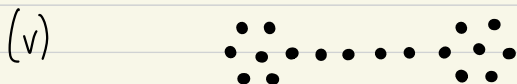
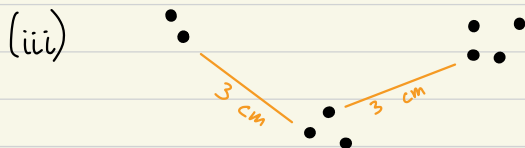
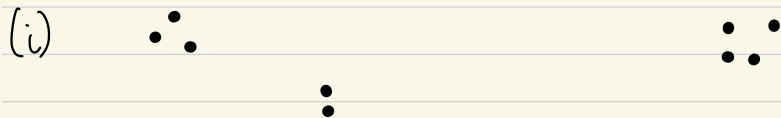


Exercises for hierarchical clustering

1. Sketch (by hand) the dendrograms for the following datasets in the plane:



2. The drastic change in the dendrogram from (iv) to (v) above is an example of the problem of "chaining". How might you address this?

2. On the course GitHub page, see the jupyter notebook "plot-linkage-comparison.ipynb". These are other forms of hierarchical (agglomerative) clustering. Which do you prefer?

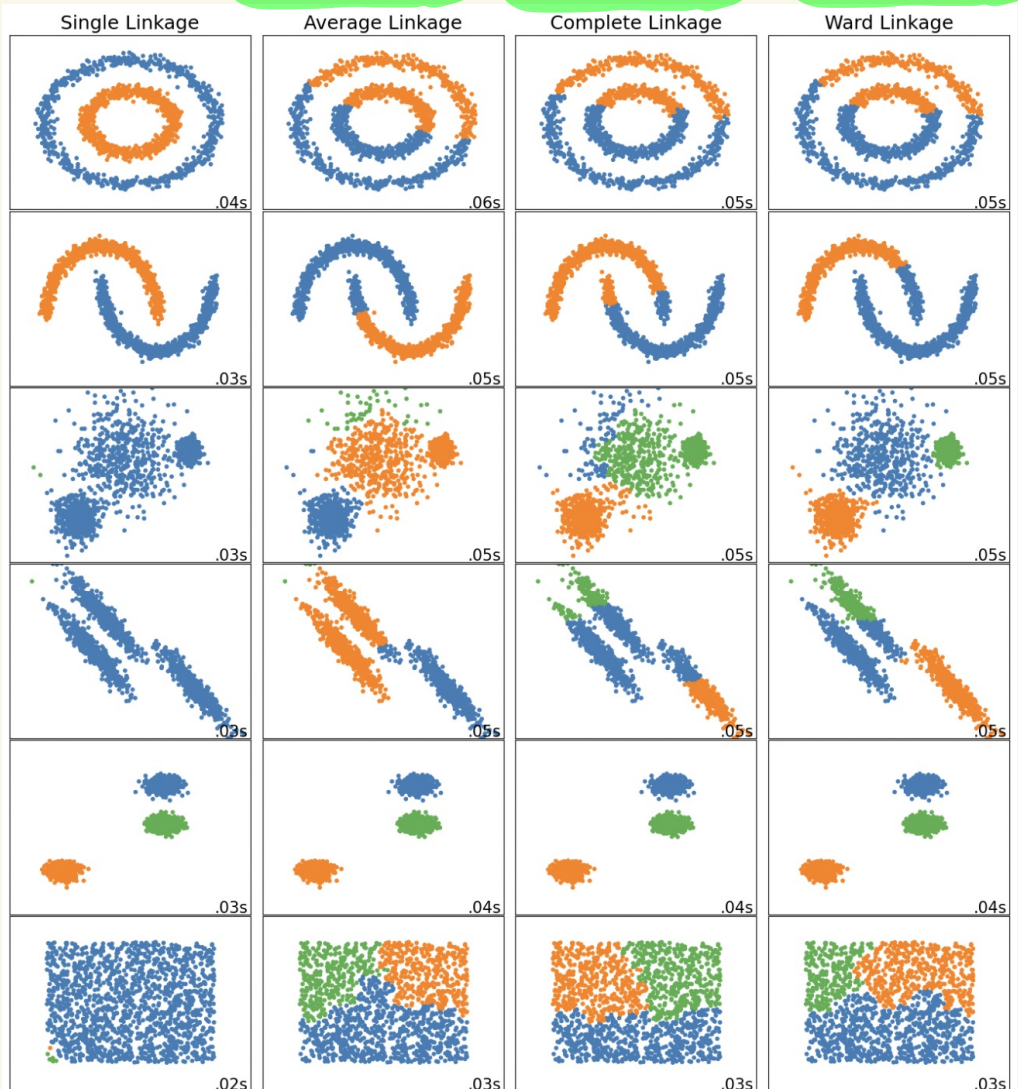
Two clusters A, B merge based on the next lowest value of:

$$\min_{a \in A, b \in B} d(a, b)$$

Fill in:

