Home Work 4

1. (a) False. Random Forese average the Var
So it has $maller var

(b) False. More weak classifier makes Adaboost focus more efficiently.

(c) False. Variance is lowered, not bias

2 (a) $\epsilon_t = \sum_{i: y_i \neq h_t(x_i)} w_{t,i}$ $\quad 1 - \epsilon_t = \sum_{i: y_i = h_t(x_i)} w_{t,i}$

given binary classifier (weak), we get -1 or 1 from sign func

thus: if $y_i = h_t(x_i)$

$$w_{t+1} = w_t e^{-\beta} / Z_t$$

else $y_i \neq h_t(x_i)$

$$w_{t+1} = w_t e^{\beta} / \Sigma_t$$

$$\Sigma_t = \sum_{i=1}^{N} w_{t,i} \times e^{-\beta_t y_i h_t(x_i)}$$

$$= \sum_{y_i = h_t(x_i)}^{N} w_{t,i} e^{-\beta_t} + \sum_{y_i \neq h_t(x_i)}^{N} w_{t,i} e^{\beta_t}$$

$$= e^{-\beta_t}(1 - \epsilon_t) + e^{\beta_t} \epsilon_t$$

(b) $\frac{\partial \Sigma}{\partial \beta} = \epsilon_t e^{\beta} - (1 - \epsilon_t) e^{-\beta} = 0$

$$\epsilon_t e^{2\beta} = (1 - \epsilon_t)$$

$$e^{2\beta} = \frac{1 - \epsilon_t}{\epsilon_t}$$

$$\beta = \frac{1}{2} \ln\left(\frac{1 - \epsilon_t}{\epsilon_t}\right)$$

(c) β is changed for each iteration to minimize the average exponential loss.

While the function is convex, the min point sit at where the gradient equals zero

3. (a) Calculate the mean of each x_i and center the matrix

$$\bar{x}_1 = \frac{4+2+5+1}{4} = \frac{12}{4} = 3$$

$$\bar{x}_2 = \frac{1+3+4+0}{4} = 2$$

$$X_C = \begin{bmatrix} 4-3 & 1-2 \\ 2-3 & 3-2 \\ 5-3 & 4-2 \\ 1-3 & 0-2 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & +1 \\ +2 & 2 \\ -2 & -2 \end{bmatrix}$$

get the Cov Matrix:

$$Cov = X_C^T \cdot X = \begin{bmatrix} 1 & -1 & +2 & -2 \\ -1 & +1 & 2 & -2 \end{bmatrix} \cdot \begin{bmatrix} 1 & -1 \\ -1 & +1 \\ +2 & 2 \\ -2 & -2 \end{bmatrix}$$

$$= \begin{bmatrix} 10 & 6 \\ 6 & 10 \end{bmatrix}$$

~~$= \begin{bmatrix} 12 & 0 \\ 0 & 12 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$~~

$$= \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix}$$

$$= \begin{bmatrix} 10 & 2 \\ -2 & 10 \end{bmatrix} = \begin{bmatrix} 5 & -1 \\ -1 & 5 \end{bmatrix}$$

$\det(C-\lambda I)v = 0$ ~~$\det(Cov - \lambda I)V = 0$~~

$(5-\lambda)^2 = 3^2$ ~~$(5-\lambda)^2 = 1^2$~~

$\lambda = 2$ or 8 ~~$\lambda = 3$ or 7~~

~~for $\lambda = 7$:~~

~~$\begin{bmatrix} -2 & -1 \\ \cdot & \cdot \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$~~

$-2x_1 - x_2 = 0$

$-x_1 - 2x_2 = 0$

$\lambda = 8$

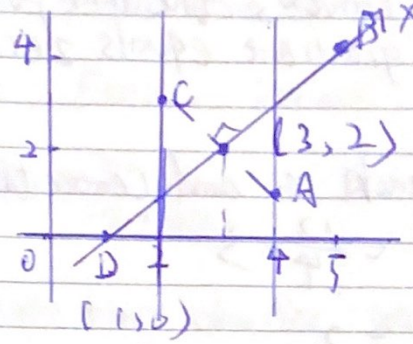
$$\begin{bmatrix} -3 & 3 \\ 3 & -3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0$$

