

Assignment4

name: Yuqi Zhou

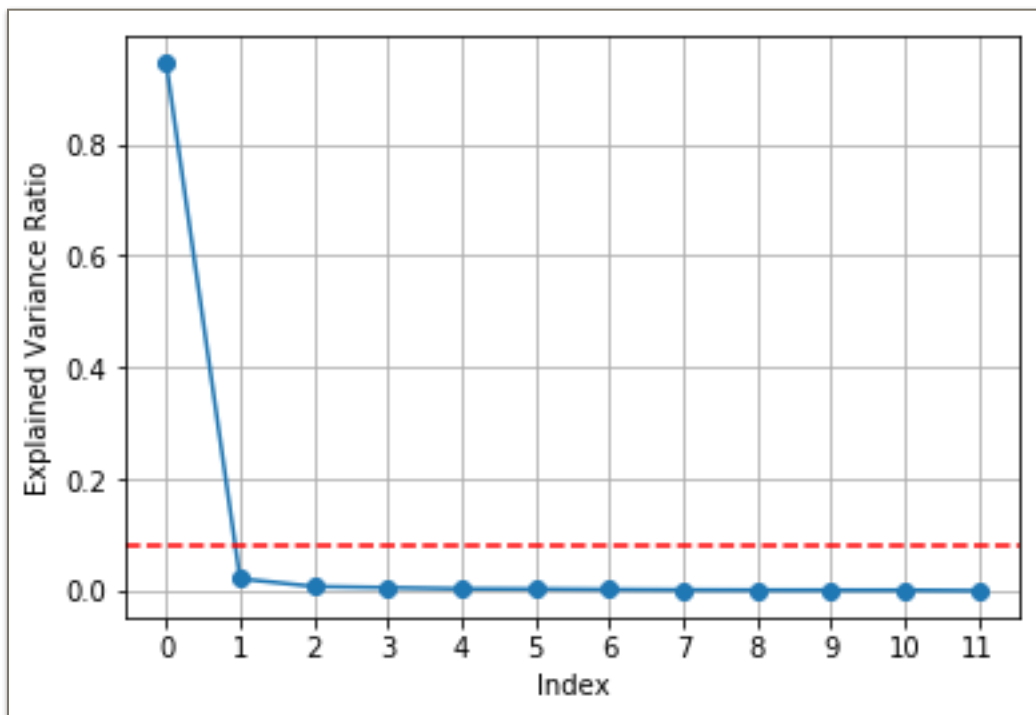
ID:A20423555

Question1:

- a) (5 points). What is the maximum number of principal components that you can get?

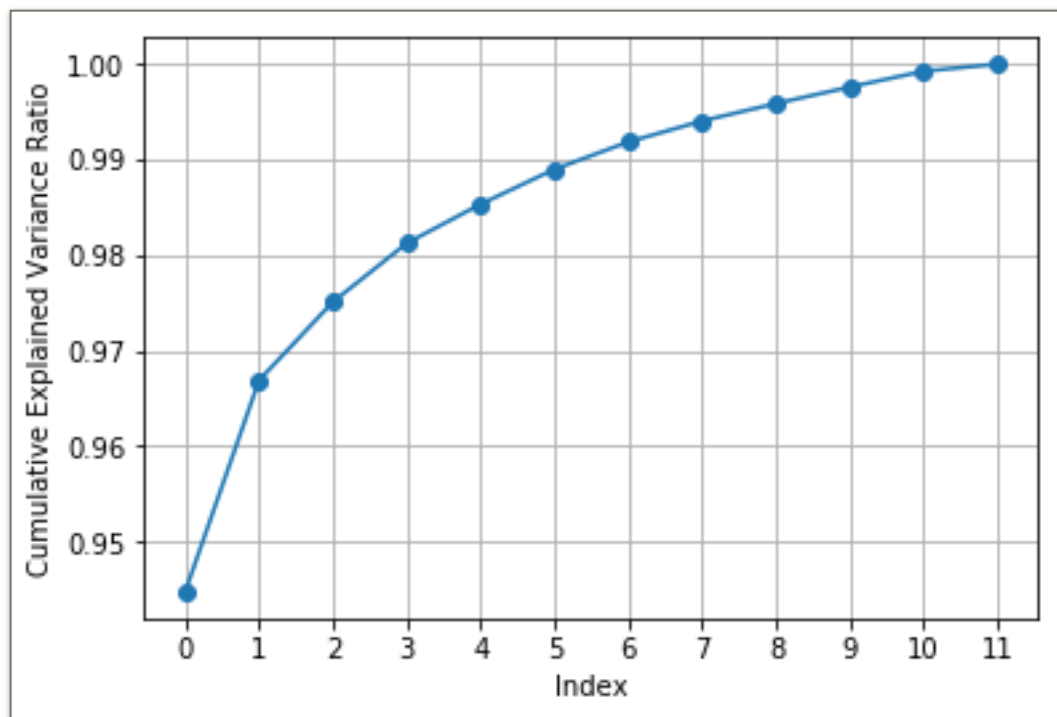
Twelve principal components.

