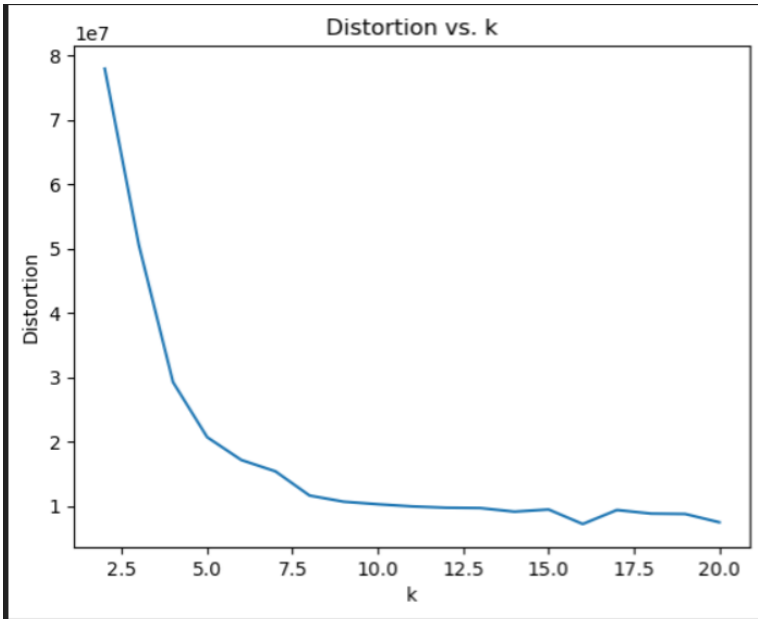
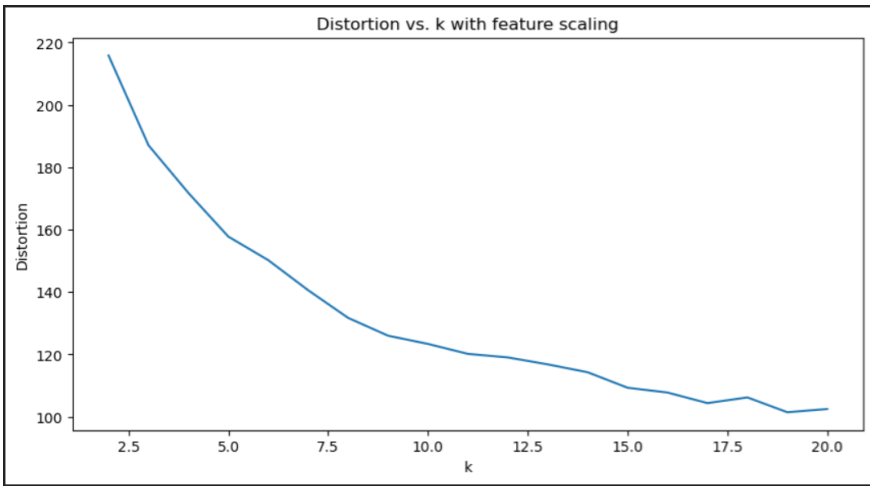


