A Comparison of Image Segmentation Methods

Marshall Grimmett

May 13, 2021

1 Part 1

1.1 a

The new objective function, where α_i is a lagrangian multiplier for each constraint, will be

$$min\ L = \frac{1}{2}w^2 - \sum_{i=1}^{n} \alpha_i (y_i(w^T \phi(x_i)) - 1)$$

1.2 b

Taking the derivative of L with respect to w we get

$$\frac{\partial}{\partial w} L = w - \sum_{i=1}^{n} \alpha_i y_i \phi(x_i)$$

1.3 c

Setting the derivative to zero and solving for w gives us

$$\frac{\partial}{\partial w} L = w - \sum_{i=1}^{n} \alpha_i y_i \phi(x_i) = 0$$

$$w = \sum_{i=1}^{n} \alpha_i y_i \phi(x_i)$$

This allows us to replace w in our objective function as a linear combination of the data points in feature space, $\phi(x_i)$, and the signed Lagrange multipliers, $\alpha_i y_i$ as coefficients.

1.4 d

Substituting w into our objective function will give us

$$
\begin{aligned}
L = \frac{1}{2}(\sum_{i=1}^{n}\alpha_i y_i \phi(x_i))^2 - \sum_{i=1}^{n}\alpha_i(y_i(\sum_{j=1}^{n}\alpha_j y_j \phi(x_j))\phi(x_i)) - 1) \\
= \frac{1}{2}\sum_{i=1}^{n}\alpha_i y_i \phi(x_i)\sum_{j=1}^{n}\alpha_j y_j \phi(x_j) - \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j \phi(x_i)^T\phi(x_j) - \alpha_i \\
= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j \phi(x_i)^T\phi(x_j) - \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j \phi(x_j)^T\phi(x_i) + \sum_{i=1}^{n}\alpha_i \\
= \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j \phi(x_i)^T\phi(x_j)
\end{aligned}
$$

So the new objective function will be

$$
max_\alpha \ L_{dual} = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j K(x_i, x_j)
$$

subject to $\alpha_i \geq 0 \ \forall i \in D$

1.5 e

let $J(\alpha) = L_{dual}$ then taking the partial derivative with repect to one of the Lagrange multipliers we get

$$
\frac{\partial J(\alpha_k)}{\partial \alpha_k} = 1 - y_k \sum_{i=1}^{n}\alpha_i y_i K(x_i, x_k)
$$

1.6 f

After calculating all of the Lagrange multipliers we can compute w using the formula we derived from part (c)

$$
w = \sum_{i=1}^{n}\alpha_i y_i \phi(x_i)
$$

We can then calculate the class of a new point z as

$$
y = sign(w^T\phi(z)) = sign(\sum_{\alpha_i \geq 0}\alpha_i y_i K(x_i, z))
$$

1.7 g

Picture (a) can not represent hard-margin kernel SVM since the dataset is non-seperable. I.e. there is a red data point inside the circle with the blue data points.
Picture (b) can represent hard-margin kernel SVM since the dataset is seperable and the kernel will just be a linear kernel.
Picture (c) can represent hard-margin kernel SVM since the dataset is seperable and the kernel will just be a quadratic kernel.
Picture (d) can not represent hard-margin kernel SVM since the dataset is non-seperable. I.e. There are 2 data points which lie on the seperating hyperplane so they cannot be classified.

2 Part 2

2.1 a

(1)

$$W(S_1, T_1) = 4$$
$$W(S_2, T_2) = 2 + 3 = 5$$
$$W(S_3, T_3) = 10 + 10 = 20$$

(2)

$$\frac{W(S_1,T_1)}{|S_1|} + \frac{W(S_1,T_1)}{|T_1|} = \frac{4}{3} + \frac{4}{3} = \frac{8}{3} \approx 2.6667$$
$$\frac{W(S_2,T_2)}{|S_2|} + \frac{W(S_2,T_2)}{|T_2|} = \frac{5}{1} + \frac{5}{5} = 5 + 1 = 6$$
$$\frac{W(S_3,T_3)}{|S_3|} + \frac{W(S_3,T_3)}{|T_3|} = \frac{20}{4} + \frac{20}{2} = 5 + 10 = 15$$