$0 = y$

3 (a) we use $\lambda = 8$ since it is bigger $\Rightarrow$ high variation

we get $v = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$

(b)

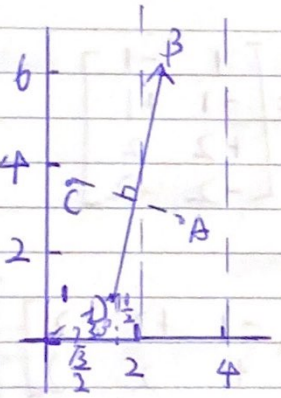


A C Project to 0 (center)

D project to $-2\sqrt{2}$

B $\rightarrow 2\sqrt{2}$

(c)



A C $\sim$ 0

D $\sim -2\sqrt{2}$

B $\sim 2\sqrt{2}$

$x_{new} = \frac{1}{2} x$

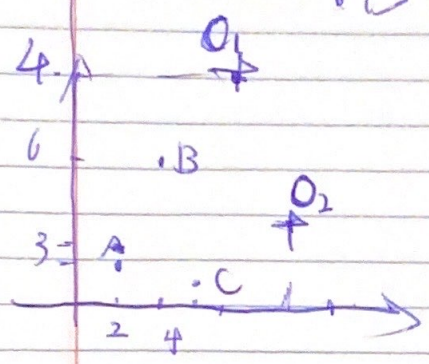
$y_{new} = \frac{\sqrt{3}}{2} y$

new $(x_B - x_D)^2 + (y_B - y_D)^2$

$= \frac{1}{4}(x_B - x_D)^2 + \frac{3}{4}(y_B - y_D)^2$

$= \frac{1}{4} \times 4^2 + \frac{3}{4} \times 4^2 = 2 \times 4^2$

$=$ old length