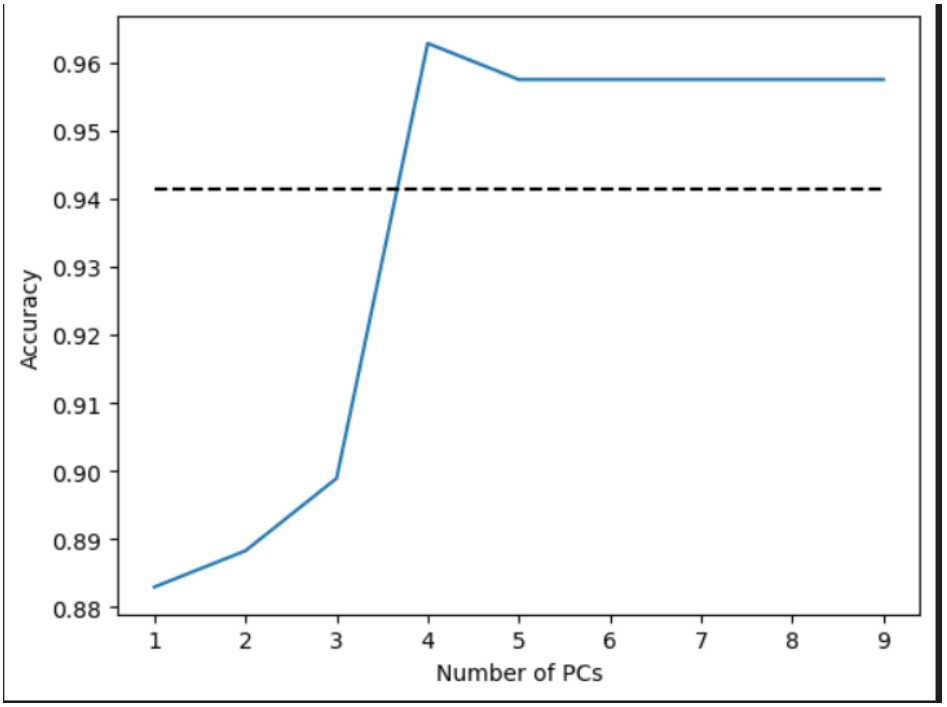
Q3.4: we choose the elbow position, thus 7 or 8



Q3.5

Q3.6:

1. without scaling, the euclidean distance would be larger.
2. yes, we literally downweight some abnormal large distance and make the data more normalized

