浙江大学爱丁堡大学联合学院

ZJU-UoE Institute

Clustering and Machine Learning

ADS2, Lecture 2.13

Dr Rob Young – robert.young@ed.ac.uk

Semester 2, 2023/24

This lecture is about...

- How to teach a computer to cluster data ("unsupervised machine learning") **and**
- How to teach a computer to classify data ("supervised machine learning")

Learning objectives

After this lecture, you should be able to:

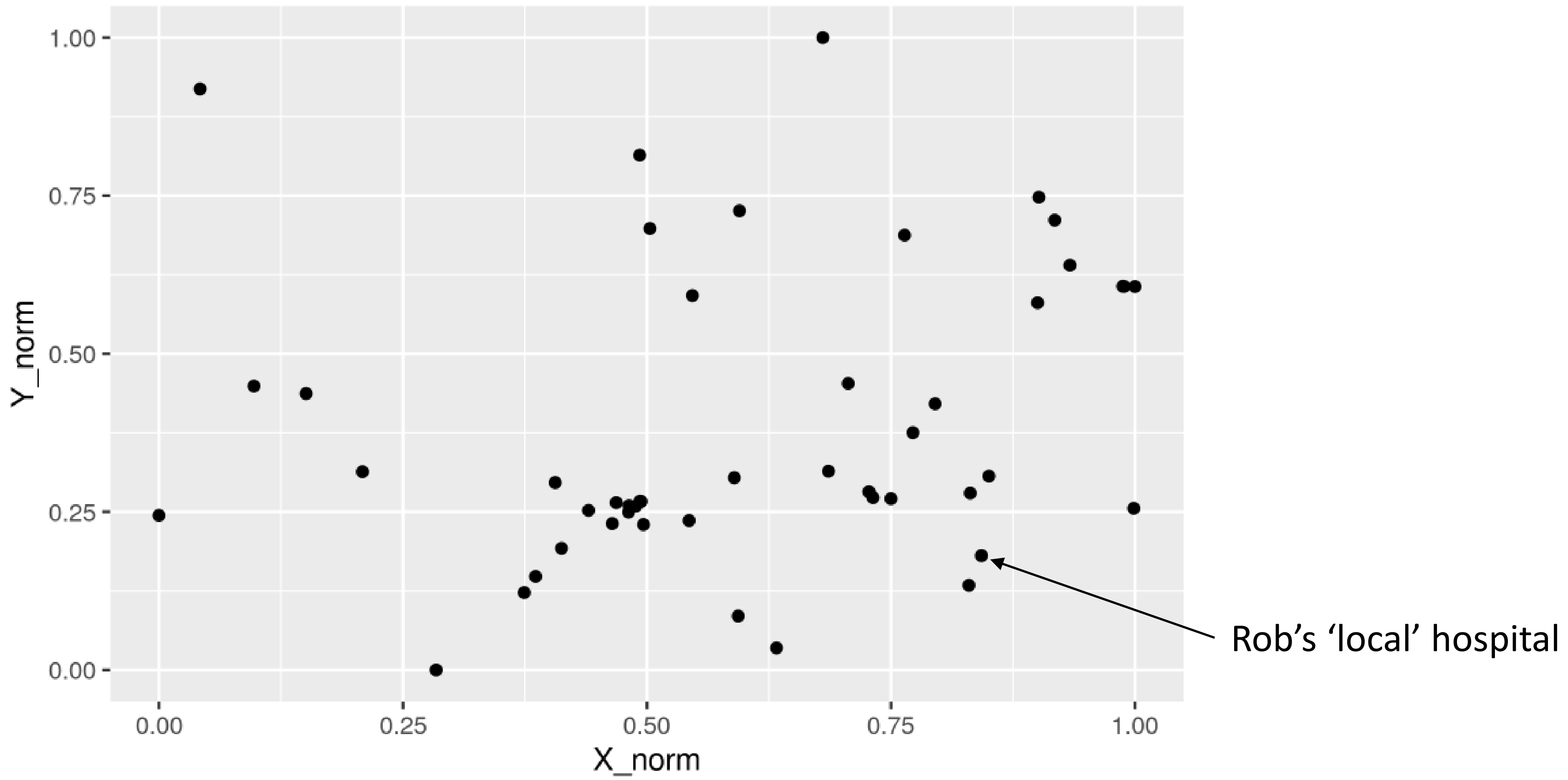
- Explain how clustering works (k-means, hierarchical clustering)
- Discuss the choices that need to be made when tackling a clustering problem
- Explain how supervised machine learning works (classification)
- How to simply evaluate a classification method

Outline

1. **What is clustering?**
2. k-means clustering
3. Hierarchical clustering
4. Machine learning