Fill in:

Fill in:

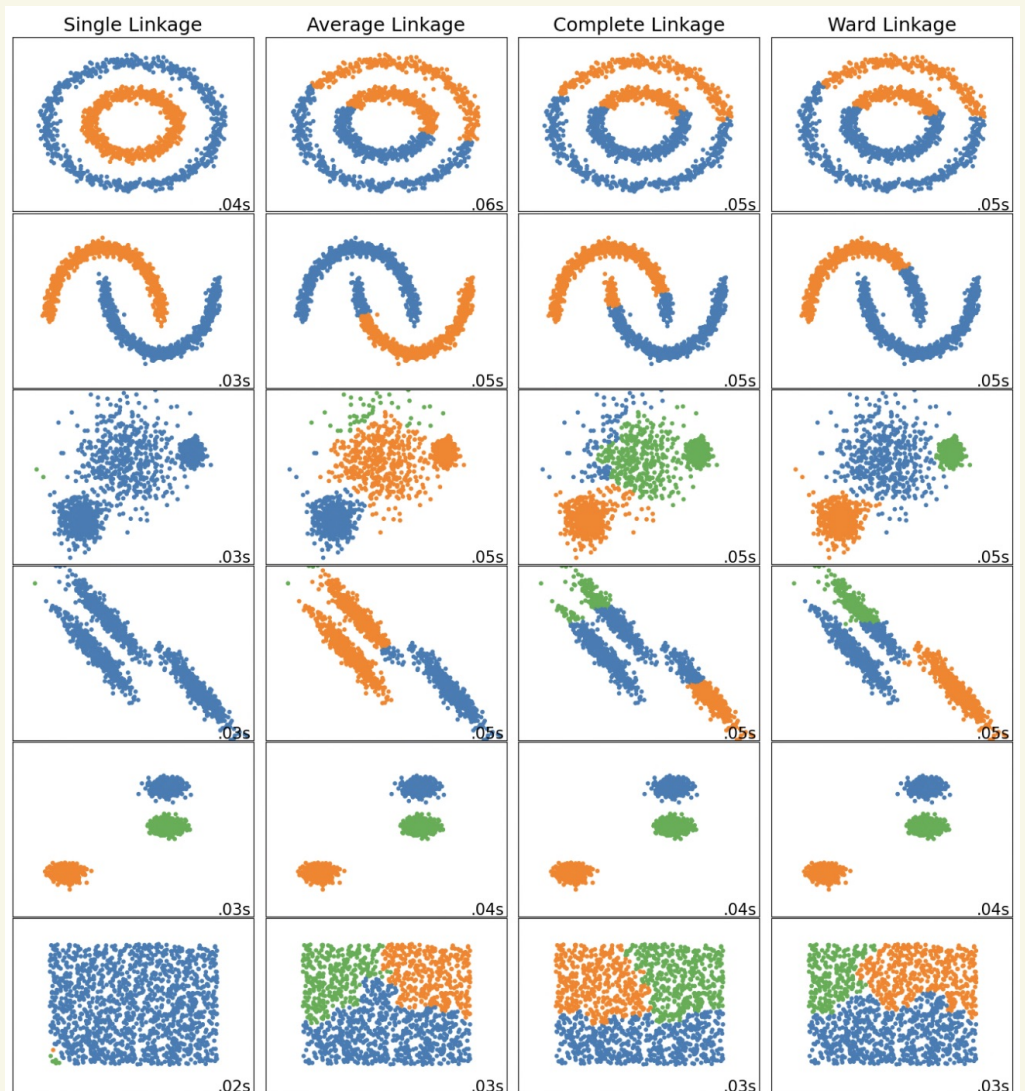


2. On the course GitHub page, see the jupyter notebook "plot-linkage-comparison.ipynb". These are other forms of hierarchical (agglomerative) clustering. Which do you prefer?

