Asmt 5: Regression

Gopal Menon
Turn in through Canvas by 2:45pm:
Wednesday, April 12

1 Singular Value Decomposition (20 points)

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

$$[U, S, V] = svd(A)$$

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

$$Uk = U(:, 1 : k)$$
$$Sk = S(1 : k, 1 : k)$$
$$Vk = V(:, 1 : k)$$
$$Ak = Uk * Sk * Vk'$$

A: (10 points): Compute and report the L_2 norm of the difference between A and Ak for each value of k using

$$norm(A - Ak, 2)$$

Table 1: L_2 norm of $A - Ak$ for each value of k

k	L_2 Norm
1	40.483
2	26.717
3	25.000
4	22.192
5	17.675
6	15.813
7	13.351
8	12.188
9	9.1206
10	9.0000

B (5 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.

The L_2 norm of A is 120.19 and 10% of that is 12.019. From table 1, we can see that the smallest value of k such that the L_2 norm of $A - Ak$ is less than 10% that of A is 9.

C (5 points): Treat the matrix as 1125 points in 30 dimensions. Plot the points in 2 dimensions in the way that minimizes the sum of residuals squared.

The first two right singular vectors were used and all 1125 points in 30 dimensions were projected on them to get 1125 points in 2 dimensions. Since the first two singular vectors represent eigen vectors, this projection will result in the least sum of residuals squared. The plot is shown below in figure 1.

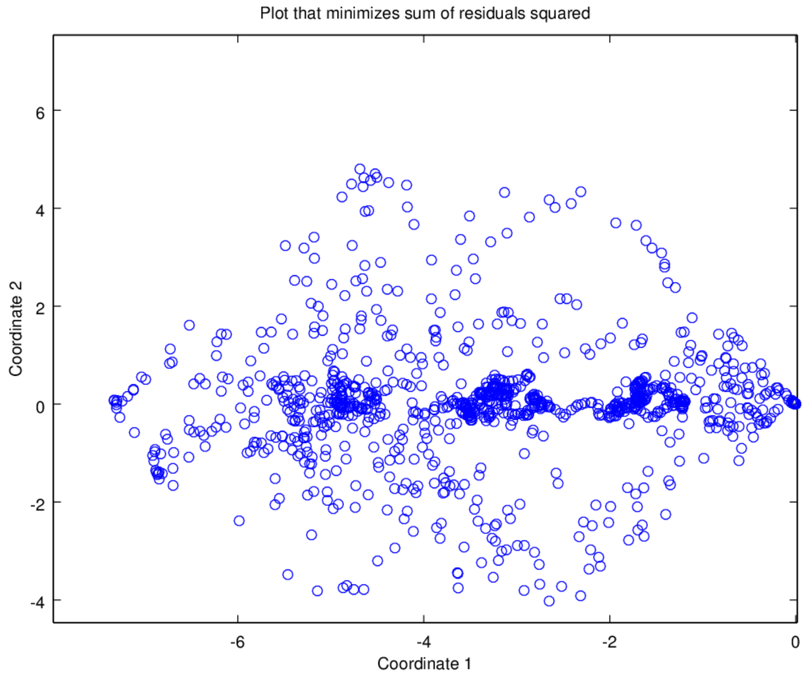

Figure 1: Points plotted in 2D to minimize sum of residuals squared

2 Frequent Directions and Random Projections (40 points)

A (20 points):

- How large does l need to be for the above error to be at most $\frac{\|A\|_F^2}{10}$?

$$\frac{\|A\|_F^2}{10} = 1903.7$$

Table 2: Error for values of l

l	Error
3	2289.9
4	1427.9
5	965.01
6	703.24
7	494.53
8	350.53
9	247.37
10	176.91

From table 2 we can see that with $l = 4$, the error is at most $\frac{\|A\|_F^2}{10}$.

- How does this compare to the theoretical bound (e.g. for $k = 0$).

The theoretical bound is given by $\frac{\|A-A_k\|_F^2}{l-k}$. When $k = 0$, the bound becomes $\frac{\|A\|_F^2}{l}$. This bound evaluates to 4759.3 when $l = 4$.

- How large does l need to be for the above error to be at most $\frac{\|A-A_k\|_F^2}{10}$ for $k = 2$?

  For $k = 2$, $\frac{\|A-A_k\|_F^2}{10} = 295.18$. From table 2, we can see that when $l = 9$, the error is at most $\frac{\|A-A_k\|_F^2}{10}$ for $k = 2$.