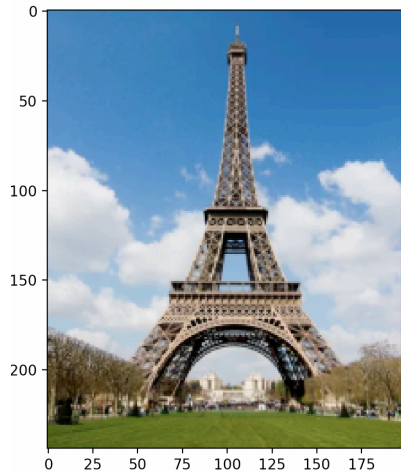


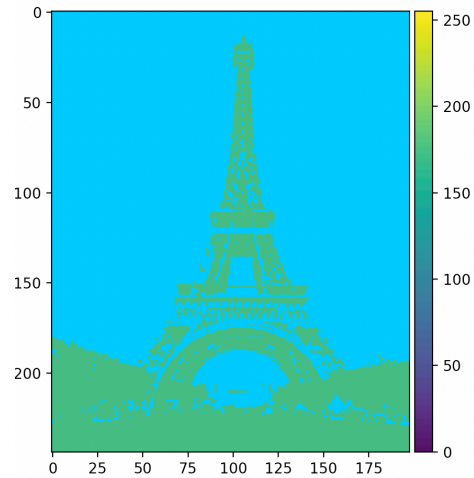
Homework 4

Question 1) (Code)

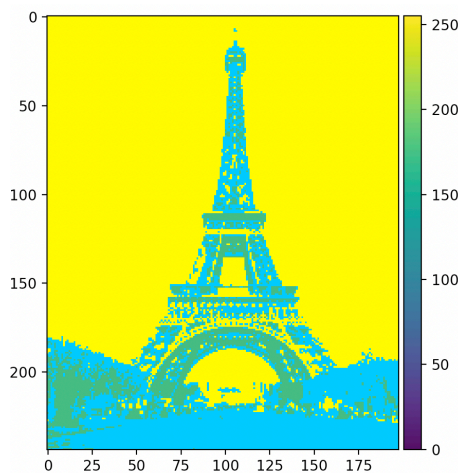
Original Image:



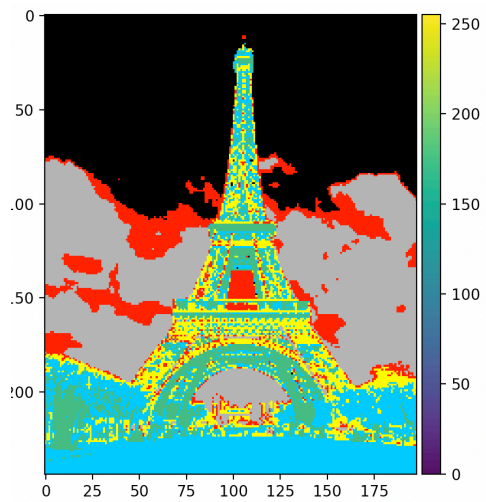
k = 2



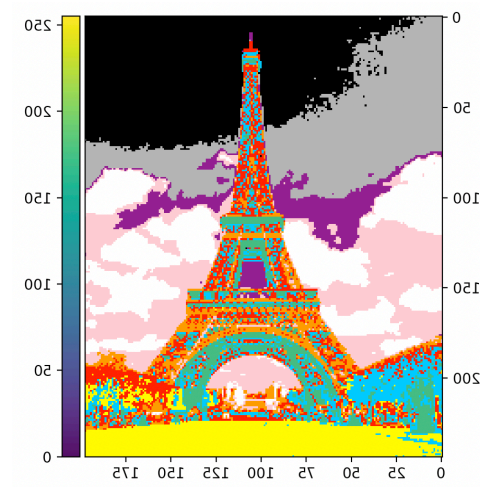
k = 3

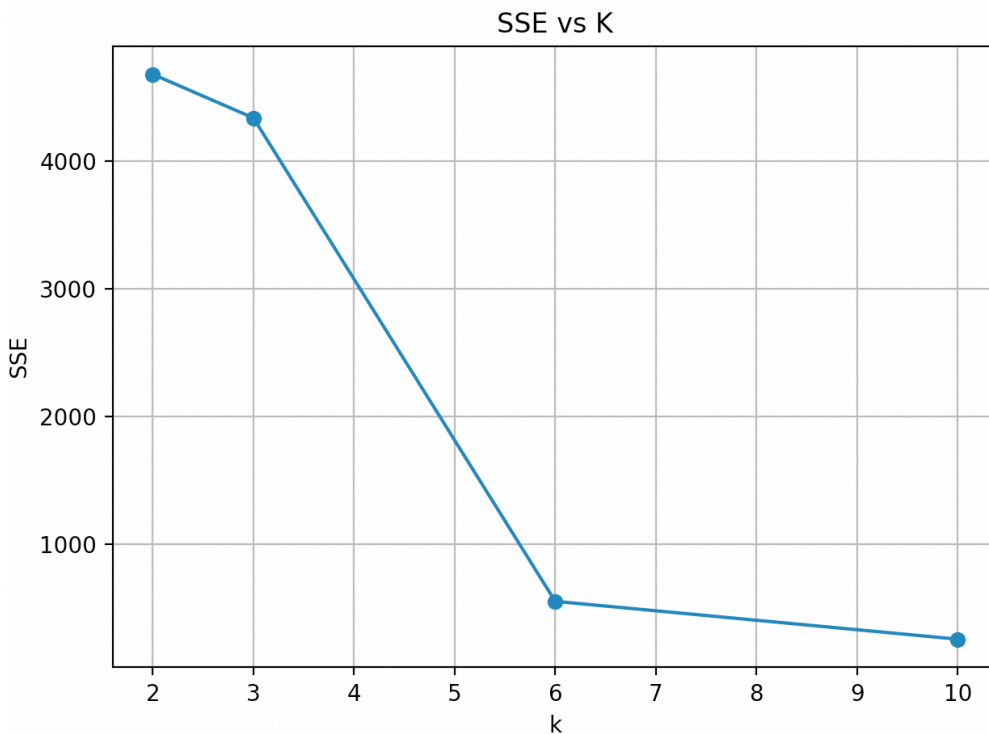


k = 6



K = 10





The centroids for each k and number of iterations needed are in code when you run python3 q1.py

```
k = 2 SSE = 4678.113042172468
k = 3 SSE = 4336.465222032642
k = 6 SSE = 550.1736259106633
k = 10 SSE = 253.84711145254835
```

Question 2)

$C1 = \{(1, 1), (2, 2), (3, 3)\}$, $C2 = \{(5, 2), (6, 2), (7, 2), (8, 2), (9, 2)\}$

(a) Mean vectors m_1 and m_2 :

$$M_1 = (1/3) * [(1, 1) + (2, 2) + (3, 3)] = (1/3) * (6, 6) = (2, 2)$$

$$M_2 = (1/5) * [(5, 2) + (6, 2) + (7, 2) + (8, 2) + (9, 2)] = (1/5) * (35, 10) = (7, 2)$$

(b) Total mean vector m :

$$M = (1/8) * [(6, 6) + (35, 10)] = (1/8) * (41, 16) = (5.125, 2)$$

(c) The scatter matrices S_1, S_2 :

$$\begin{aligned} S_1 &= ((1,1) - (2,2)) * ((1,1) - (2,2))^T + ((2,2) - (2,2)) * ((2,2) - (2,2))^T + ((3,3) - (2,2)) * ((3,3) - (2,2))^T \\ &= \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} \end{aligned}$$

$$\begin{aligned} S_2 &= ((5,2) - (7,2)) * ((5,2) - (7,2))^T + ((6,2) - (7,2)) * ((6,2) - (7,2))^T + ((7,2) - (7,2)) * ((7,2) - (7,2))^T + \\ &((8,2) - (7,2)) * ((8,2) - (7,2))^T + ((9,2) - (7,2)) * ((9,2) - (7,2))^T = \\ &= \begin{pmatrix} 4 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 4 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 10 & 0 \\ 0 & 0 \end{pmatrix} \end{aligned}$$

(d) The within-cluster scatter matrix S_W :

$$S_W = S_1 + S_2$$

$$\begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} + \begin{pmatrix} 10 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 12 & 2 \\ 2 & 2 \end{pmatrix}$$

(e) The between-cluster scatter matrix S_B :

$$(3 * [(2, 2) - (5.125, 2)] * [(2, 2) - (5.125, 2)]^T) + (5 * [(7, 2) - (5.125, 2)] * [(7, 2) - (5.125, 2)]^T) =$$
$$\begin{pmatrix} 29.297 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 17.578 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 46.875 & 0 \\ 0 & 0 \end{pmatrix}$$

(f) The scatter criterion $\text{tr}(S_B)/\text{tr}(S_W)$:

$$46.875 / 14 = 3.348$$

Question 3) (Code)

$$\sum_{k > n./2}^n \binom{n}{k} p^k (1-p)^{n-k}$$

- a) $(0.4^3) + 3(0.6)(0.4^2) = 0.352$ which is 35.2%
- b) $(0.4^5) + 5(0.6)(0.4^4) + 10(0.6^2)(0.4^3) = 0.31744$ which is approximately 31.74%
- c) For all 25 models answer is approximately $0.1537 = 15.4\%$
- d) The reason the error rate is small for part c is because all models are assumed to be independent which isn't always true.
- e) For all 25 models at 0.45% accuracy the answer is approximately 0.694 or 69.4%

Question 4)

Confusion Matrix:

	Predicted: Positive	Predicted: Negative
Actual: Positive	4 = True Positive (40%)	1 = False Negative (10%)
Actual: Negative	2 = False Positive (20%)	3 = True Negative (30%)

Accuracy: $(TP + TN) / (TP + TN + FP + FN) = (4+3) / (4 + 3 + 2) = 1 = 7/10 = \mathbf{70\%}$

Precision: $TP / (TP+FP) = 4/(4+3) = 4/7 = 0.5714 = \mathbf{57.14\%}$

Recall: $TP/(TP+FN) = 4/(4+1) = 4/5 = 0.8 = \mathbf{80\%}$

F1 Score: $2 * \text{Recall} * \text{Precision} / (\text{Recall} + \text{Precision}) = 2(4/5)(4/7) / (4/5 + 4/7) = 0.7272 = \mathbf{72.72\%}$

Specificity: $TN / N = 3/(3+2) = 3/5 = 0.6 = \mathbf{60\%}$