$k=1$

$O_1 = (6,9)$ $O_2 = (8,4)$

$d(A,1) = 7.2$ $d(A,2) = 6.1$

$d(B,1) = 3.6$ $d(B,2) = 4.8$

$d(C,1) = 8.1$ $d(C,2) = 4.2$

$d(D,1) = 5$ $d(B,2) = 8.2$

$O_1 \{B, D\}$ $O_2 \{A, C\}$

$O_1 = (B+D)/2$ $O_2 = (A+C)/2$

$= (7,9)$ $= (3.5, 2)$

$k=2$

$d(A,1) = 7.8$ $d(A,2) = 1.8$

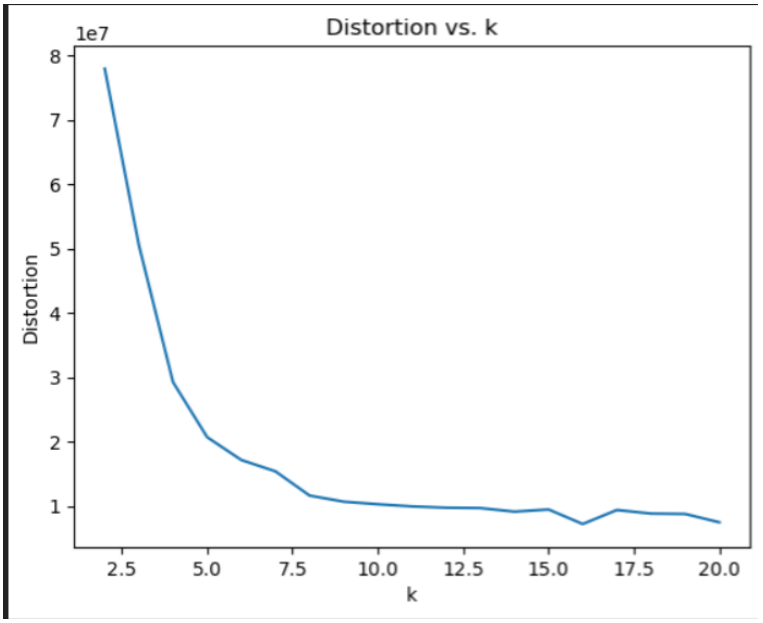
$d(B,1) = 4.2$ $d(B,2) = 4$

$d(C,1) = 8.2$ $d(C,2) = 1.8$

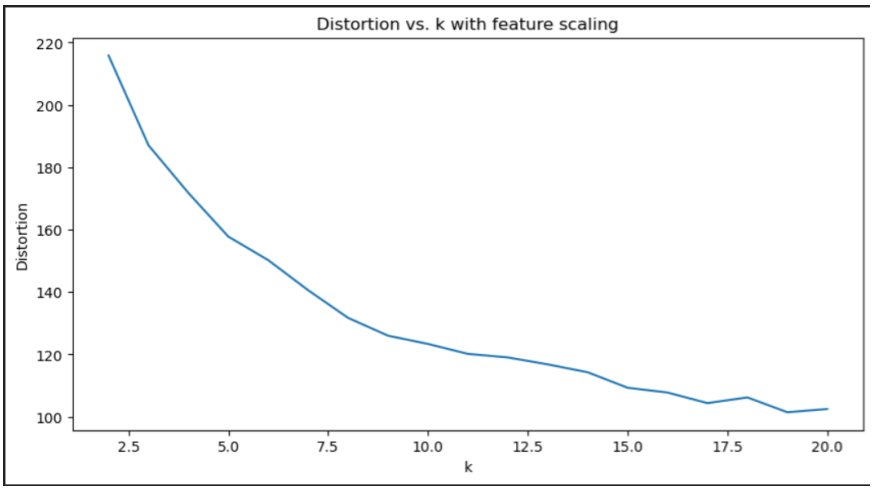
$d(D,1) = 4.2$ $d(D,2) = 11.9$

$O_1 \{D\}$ $O_2 \{A, B, C\}$

$O_1 = (10, 12)$ $O_2 = (3.6, 3.3)$



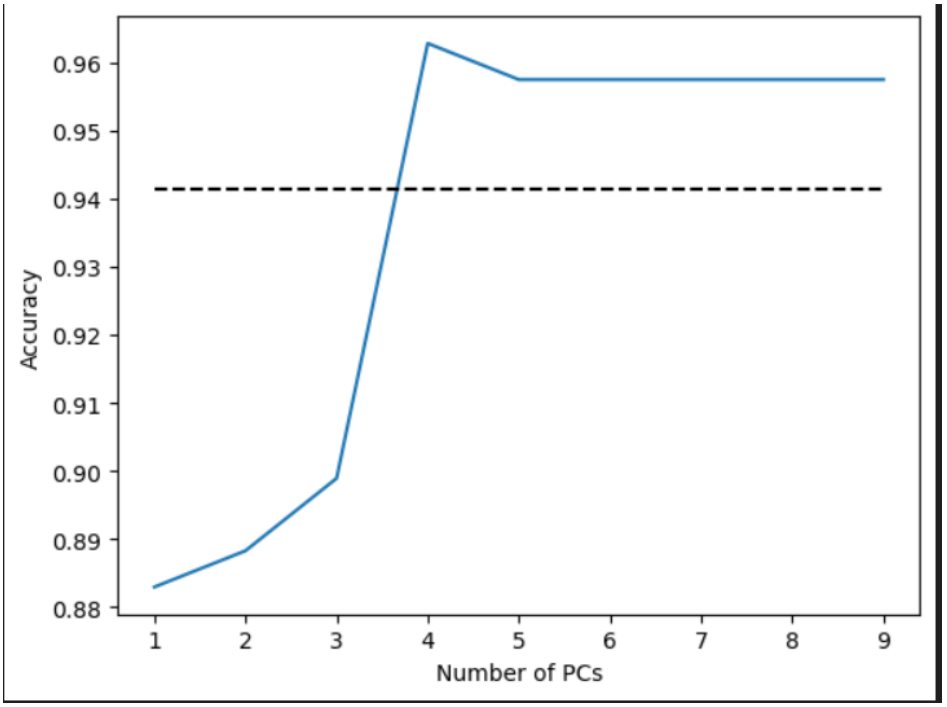
Q3.4: we choose the elbow position, thus 7 or 8