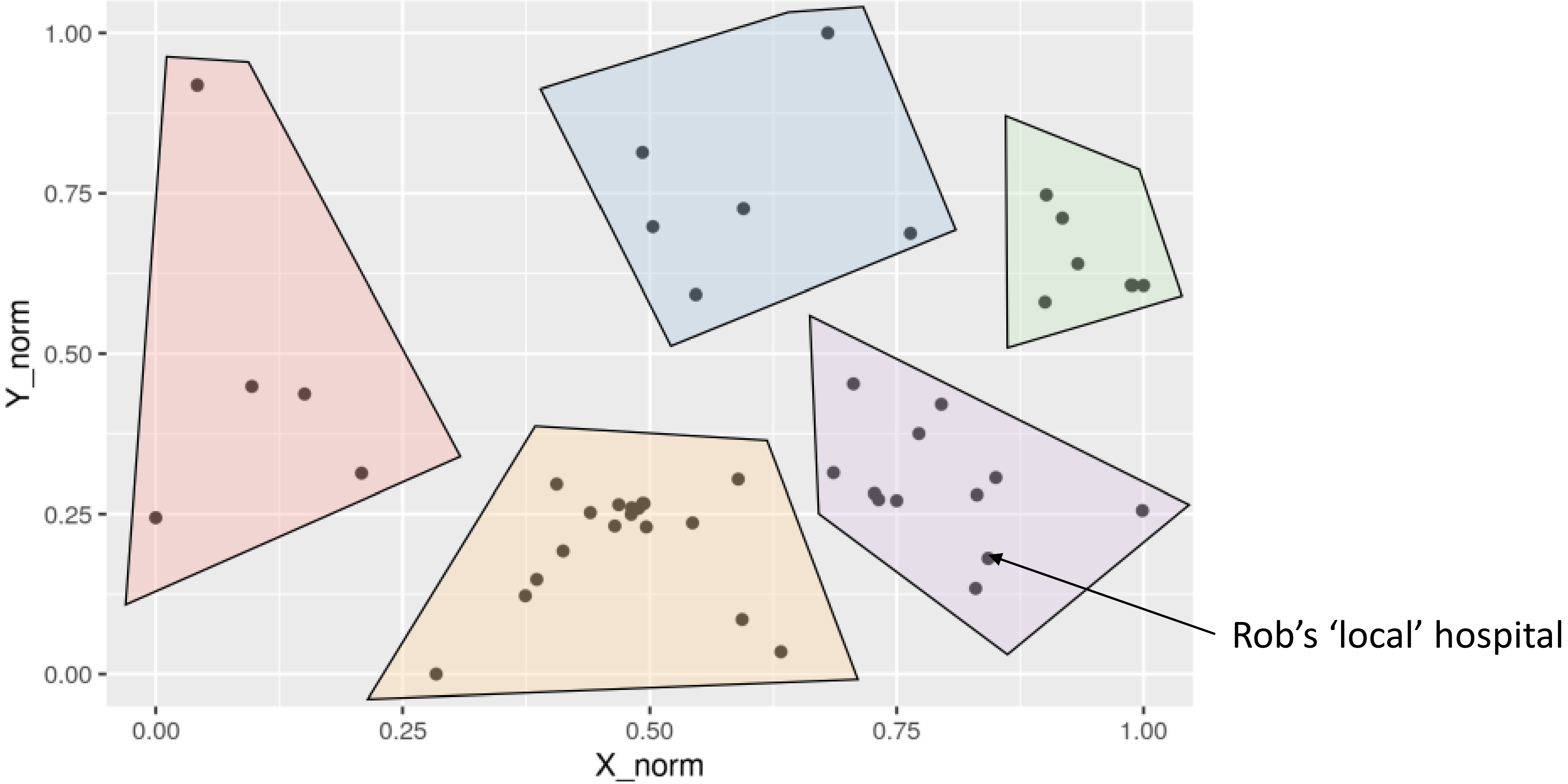
Managing Scottish Hospitals

Each dot on this map is one of a group of 50 hospitals in Scotland. The National Health Service wants to divide these 50 hospitals into regional groups to facilitate management. How would you divide them?



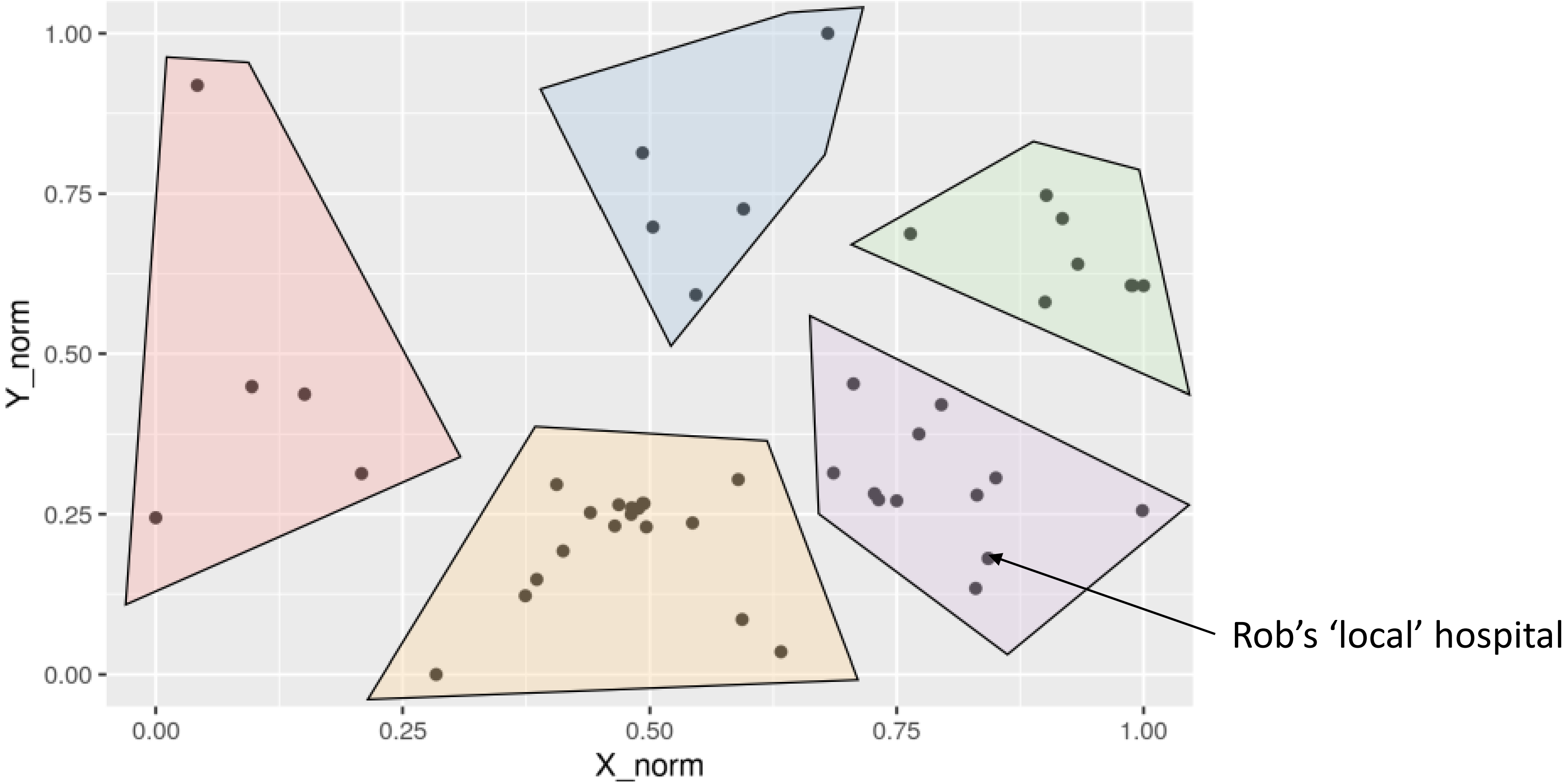
Managing Scottish Hospitals

Each dot on this map is one of a group of 50 hospitals in Scotland. The National Health Service wants to divide these 50 hospitals into regional groups to facilitate management. How would you divide them?