(3)

$$\frac{W(S_1,T_1)}{vol(S_1)} + \frac{W(S_1,T_1)}{vol(T_1)} = \frac{4}{3+2+3+4} + \frac{4}{10+10+100+4} = \frac{4}{12} + \frac{4}{124} = \frac{34}{93} \approx 0.3656$$
$$\frac{W(S_2,T_2)}{vol(S_2)} + \frac{W(S_2,T_2)}{vol(T_2)} = \frac{5}{3+2} + \frac{5}{3+2+3+10+10+100+4} = \frac{5}{5} + \frac{5}{136} = \frac{141}{136} \approx 1.0368$$
$$\frac{W(S_3,T_3)}{vol(S_3)} + \frac{W(S_3,T_3)}{vol(T_3)} = \frac{20}{3+2+3+10+10+4} + \frac{20}{10+10+100} = \frac{20}{32} + \frac{20}{120} = \frac{19}{24} \approx 0.7917$$

From smallest to largest for cut weight: S_1, S_2, S_3
From smallest to largest for ratio cut: S_1, S_2, S_3
From smallest to largest for normalized cut: S_1, S_3, S_2
In general, the rank from smallest to largest is: S_1, S_2, S_3
The only difference was for the normalized cut since the second cut had only a single point and so $\frac{W(S_2,T_2)}{vol(S_2)}$ was equal to 1.

2.2 b

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

$$D = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}$$

$$L = \begin{bmatrix} 2 & -1 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ -1 & -1 & 3 & -1 & 0 & 0 \\ 0 & 0 & -1 & 3 & -1 & -1 \\ 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & -1 & 2 \end{bmatrix}$$

$$\begin{aligned} L_s &= \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{3}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \odot \begin{bmatrix} 2 & -1 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ -1 & -1 & 3 & -1 & 0 & 0 \\ 0 & 0 & -1 & 3 & -1 & -1 \\ 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & -1 & 2 \end{bmatrix} \odot \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{3}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \\ &= \begin{bmatrix} \frac{2}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{3}} & 0 & 0 & 0 \\ \frac{-1}{\sqrt{2}} & \frac{2}{\sqrt{2}} & \frac{-1}{\sqrt{3}} & 0 & 0 & 0 \\ \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & \frac{3}{\sqrt{3}} & \frac{-1}{\sqrt{3}} & 0 & 0 \\ 0 & 0 & \frac{-1}{\sqrt{3}} & \frac{3}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ 0 & 0 & 0 & \frac{-1}{\sqrt{3}} & \frac{2}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ 0 & 0 & 0 & \frac{-1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{2}{\sqrt{2}} \end{bmatrix} \odot \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{3}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \\ &= \begin{bmatrix} 2 & \frac{-1}{2} & \frac{-1}{3} & 0 & 0 & 0 \\ \frac{-1}{2} & 2 & \frac{-1}{3} & 0 & 0 & 0 \\ \frac{-1}{2} & \frac{-1}{2} & 3 & \frac{-1}{3} & 0 & 0 \\ 0 & 0 & \frac{-1}{3} & 3 & \frac{-1}{2} & \frac{-1}{2} \\ 0 & 0 & 0 & \frac{-1}{3} & 2 & \frac{-1}{2} \\ 0 & 0 & 0 & \frac{-1}{3} & \frac{-1}{2} & 2 \end{bmatrix} \end{aligned}$$

dendogram1.png

3 Part 3

For the set, $X = \{0, 1, 2, 2, 10\}$, the mean is

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{5}(0 + 1 + 2 + 2 + 10) = \frac{15}{5} = 3$$

And the median is the value m where $P(X \leq m) \geq \frac{1}{2}$ and $P(X \geq m) \geq \frac{1}{2}$
Which in this case will just be 2 since it is the middle-most value.
so $m = 2$

3.1 a

(1) The sum of squared distances for the mean is

$$\mu = \sum_{i=1}^{n}(x_i - \mu)^2 = [(0-3)^2 + (1-3)^2 + (2-3)^2 + (2-3)^2 + (10-3)^2] = (9+4+1+1+49) = 64$$

And the sum of squared distances for the median is

$$\mu = \sum_{i=1}^{n}(x_i - m)^2 = [(0-2)^2 + (1-2)^2 + (2-2)^2 + (2-2)^2 + (10-2)^2] = (4+1+0+0+64) = 69$$

Therefore, $\sum_{i=1}^{n}(x_i - \mu)^2 \leq \sum_{i=1}^{n}(x_i - m)^2$ Since $64 \leq 69$