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

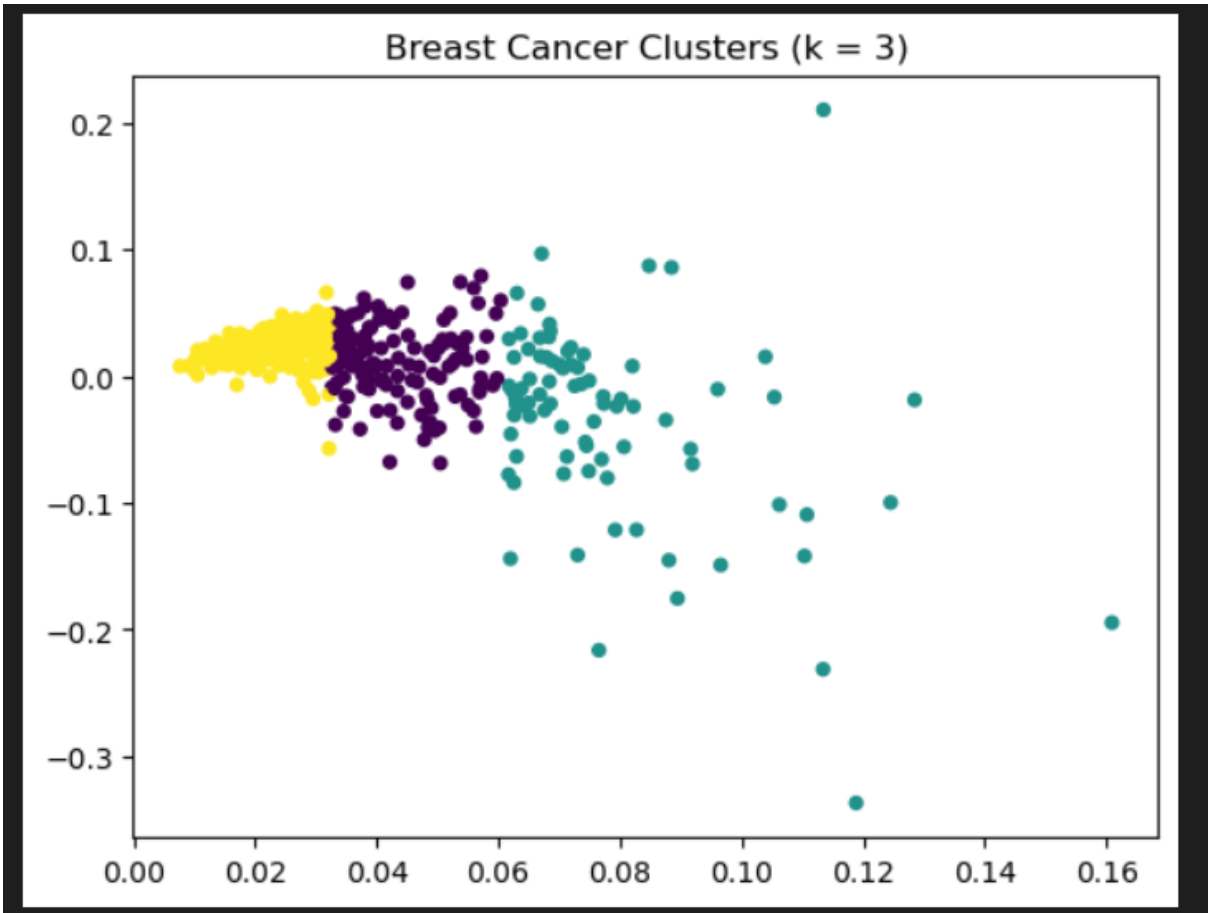


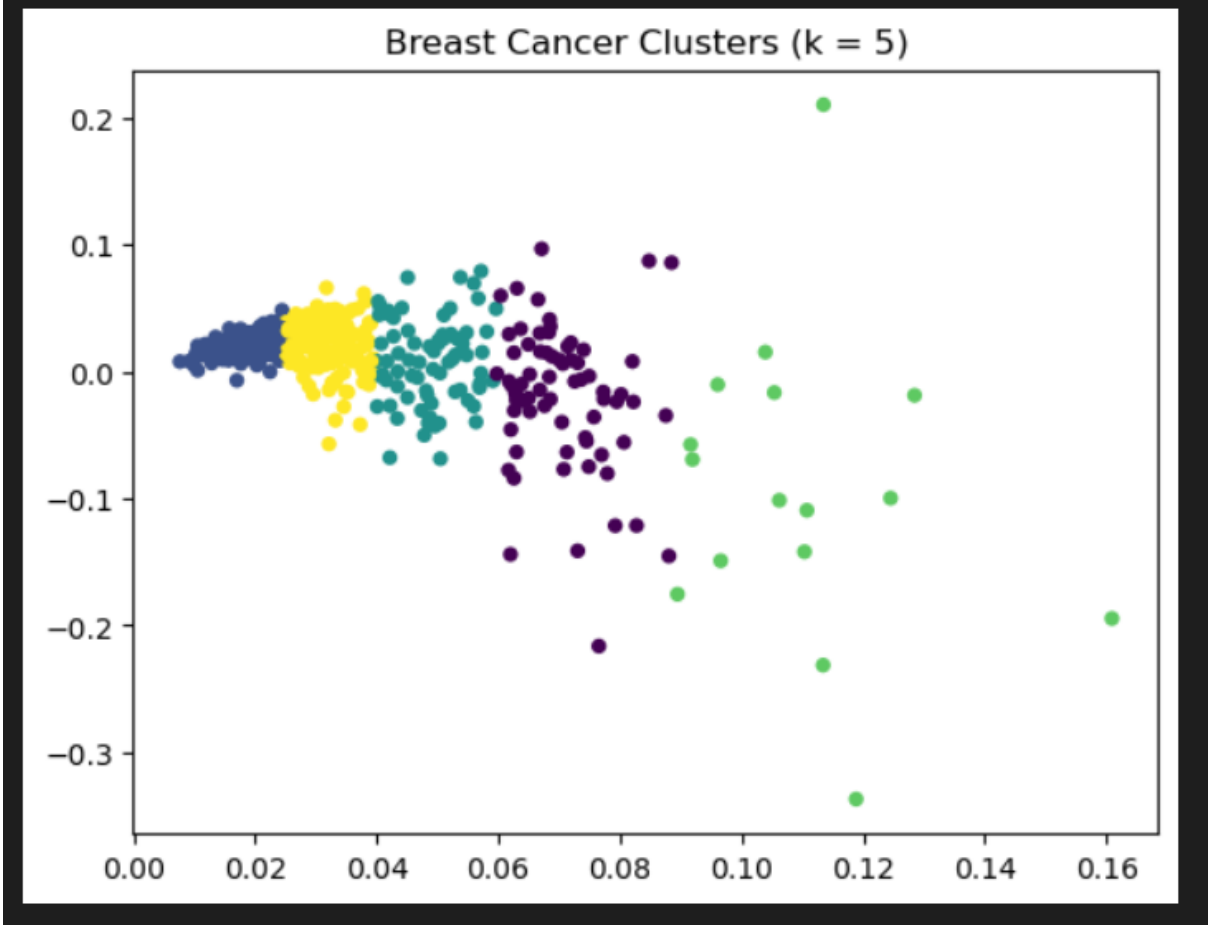
Q4.2: n_component = 4 would be best. less PC may fail to capture important information while more PC may be overfit and induce noise.