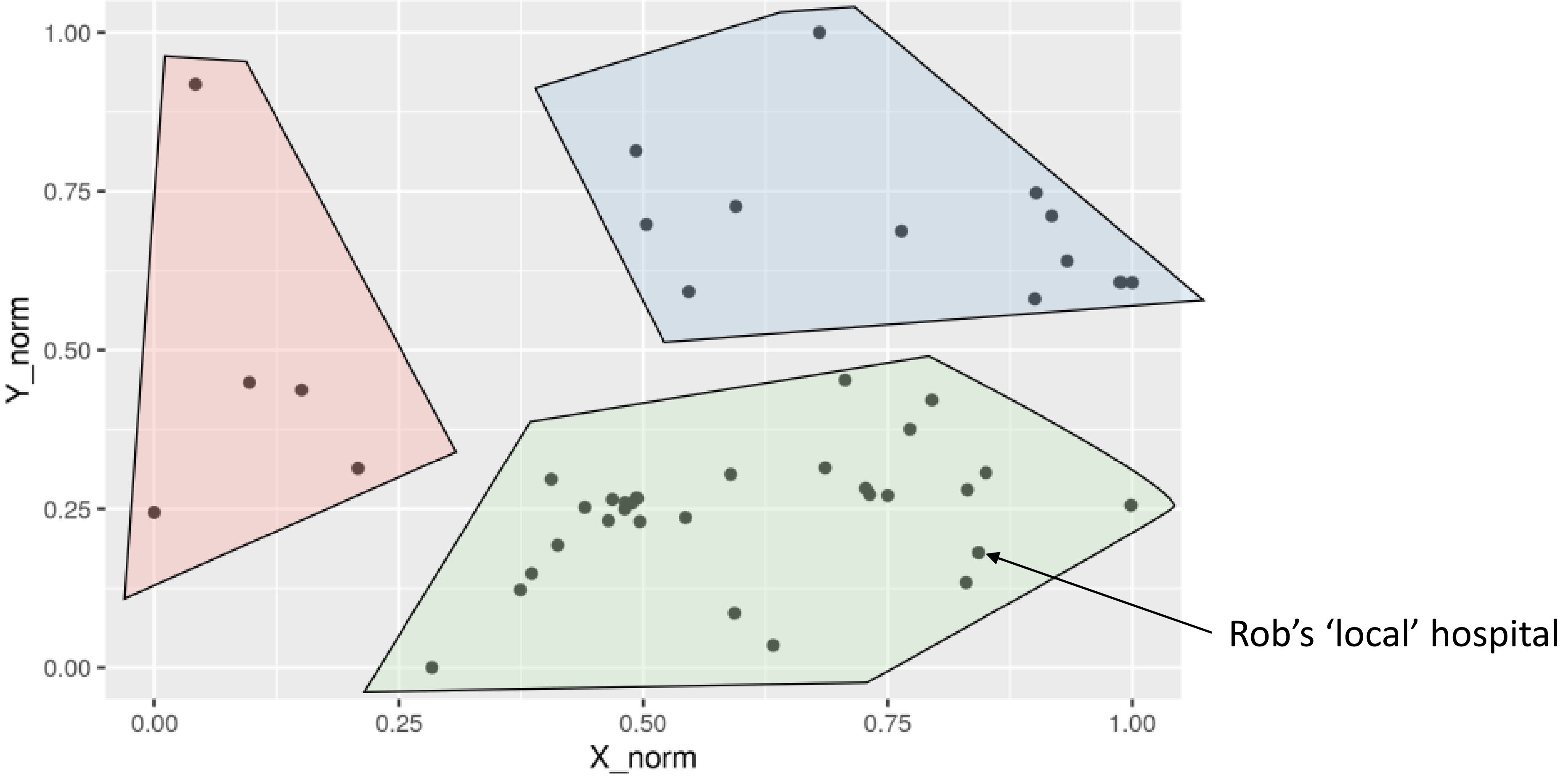
Managing Scottish Hospitals

Each dot on this map is one of a group of 50 hospitals in Scotland. The National Health Service wants to divide these 50 hospitals into regional groups to facilitate management. How would you divide them?