- b) (5 points). Plot the Explained Variances against their indices. Add a horizontal reference line whose value is the reciprocal of the number of variables. Label the axes and add grid lines to the axes.



- c) (5 points). Suppose I am required to explain at least 95% of the total variance, then which major (i.e., top k) principal components should I select?

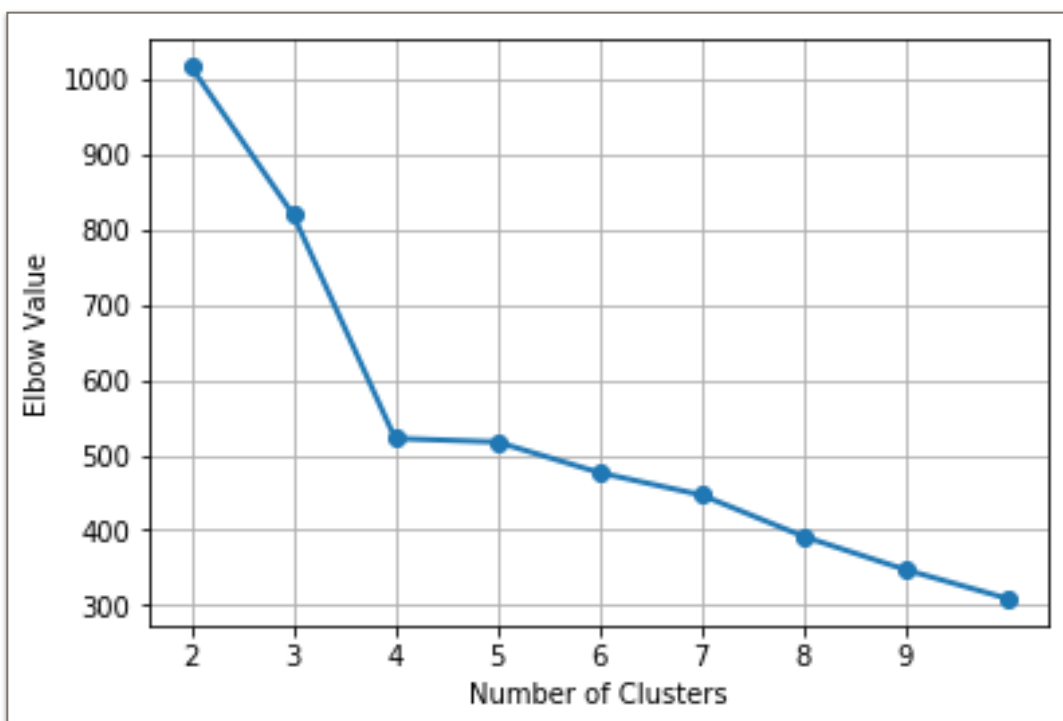
The top two principal components should be selected.

