

Kelly Cheng
k.cheng@gatech.edu

2.3 kMeans Clustering

a.

k=	All Features kMeans	Filtered Features kMeans
2	0.54000	0.35943
3	0.48373	0.56847
4	0.55992	0.37668
5	0.60710	0.59848

b.

k=3 for all features	Cluster 1	Cluster 2	Cluster 3
Percentage Case	100	0	0
Percentage Control	97.78	0	2.22
Percentage Unknown	96.6	3.34	0

k=3 for filtered features	Cluster 1	Cluster 2	Cluster 3
Percentage Case	100	0	0
Percentage Control	100	0	0
Percentage Unknown	67.49	31.08	1.44

2.4 Gaussian Clustering

a.

k=	All Features Gaussian	Filtered Features Gaussian
2	0.47831	0.43084
3	0.47831	0.82718
4	0.48048	0.85236
5	0.54420	0.86306

b.

k=3 for all features	Cluster 1	Cluster 2	Cluster 3
Percentage Case	95.29	4.71	0
Percentage Control	96.2	3.8	0
Percentage Unknown	84.69	13.21	2.10

k=3 for filtered features	Cluster 1	Cluster 2	Cluster 3
Percentage Case	1.95	81.86	16.19
Percentage Control	98.84	1.16	0

k=3 for filtered features	Cluster 1	Cluster 2	Cluster 3
Percentage Unknown	28.1	67.9	0

3.0 NMF

a.

k=	All Features NMF	Filtered Features NMF
2	0.47831	0.45088
3	0.49268	0.47531
4	0.49349	0.53984
5	0.49438	0.55125

b.

k=3 for all features	Cluster 1	Cluster 2	Cluster 3
Percentage Case	2.15	2.25	95.59
Percentage Control	8.33	6.43	85.23
Percentage Unknown	2.15	2.78	95.07

k=3 for filtered features	Cluster 1	Cluster 2	Cluster 3
Percentage Case	0.01	99.08	0
Percentage Control	0	0	100
Percentage Unknown	53.54	11.49	34.97