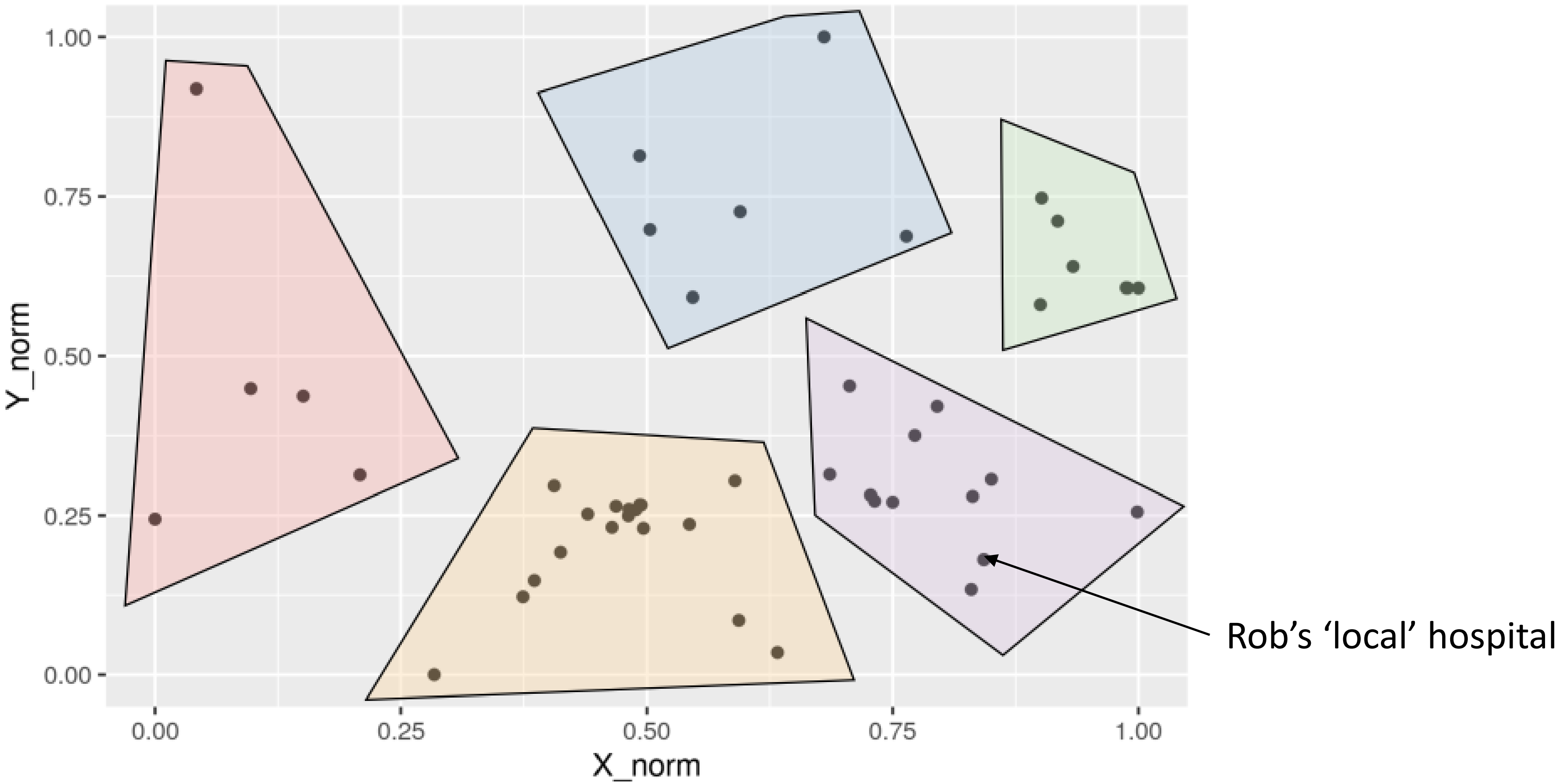
Managing Scottish Hospitals

Each dot on this map is one of a group of 50 hospitals in Scotland. The National Health Service wants to divide these 50 hospitals into regional groups to facilitate management. How would you divide them?



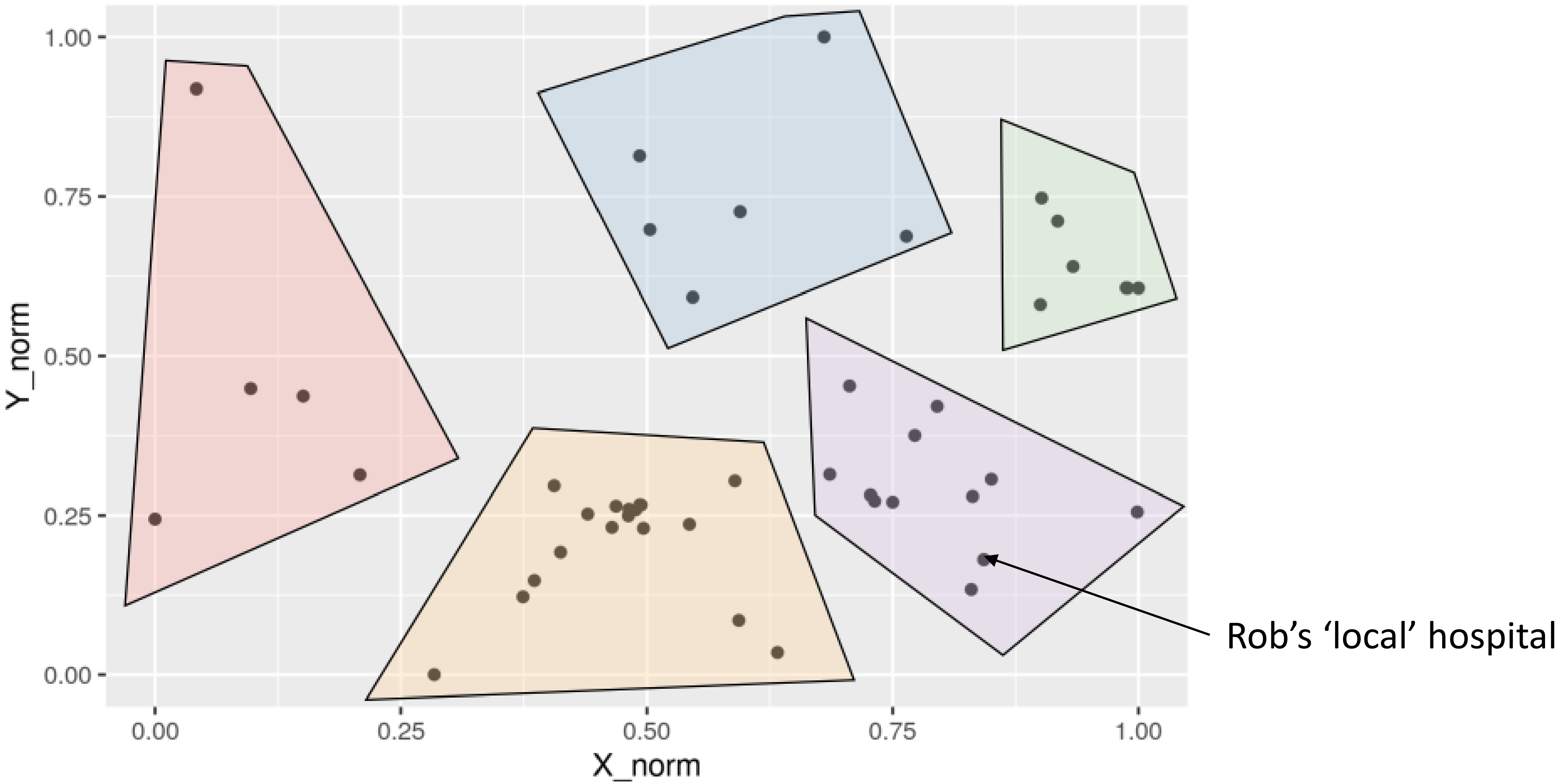
What exactly did you do?

What exactly happens when we cluster data? How would you explain it to a computer?