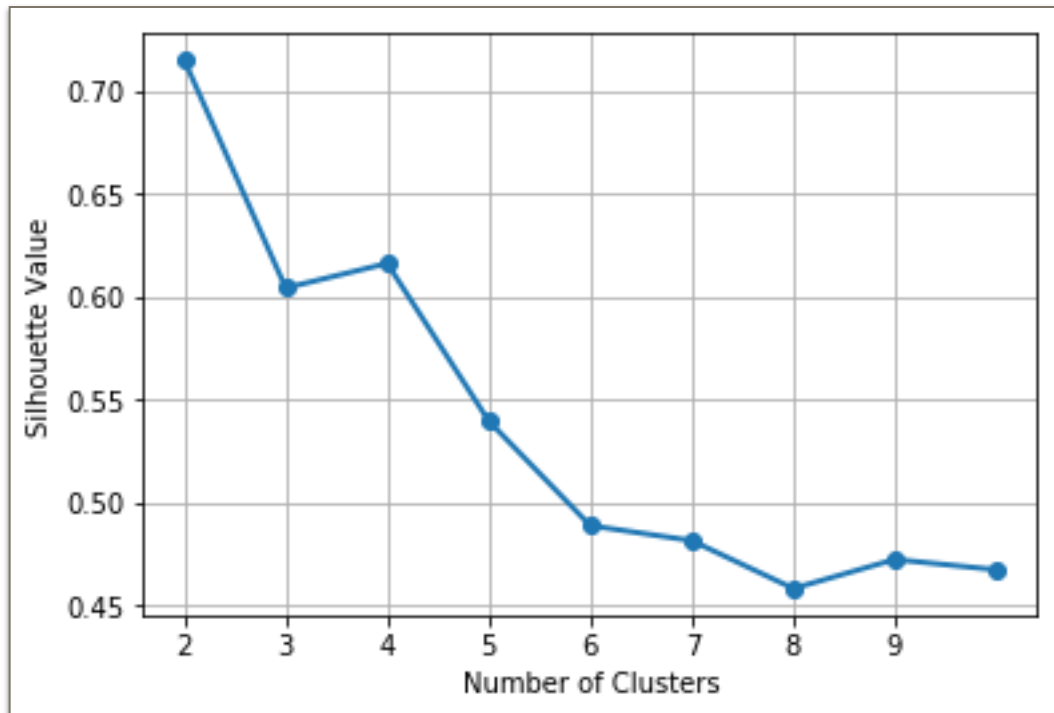


- d) (5 points). What is the cumulative explained variance ratio accounted by the major principal components that you selected in c)?

The cumulative explained variance ratio accounted by the first two principal components is 0.9669018069364491.

- e) (5 points). Plot the Elbow and the Silhouette charts against the number of clusters.





f) (5 points). What is the number of clusters that you will choose based on the charts in e)?

4 clusters will be chosen based on the charts in e)

g) (5 points). List the names of the communities in each cluster.

The names of the communities in cluster0: ['Downtown', 'West Loop', 'Near North Side', 'Lake View', 'Lincoln Park', 'Avondale', 'Albany Park', 'Jefferson Park', 'Edison Park', 'Archer Heights', 'Dunning', 'Portage Park', 'West Ridge', 'Edgebrook', 'Mount Greenwood', 'Norwood Park', 'Belmont Harbor', 'North Park']

The names of the communities in cluster1: ['Chatham', 'Auburn Gresham', 'Englewood', 'West Garfield Park', 'Roseland', 'West Englewood', 'Austin', 'South Shore', 'Kenwood']

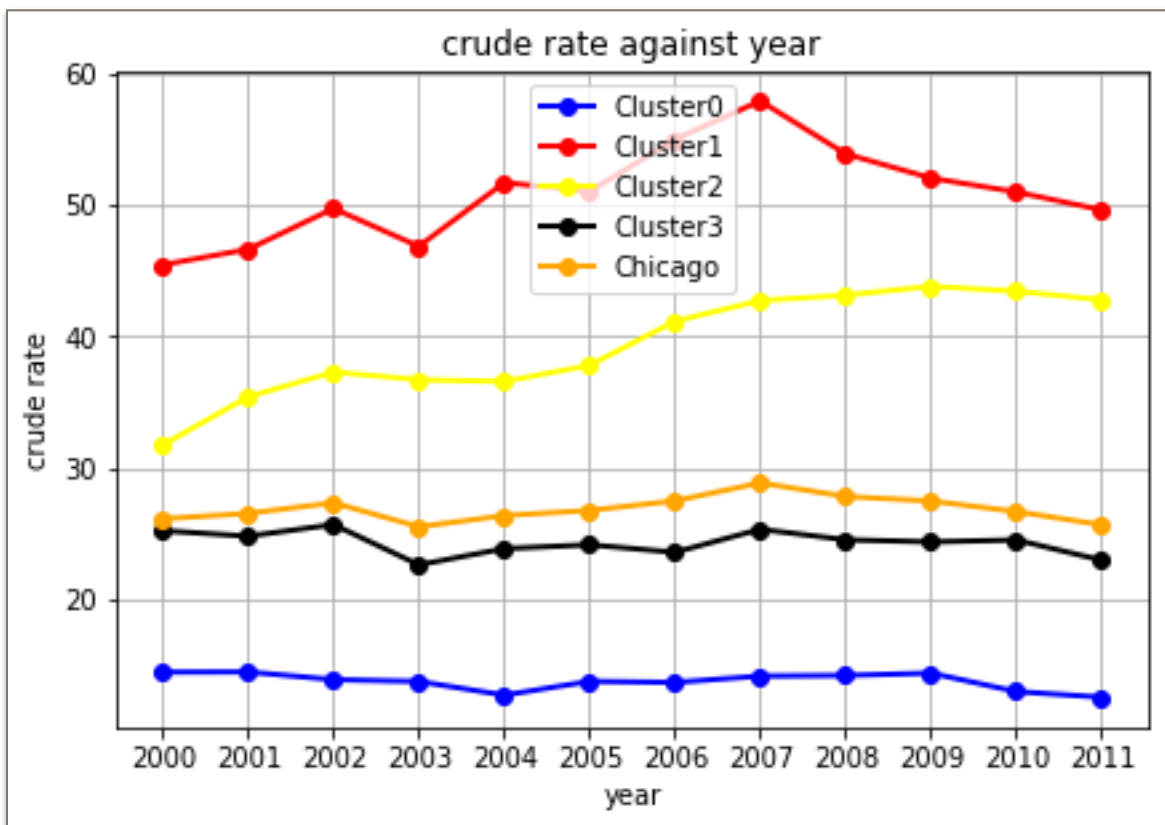
The names of the communities in cluster2: ['Near West Side', 'South Chicago', 'Woodlawn', 'Beverly', 'West Humboldt Park']

