 75.57170361248677
K=5 Error: 62.986805432603646
K=6 Error: 61.56671899951177
K=7 Error: 27.67987618113234
K=8 Error: 26.445160689201842
K=9 Error: 26.26866531890766
K=10 Error: 24.60274852950896
K=11 Error: 23.165638822024004
K=12 Error: 22.449606464686596
K=13 Error: 20.338010577561334
K=14 Error: 20.028588680980608
K=15 Error: 19.55224006699753
K=16 Error: 18.964504214996836
K=17 Error: 15.056429412265645
K=18 Error: 14.281112088474206
K=19 Error: 5.899057236177754

Cutoff reached at K=19
```

Figure 2: Searching for the Specified K

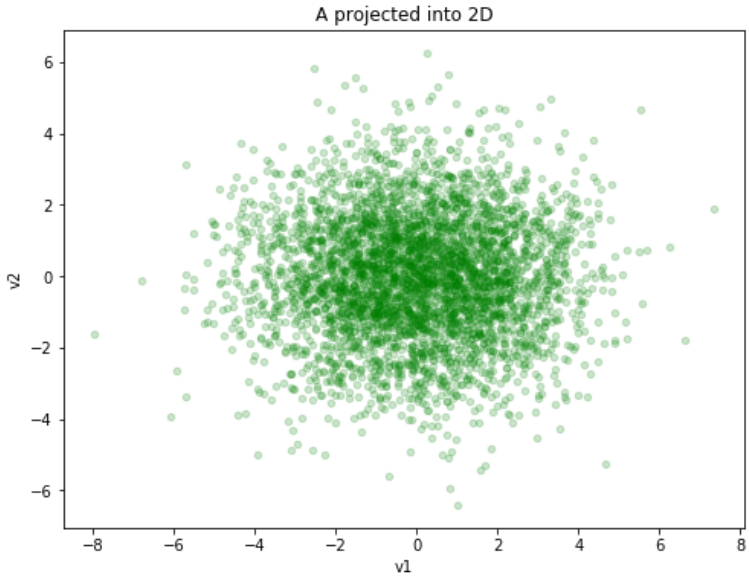

Figure 3: A Projected onto Two Dimensions

```
L=1 Error: 24588.196040625095
L=2 Error: 24573.59466303805
L=3 Error: 24569.601066983083
L=4 Error: 24532.90120524591
L=5 Error: 22738.28438886213
L=6 Error: 19976.98957779534
L=7 Error: 15068.788940405282
L=8 Error: 11593.135341230645
L=9 Error: 7850.558263746615
L=10 Error: 5413.748309828258

Cutoff reached at L=10
```

Figure 4: Error Bound by $\|A\|_F^2/10$

```
L=1 Error: 24588.196040625095
L=2 Error: 24573.59466303805
L=3 Error: 24569.601066983083
L=4 Error: 24532.90120524591
L=5 Error: 22738.28438886213
L=6 Error: 19976.98957779534
L=7 Error: 15068.788940405282
L=8 Error: 11593.135341230645
L=9 Error: 7850.558263746615
L=10 Error: 5413.748309828258
L=11 Error: 4230.324041245445
L=12 Error: 3547.281179694719

Cutoff reached at L=12
```

Figure 5: Error Bound by $\|A - A_{k=2}\|_F^2/10$