What happens when we cluster data?

We want data points within the same cluster to be 'close together', and we want them to be 'further away' and separate from other clusters.



What happens when we cluster data?

This requires an idea of ‘distance’.

Here we use geographical coordinates (normalised) and Euclidian distance in two dimensions. But we could use any number of dimensions, any measure and any (reasonable) metric of distance.

What happens when we cluster data?

This requires an idea of 'distance'.

Here we use geographical coordinates (normalised) and Euclidian distance in two dimensions. But we could use any number of dimensions, any measure and any (reasonable) metric of distance.

- Distance along roads/driving time (hospitals)
- (Differences in) height, weight, blood pressure, temperature (patients)
- Number of base substitutions (DNA sequences)
- …

What happens when we cluster data?

This requires an idea of ‘distance’.

Here we use geographical coordinates (normalised) and Euclidian distance in two dimensions. But we could use any number of dimensions, any measure and any (reasonable) metric of distance.

- Distance along roads/driving time (hospitals)
- (Differences in) height, weight, blood pressure, temperature (patients)
- Number of base substitutions (DNA sequences)
- …
- *Pay attention to scaling!*

Outline

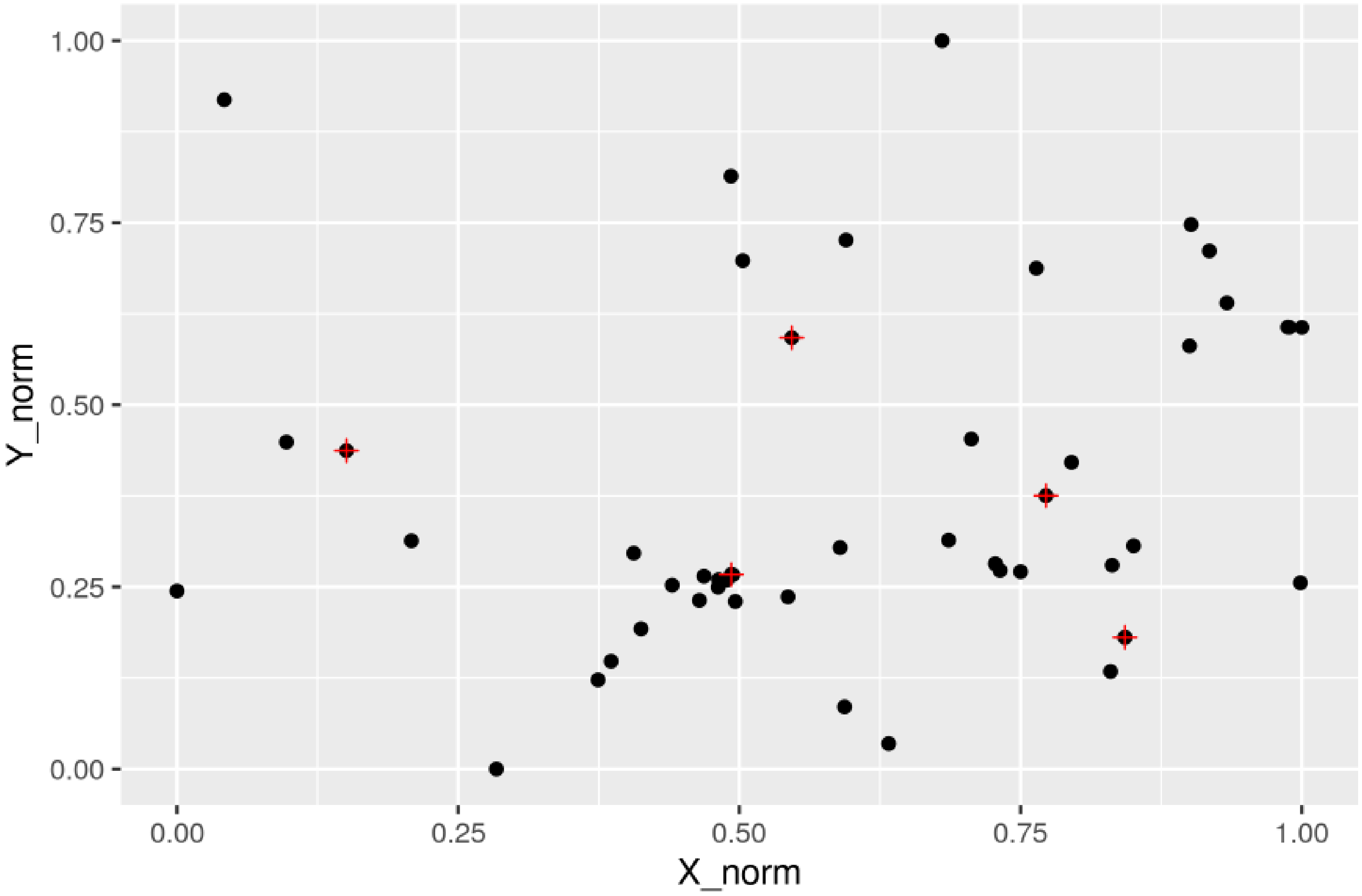
1. **What is clustering?**
2. **k-means clustering**
3. Hierarchical clustering
4. Machine learning