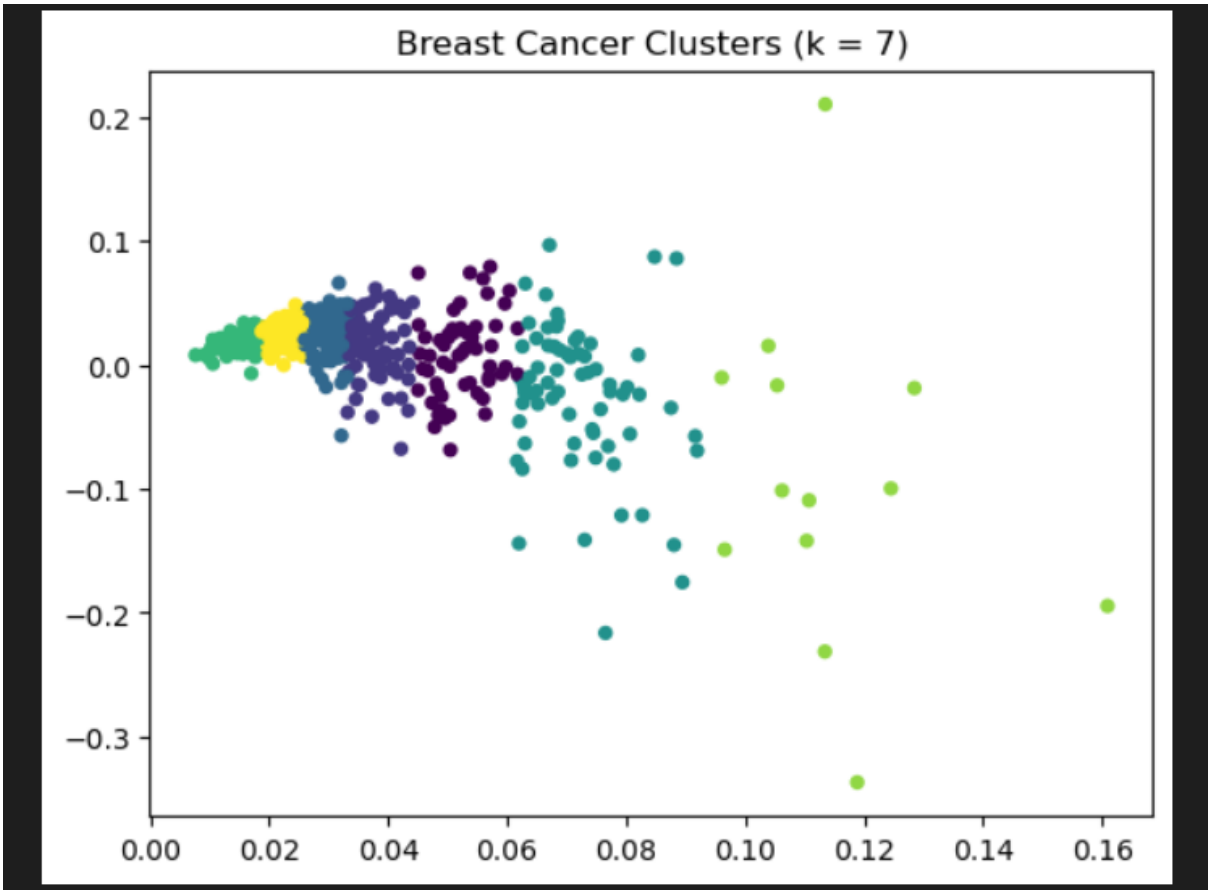
Q4.2.2

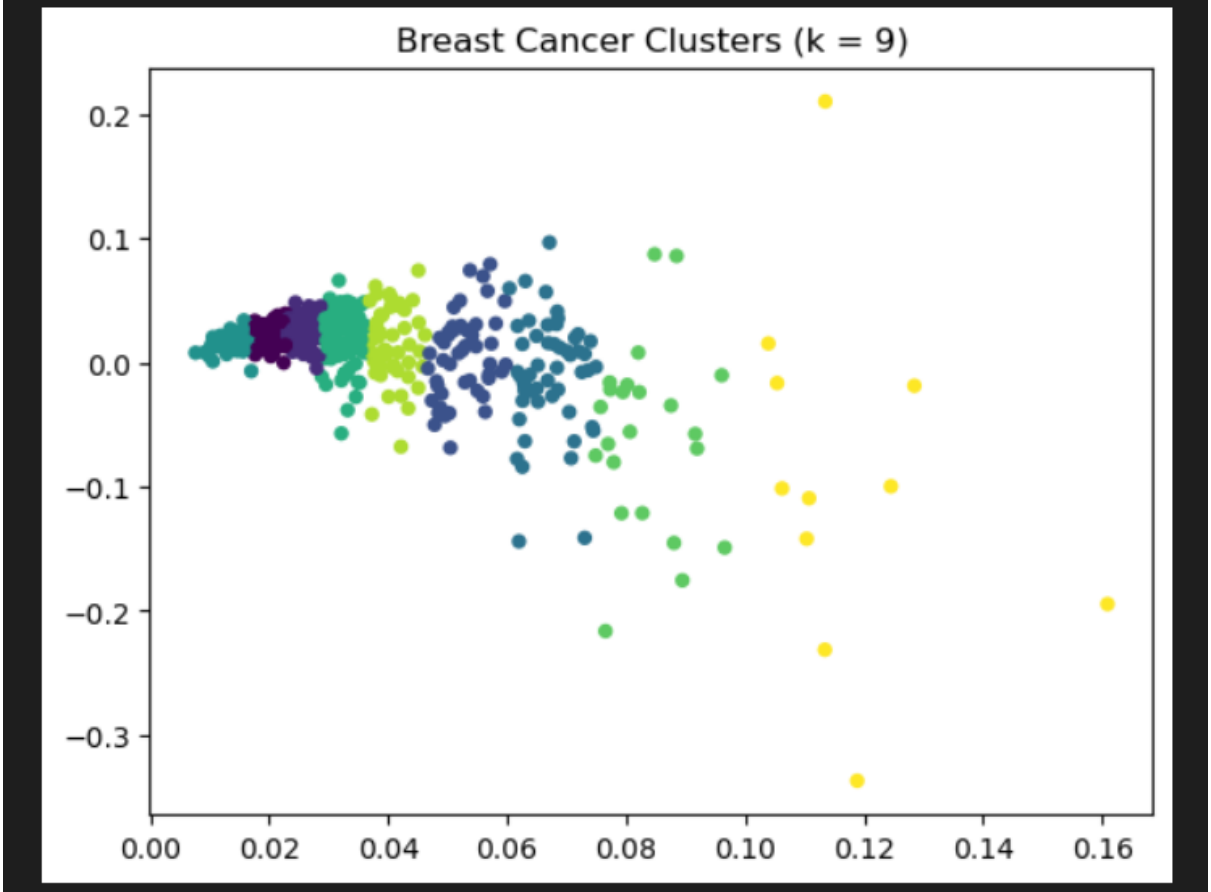
```
Top 3 features contributing to PC1: [(0.852063391798144, 'worst area'),
(0.5168264687224632, 'mean area'), (0.05572716691107061, 'area error')]
Top 3 features contributing to PC2: [(0.8518237204834174, 'mean area'),
(0.0627480827489321, 'mean perimeter'), (0.00928705649723585, 'mean
radius')]
Top 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')]
```

