Matlab Midterm

Freya Genesis D Mello

November 5, 2018

1 Problem 1

1.1 (a)

The linear regression is modelled by constructing a Vandermonde Matrix.

1.2 (b)

The coefficient values are computed as -2.6356 0.1436 0.5514 3.2229 -0.4329.

1.3 (c)

The ezplot command was used to plot the data points to form the elliptical orbit depicting the planet's positions as the fit curve on the same graph.

2 Problem 2

2.1 (a)

2.1.1 (i)

The rank is 2 as the matrix A has 3 columns but the third column is dependant on the other two columns. Hence the number of pivot columns is 2 and hence the rank is 2. The exact rank has a value of 3 as computed from rank(A, 1e-16), as it considers the double precision floating point errors.

2.1.2 (ii)

The SVD is computed for the matrix A and U, S, V matrices are derived by the command. The diagonal matrix S has only 2 singular values and hence the rank is 2, even on considering the double precision.

On account to the floating point double precision error, it is observed that the 3rd singular value (1.926123792480998e-16) is lesser than eps.

2.1.3 (iii)

Assigning singular values across the diagonal elements in powers of (0.9) from 1 to n. On running the designated rank(B) command, a value of rank is 276. On tracking, manually the value on the 276th column and row, a value of 2.34926822809294e-13 is found. For a double precision floating point, the command rank(B,eps) is used to infer a rank value of 342.

When the tolerance is equal to the eps (2.22*e-16) rank is 342 but when the default tolerance (from MATLAB) is considered, it is 276. It depends on the tolerance.

2.2 (b)

2.2.1 (i)

Depending on the values of the random number generated between [-0.005, 0.005], the ellipses on repeated iterations either show a trend of increasing or decreasing elliptical orbit size.

2.2.2 (ii)

On including the tolerance values in the equation to plot the ellipses, for a value of k = 1, rank is 3 and we obtain a hyperbola as a and c values are omitted and hence the equation no longer stands for an ellipse hence making it a hyperbola. When k=2, we obtain a parabola which gives rank value 4 which will result in a new equation thus formed depending on the changing coefficient value. When k = 3,4,5, a rank 5 matrix is obtain which gives all non zero coefficients which enabled the creation of the ellipse for all three cases (k=3,4,5).

3 Problem 3

3.1 (a)

MATLAB script that parses the Senate data.

3.2 (b)

The singular values are maximum in the beginning till they gradually turn into comparable values(almost identical) after the 10th value.

3.3 (c)

Scatter plot of the first and second columns of U.

3.4 (d)

Because U1 and U2 are corresponding to first and second largest singular values, they show the maximum rotation of the unit circle. The U matrix on being applied to the singular vale matrix, rotates and stretches the unit circle in the clock wise direction. The sigma 1 and sigma 2 values will scale the unit vectors by length sigma 1 and sigma 2 respectively. Whereas the V transpose will transform the scaled ellipse by rotating and stretching in the anti clockwise direction.

3.5 (e)

The Democrats are majorly clustered at the higher end of the accuracy as compared to the Republicans. Both the Democrats and Republicans have higher accuracy. 90% of the data points have accuracy greater than .8. The plot is versus the U1 which uses the maximum variance and because of which the classes are far from each other but the respective data points are very closely clustered to each other.

3.6 (f)

Exercise repeated for for the House data.

3.7 (g)

Negative U1 value depicts the Republicans (marked as green) and positive U1 value depicts the Democrats, for House and vice versa for Senate. This gives us a brief idea of U1 and how it affects the behaviour of Republicans and Democrats. The data points in either of the classes(Republicans and Democrats) are well separated but the respective data points of each class are closely clustered.(As can be seen from the plots). This is inferred when the first column of the largest singular value is considered to identify trends in the data.