k-means clustering

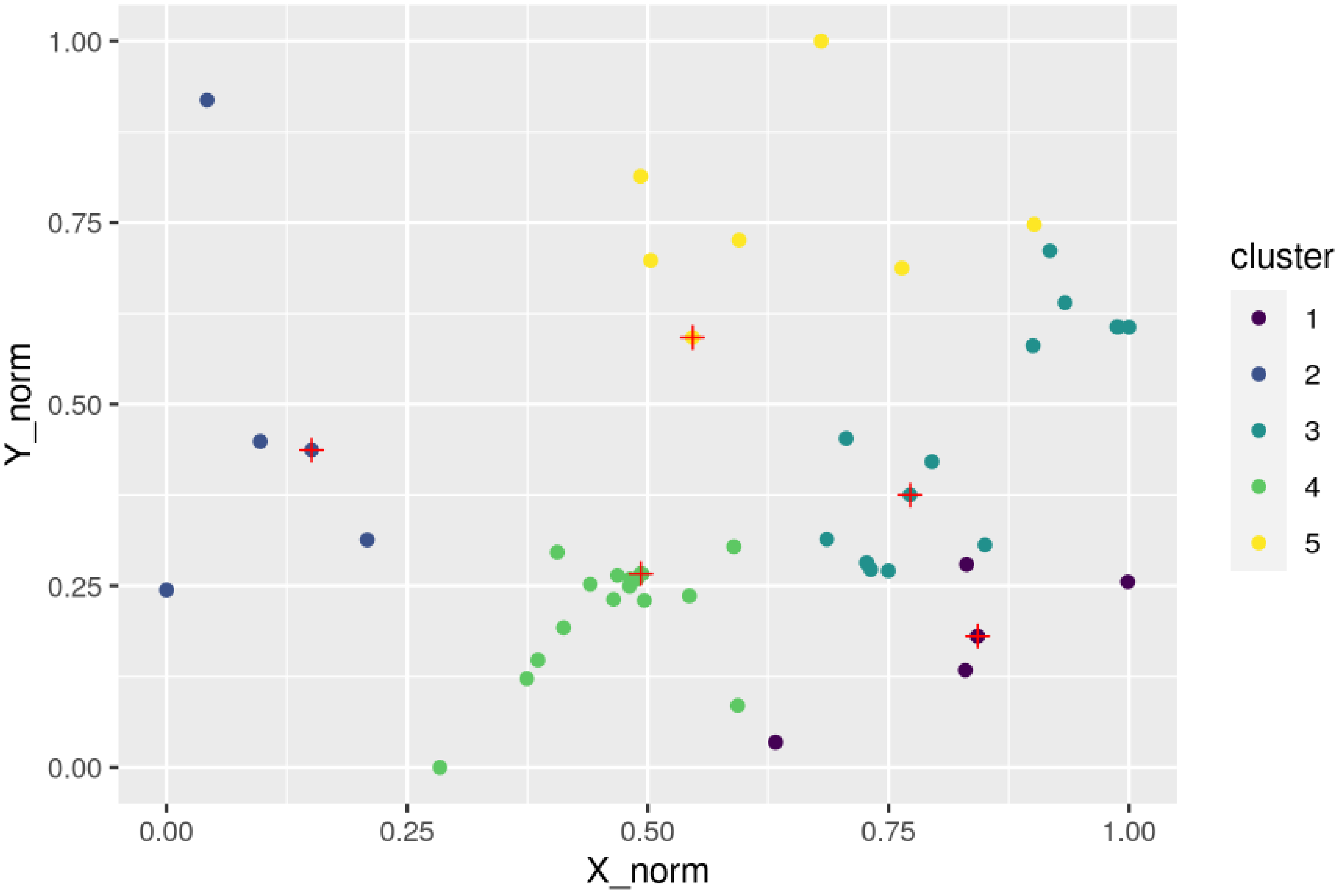
This idea behind this is that you:

- Decide beforehand on the number of clusters you want (*k*)
- Select *k* data points at random, and make them the *k* cluster centres
- Assign each data point to the nearest cluster
- Re-compute the centre of each cluster
- Re-assign data points to the nearest cluster
- Repeat until this converges to stable clusters