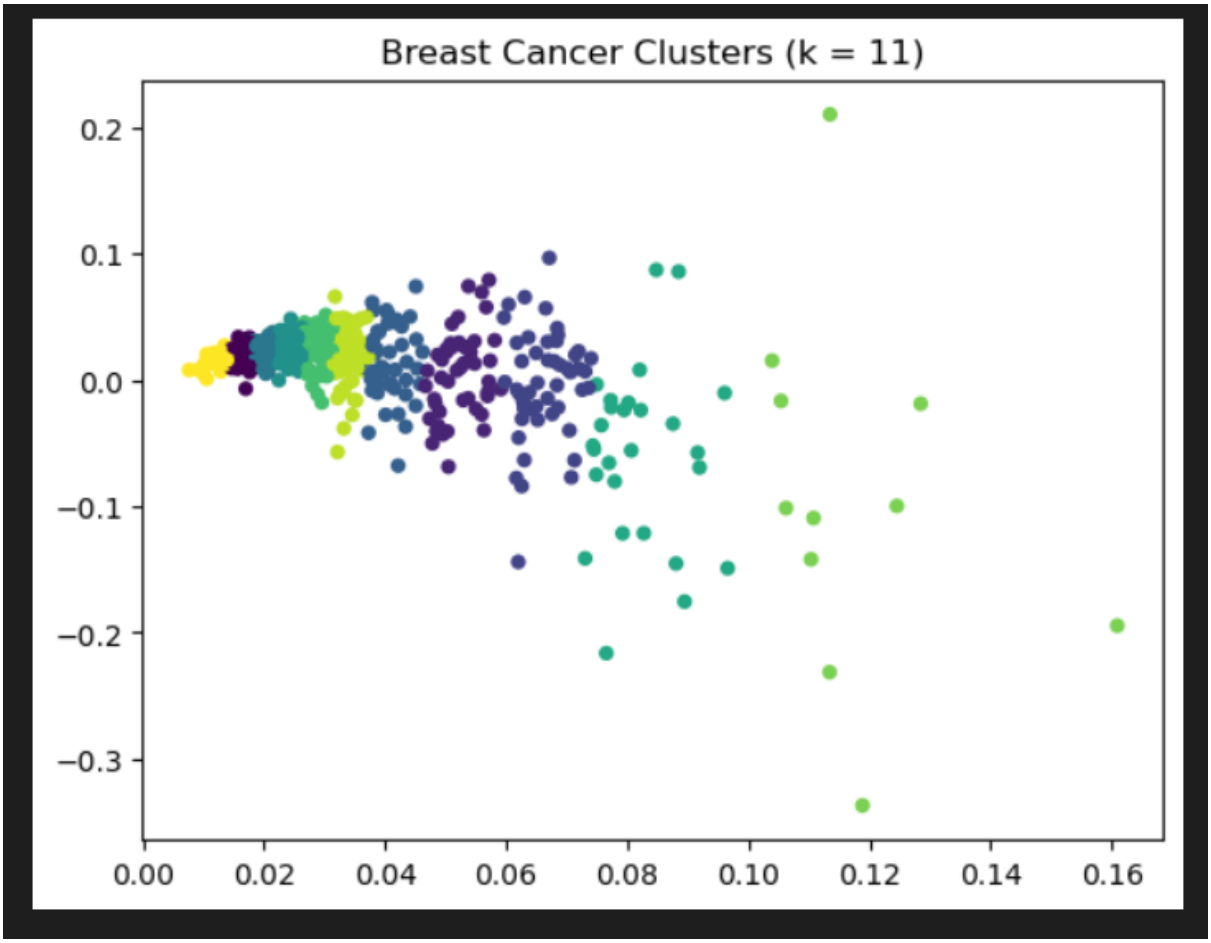
Q4.3


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