The names of the communities in cluster3: ['Lower West Side', 'New City', 'Hyde Park', 'Chinatown', 'West Town', 'South Lawndale', 'Rogers Park', 'West Lawn', 'Garfield Ridge', 'Belmont Gardens', 'Edgewater', 'Bucktown', 'Ashburn', 'Edgewater Glen']

h) (5 points). What are Chicago's annual crude hospitalization rates from 2000 to 2011? Please present your answer in a table.

Index	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Chicago's annual crude hospitalization rate	26.1468	26.5527	27.3606	25.5392	26.389	26.7807	27.5083	28.8725	27.8378	27.4936	26.693	25.6865

i) (5 points). Plot the crude hospitalization rates in each cluster against the years. You also plot the Chicago's annual crude hospitalization rates (in your answer in h) against the years as the reference curve.



- j) (5 points) Based on the graph in i), what will you conclude about the trend of crude hospitalization rate in each cluster relative to the Chicago's rates?

The four clusters have similar trend to the Chicago's rates trend. When Chicago's rates rise, the trends of the four clusters rise to varying degrees. When the Chicago's rates decline, the trends of the four clusters decline to varying degrees. Based on this graph, we can see that the trends in Clusters 1 and 2 have a stronger response to changes in the Chicago's rate.

Question2:

- a) (15 points). If we restrict the values of the parameters w_1 , w_2 , and c to positive integers, then specify the lowest possible values for these parameters such that the perceptron can implement the logical OR function. You have to prove that your solution does work.

The lowest possible values for these parameters:

$$w_1 = 1, w_2 = 1, c = 1$$

x1	x2	$w_1 \cdot x_1 + w_2 \cdot x_2$	$\geq c?$	y
0	0	0	no	0
1	0	1	yes	1
0	1	1	yes	1
1	1	2	yes	1

- b) (15 points). If we restrict the values of the parameters w_1 , w_2 , and c to positive integers, then specify the lowest possible values for these parameters such that the perceptron can implement the logical AND function which can be represented by the following table. You have to prove that your solution does work.

$$w_1 = 1, w_2 = 1, c = 2$$

x1	x2	$w_1 \cdot x_1 + w_2 \cdot x_2$	$\geq c?$	y
0	0	0	no	0
1	0	1	no	0
0	1	1	no	0
1	1	2	yes	1

- c) (20 points). The logical XAND function (i.e., the Exclusive AND) returns TRUE only when both arguments are the same (e.g., both TRUE or both FALSE). Otherwise, it returns FALSE. Specify the six synaptic weights and the three threshold values such that the above neural network can implement the XAND function. The parameters are still integers but we allow negative integers. You have to prove that your solution does work.

$$w_{11} = -1, w_{21} = 1, c_1 = 0$$

$$w_{12} = 1, w_{22} = -1, c_2 = 0$$

$$w_{31} = 1, w_{32} = 1, c_3 = 2$$

x1	x2	$w_{11}x_1 + w_{21}x_2$	$\geq c_1?$	h1
0	0	0	yes	1
1	0	-1	no	0
0	1	1	yes	1
1	1	0	yes	1

x1	x2	$w_{12}x_1 + w_{22}x_2$	$\geq c_2?$	h2
0	0	0	yes	1
1	0	1	yes	1
0	1	-1	no	0
1	1	0	yes	1

h1	h2	$w_{31}h_1 + w_{32}h_2$	$\geq c_3?$	y
1	1	2	yes	1
0	1	1	no	0
1	0	1	no	0
1	1	2	yes	1