(2) The sum of absolute distances for the mean is

$$\mu = \sum_{i=1}^{n}|x_i - \mu| = [|0-3| + |1-3| + |2-3| + |2-3| + |10-3|] = (3+2+1+1+7) = 14$$

And the sum of absolute distances for the median is

$$\mu = \sum_{i=1}^{n}|x_i - m| = [|0-2| + |1-2| + |2-2| + |2-2| + |10-2|] = (2+1+0+0+8) = 11$$

Therefore, $\sum_{i=1}^{n}(x_i - m)^2 \leq \sum_{i=1}^{n}(x_i - \mu)^2$ Since $11 \leq 14$

3.2 b

Let

$$f = argmin_a \sum_{i=1}(x_i - a)^2$$

Also,

$$\sum_{i=1}(x_i - a)^2 = \sum_{i=1} x_i^2 - 2a\sum_{i=1} x_i + na^2$$

f reaches a minimum value when $\frac{\partial f}{\partial a} = 0$.

$$\frac{\partial f}{\partial a} = \frac{\partial}{\partial a}[\sum_{i=1} x_i^2 - 2a\sum_{i=1} x_i + na^2]$$

$$= -2\sum_{i=1} x_i + 2na$$

Set to zero.

$$-2\sum_{i=1} x_i + 2na = 0$$

$$a = \frac{1}{n}\sum_{i=1} x_i$$

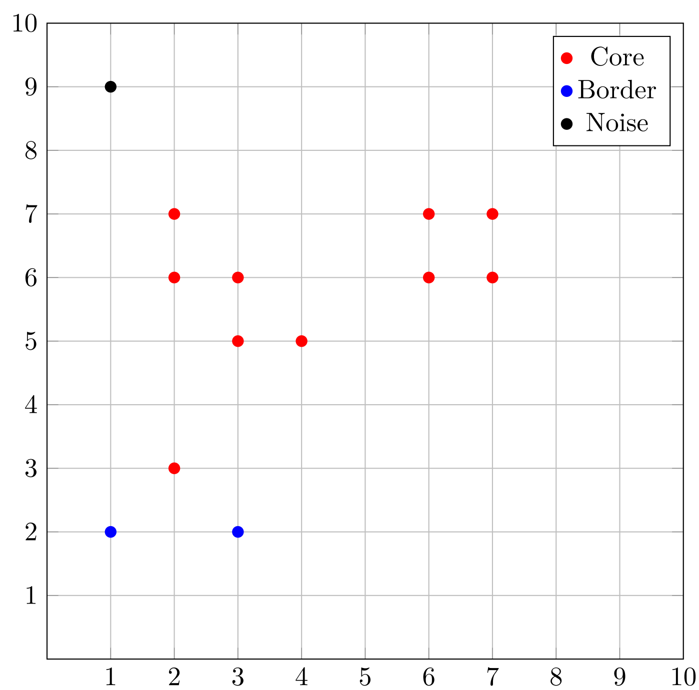
$$= \mu$$

3.3 c

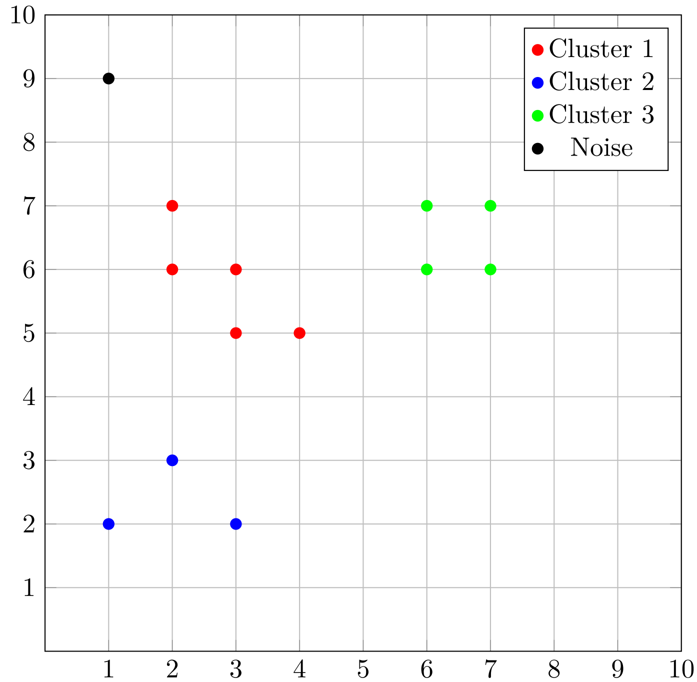
todo

4 Part 4