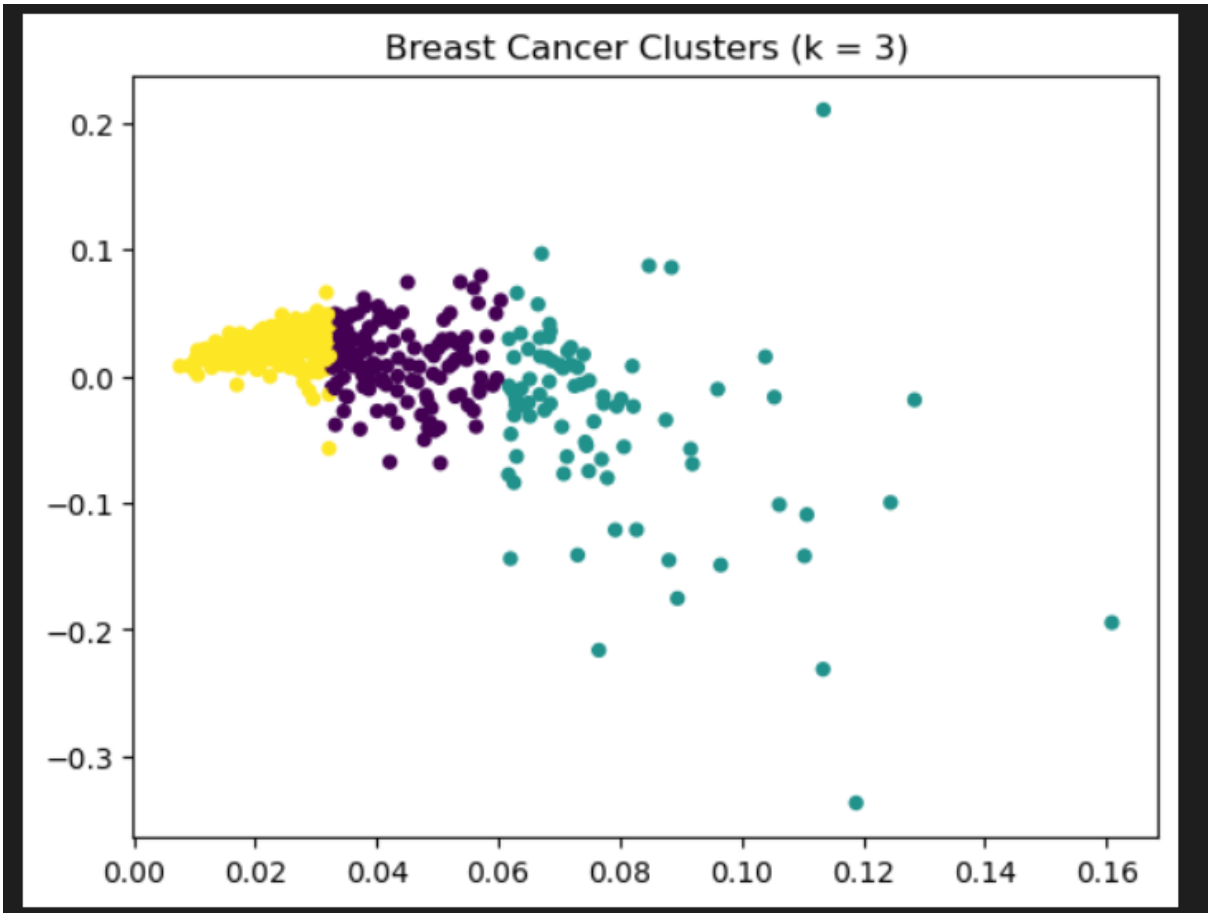


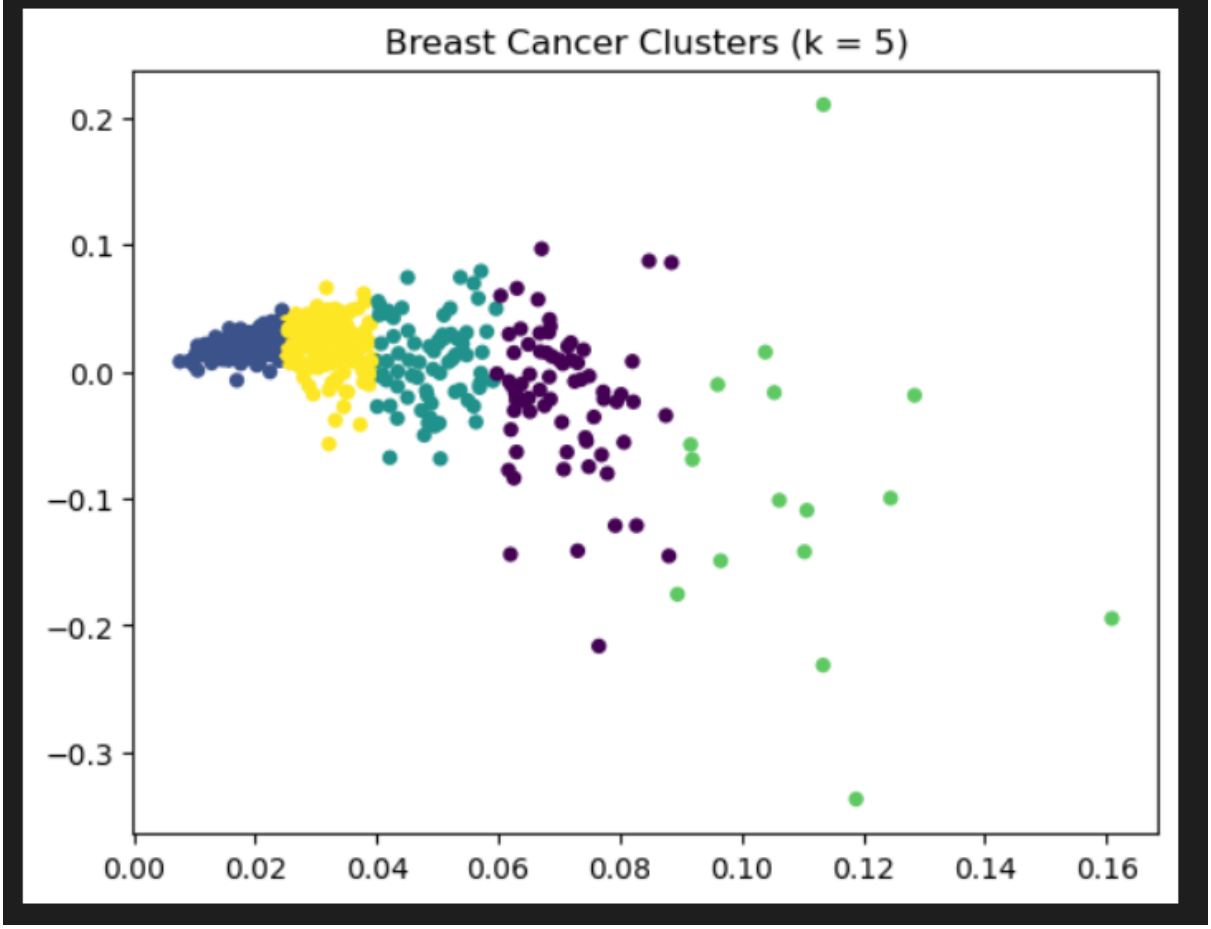
Q4.2: n_component = 4 would be best. less PC may fail to capture important information while more PC may be overfit and induce noise.