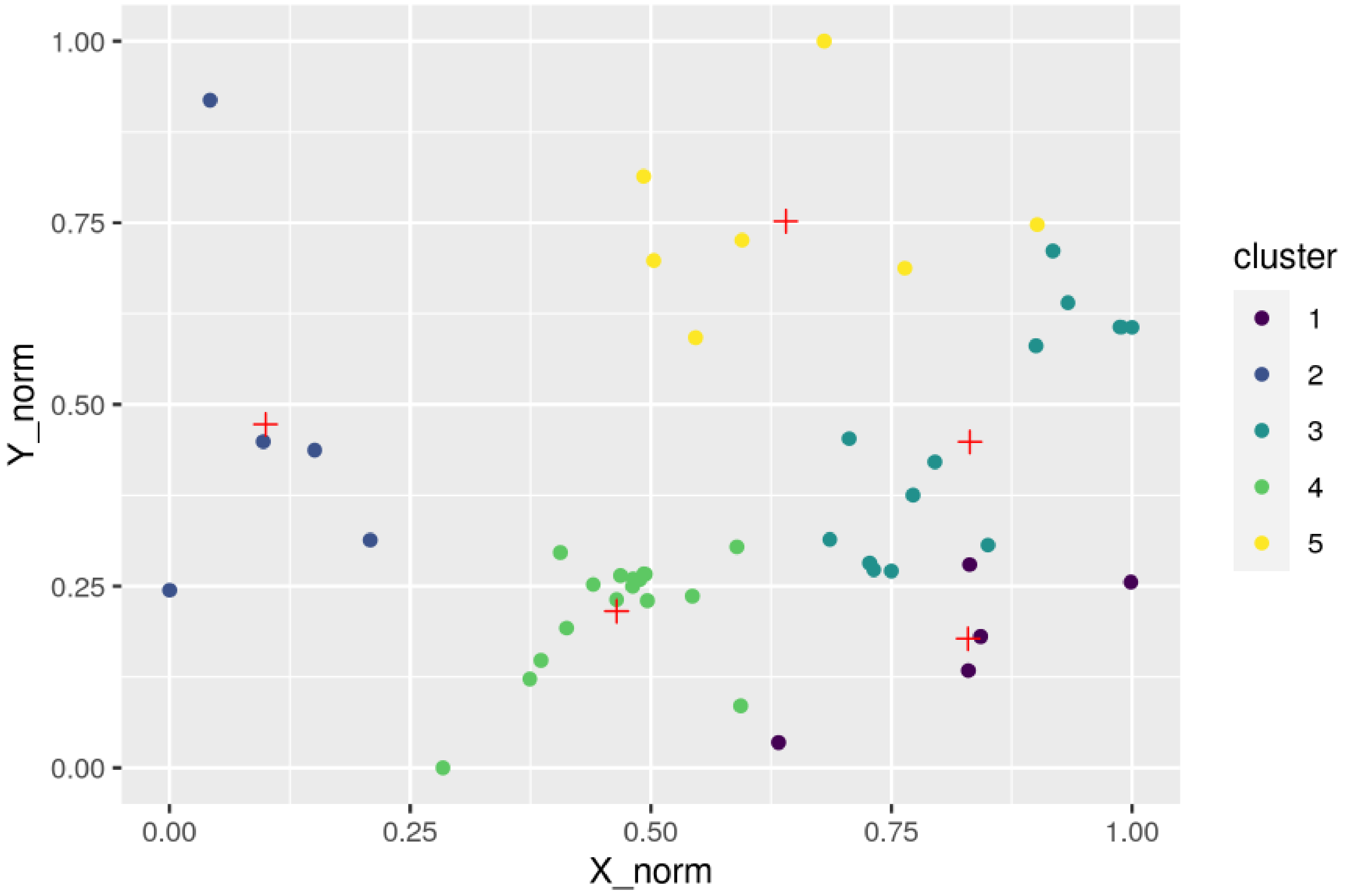
k-means clustering



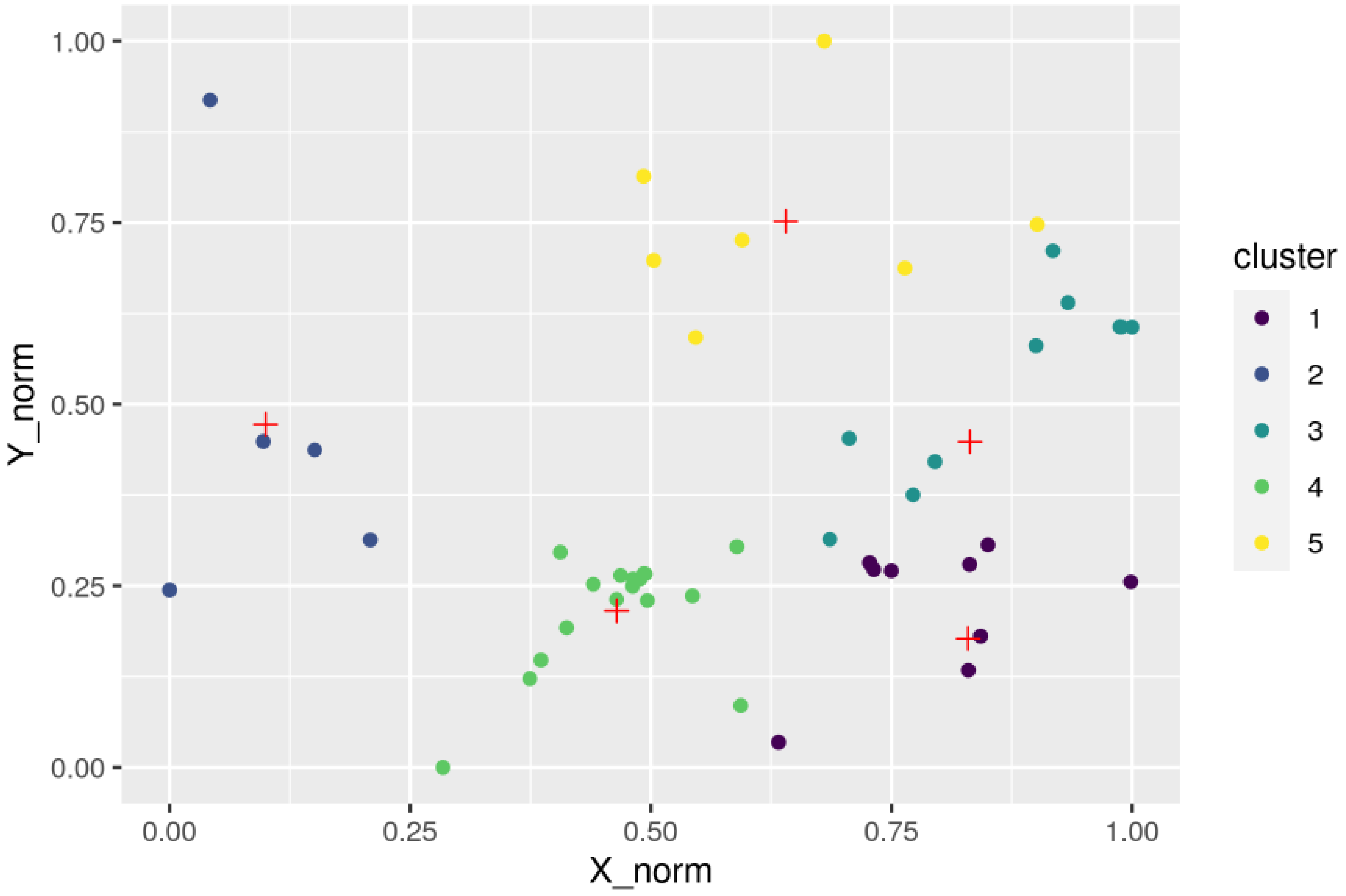
k-means clustering