Two clusters A, B merge based on the next lowest value of:

$$\min_{a \in A, b \in B} d(a, b)$$

$$\max_{a \in A, b \in B} d(a, b)$$



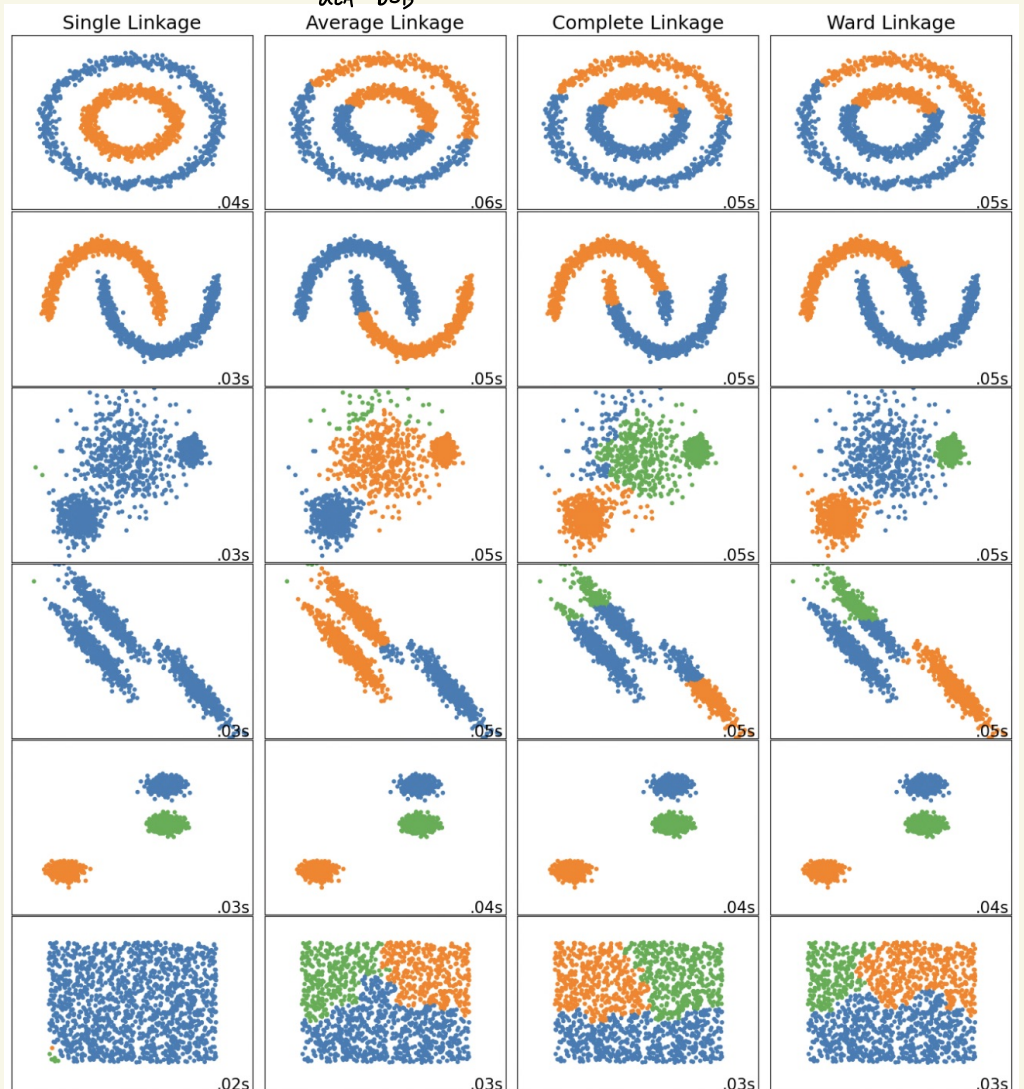
2. On the course GitHub page, see the jupyter notebook "plot-linkage-comparison.ipynb". These are other forms of hierarchical (agglomerative) clustering. Which do you prefer?

Two clusters A, B merge based on the next lowest value of:

$$\min_{a \in A, b \in B} d(a, b)$$

$$\frac{1}{|A||B|} \sum_{a \in A} \sum_{b \in B} d(a, b)$$

$$\max_{a \in A, b \in B} d(a, b)$$



2. On the course GitHub page, see the jupyter notebook "plot-linkage-comparison.ipynb". These are other forms of hierarchical (agglomerative) clustering. Which do you prefer?

minimum increase of error

Two clusters A, B merge based on the next lowest value of:

$$\min_{a \in A, b \in B} d(a, b) \quad \frac{1}{|A||B|} \sum_{a \in A} \sum_{b \in B} d(a, b) \quad \max_{a \in A, b \in B} d(a, b) \quad \frac{|A| \cdot |B|}{|A \cup B|} \|C_A - C_B\|^2$$