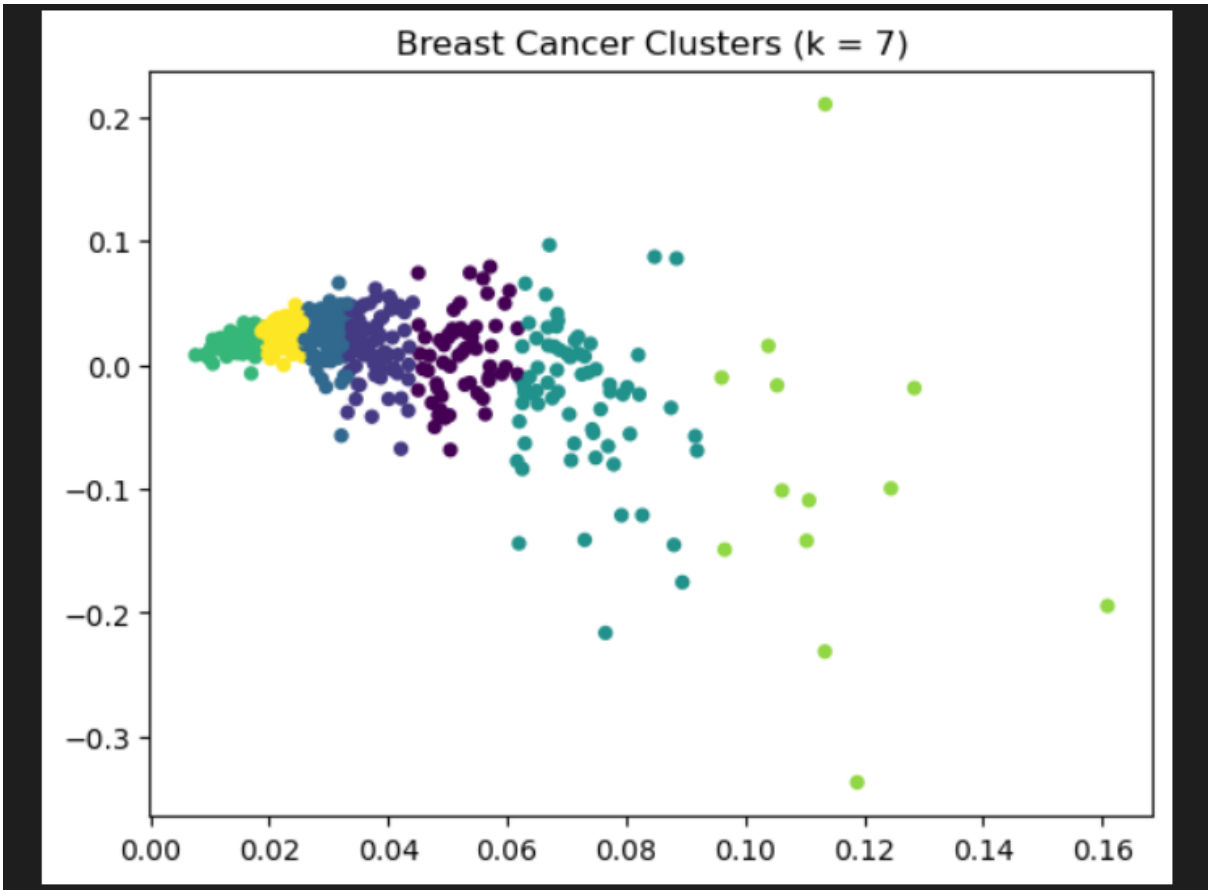
Q4.2.2

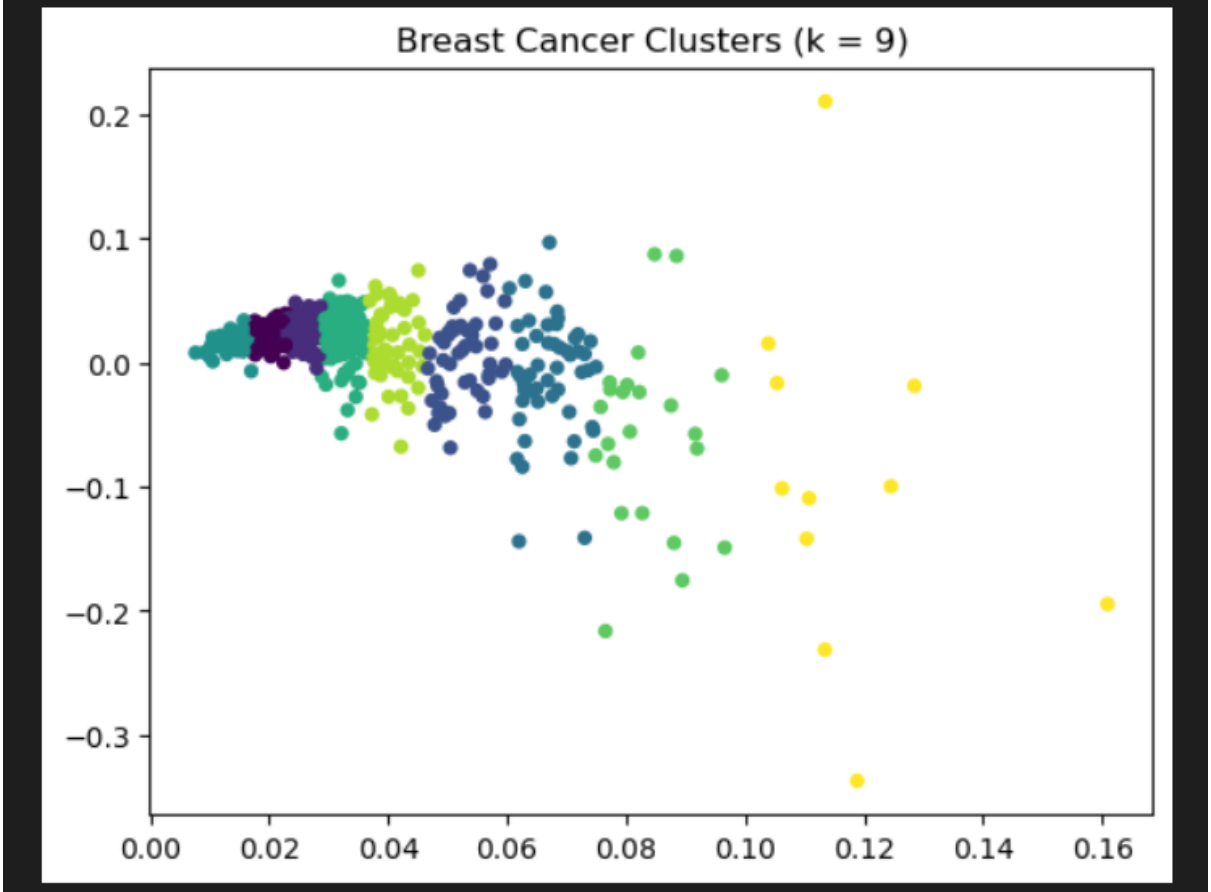
```
Top 3 features contributing to PC1: [(0.852063391798144, 'worst area'),
(0.5168264687224632, 'mean area'), (0.05572716691107061, 'area error')]
Top 3 features contributing to PC2: [(0.8518237204834174, 'mean area'),
(0.0627480827489321, 'mean perimeter'), (0.00928705649723585, 'mean
radius')]
Top 3 features contributing to PC3: [(0.9902458782833069, 'area
error'), (0.04385603691150635, 'perimeter error'),
(0.006233776347976855, 'texture error')]
Top 3 features contributing to PC4: [(0.039412249355513615, 'mean
area'), (0.03876915240528661, 'worst area'), (6.6911461943283174e-06,
'smoothness error')]
```