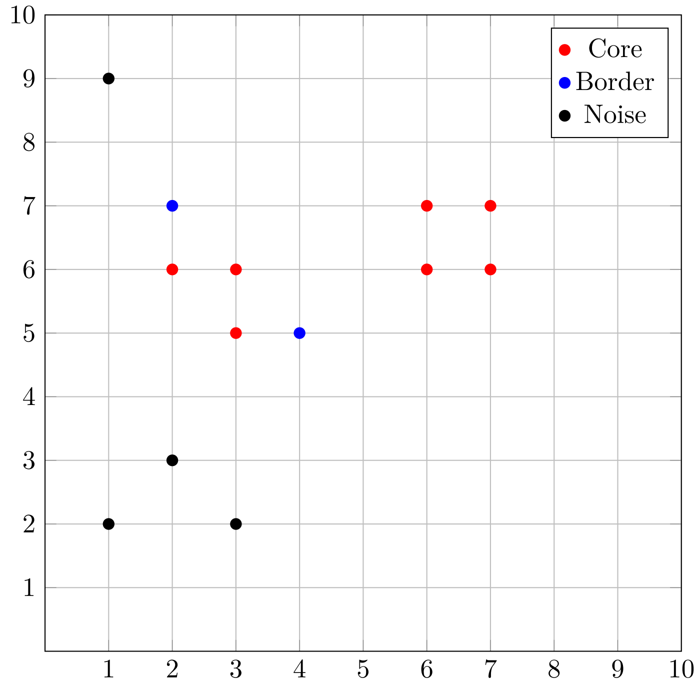
4.1 a



4.2 b

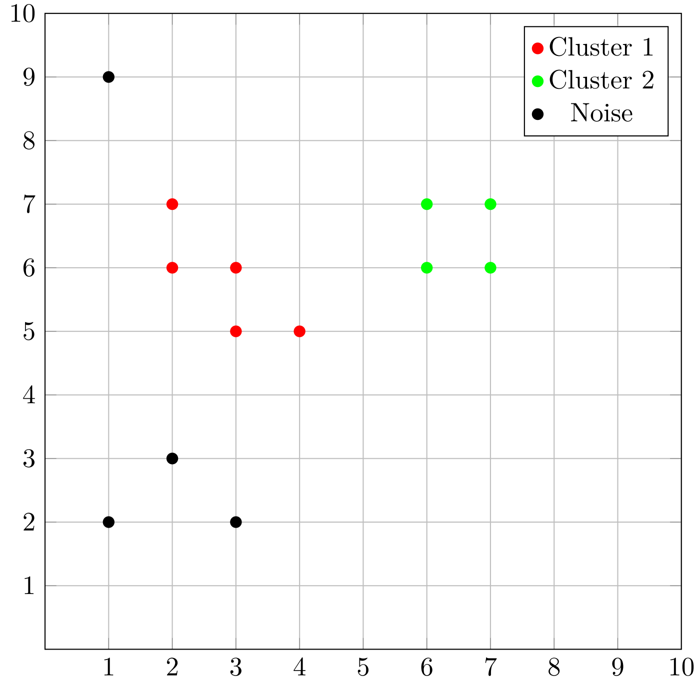

Cluster 1
Cluster 2
Cluster 3
Noise
10
9
8
7
6
5
4
3
2
1
1 2 3 4 5 6 7 8 9 10

4.3 c



Compared to part (a), the 3 points in the bottom left have all become outliers since the point (2,3) is no longer within distance of the other 2 points (Since the distance between diagonal points is $\sqrt{2}$ away from each other and that is greater than $\epsilon = 1$). Also, the points (2,7) and (4,5) are now border points because of the same issue with diagonal points. For all of these points, the number of points within distance $\epsilon = 1$ has changed and is now less than $minpts = 3$, causing them to become an outlier if they are within distance of a core point and they become a border point otherwise.

4.4 d



Compared to part (b), the bottom left points are no longer a single cluster since these points became outliers and are not connected.