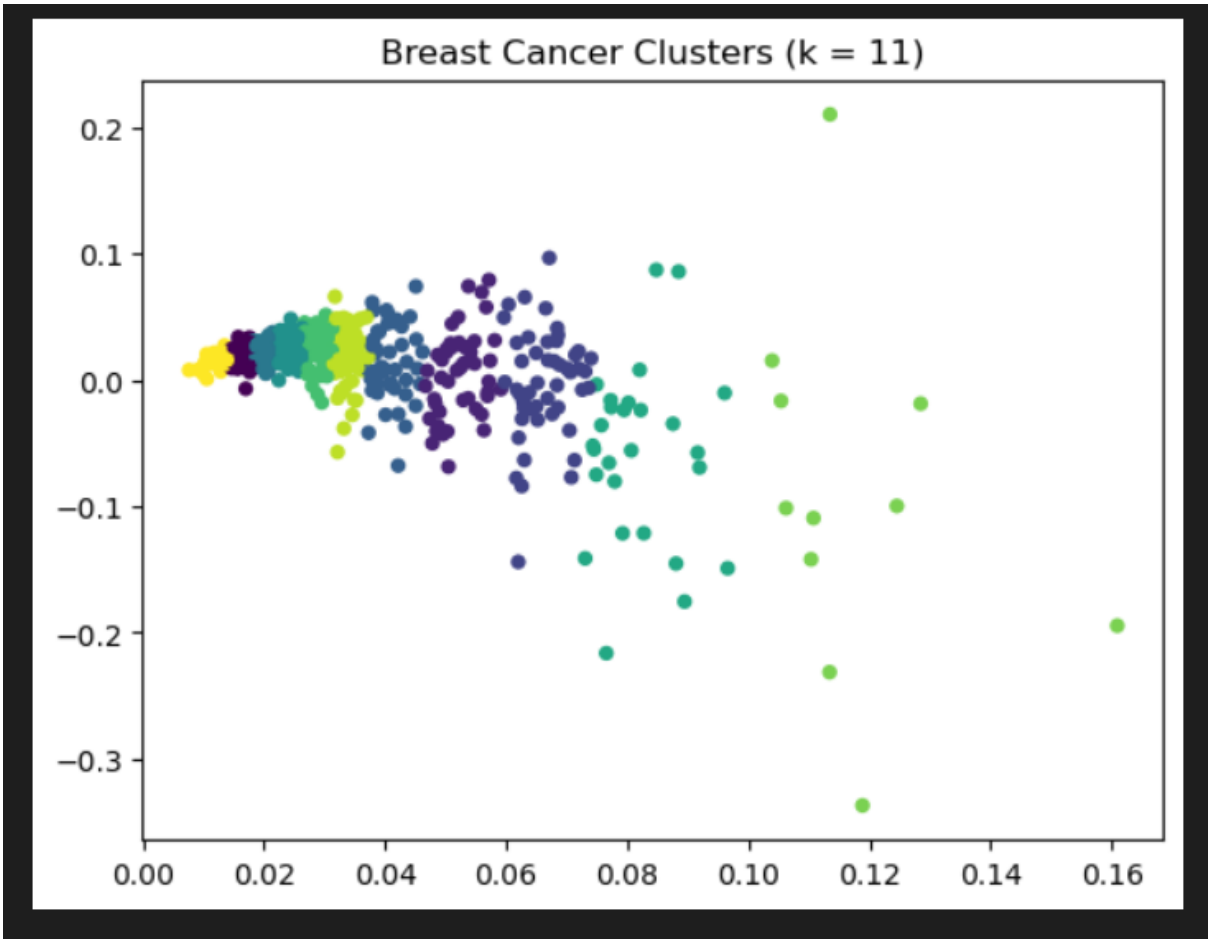
Q4.3


Breast Cancer Clusters (k = 3)
0.2
0.1
0.0
−0.1
−0.2
−0.3
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.16


Breast Cancer Clusters (k = 5)
0.2
0.1
0.0
−0.1
−0.2
−0.3
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.16


Breast Cancer Clusters (k = 7)
0.2
0.1
0.0
−0.1
−0.2
−0.3
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.16


Breast Cancer Clusters (k = 9)
0.2
0.1
0.0
−0.1
−0.2
−0.3
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.16


Breast Cancer Clusters (k = 11)
0.2
0.1
0.0
−0.1
−0.2
−0.3
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.16