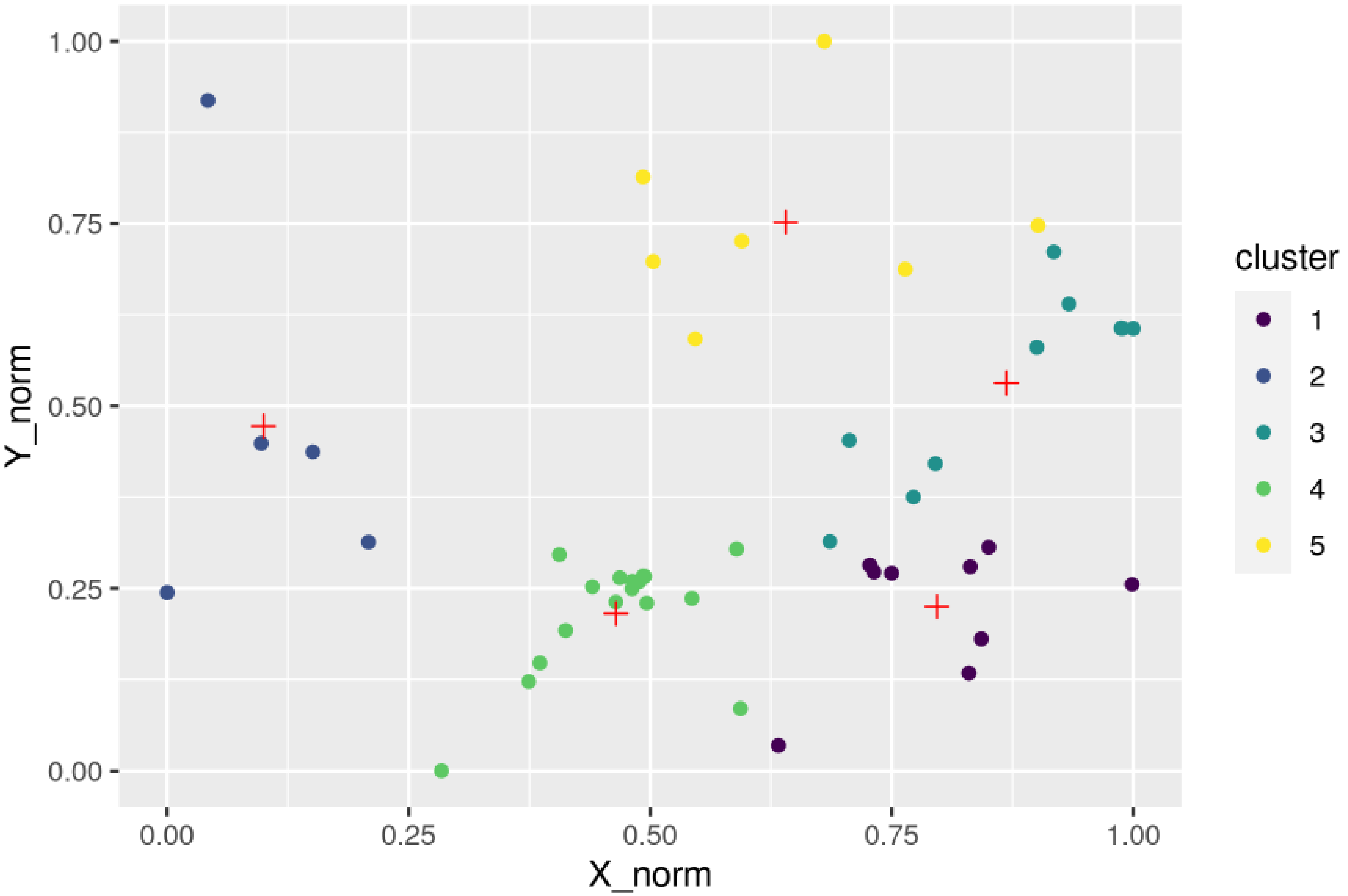
k-means clustering



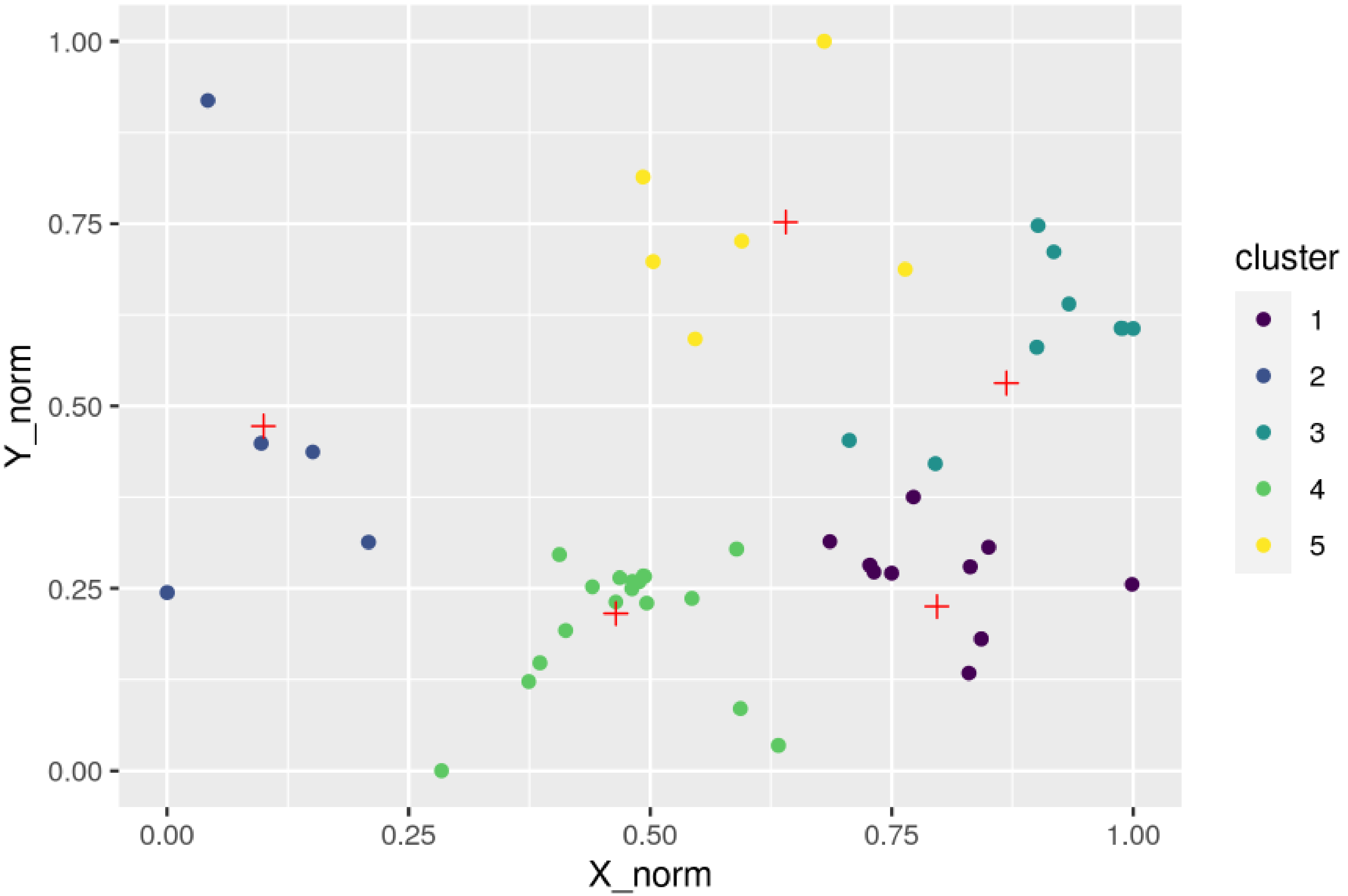
k-means clustering



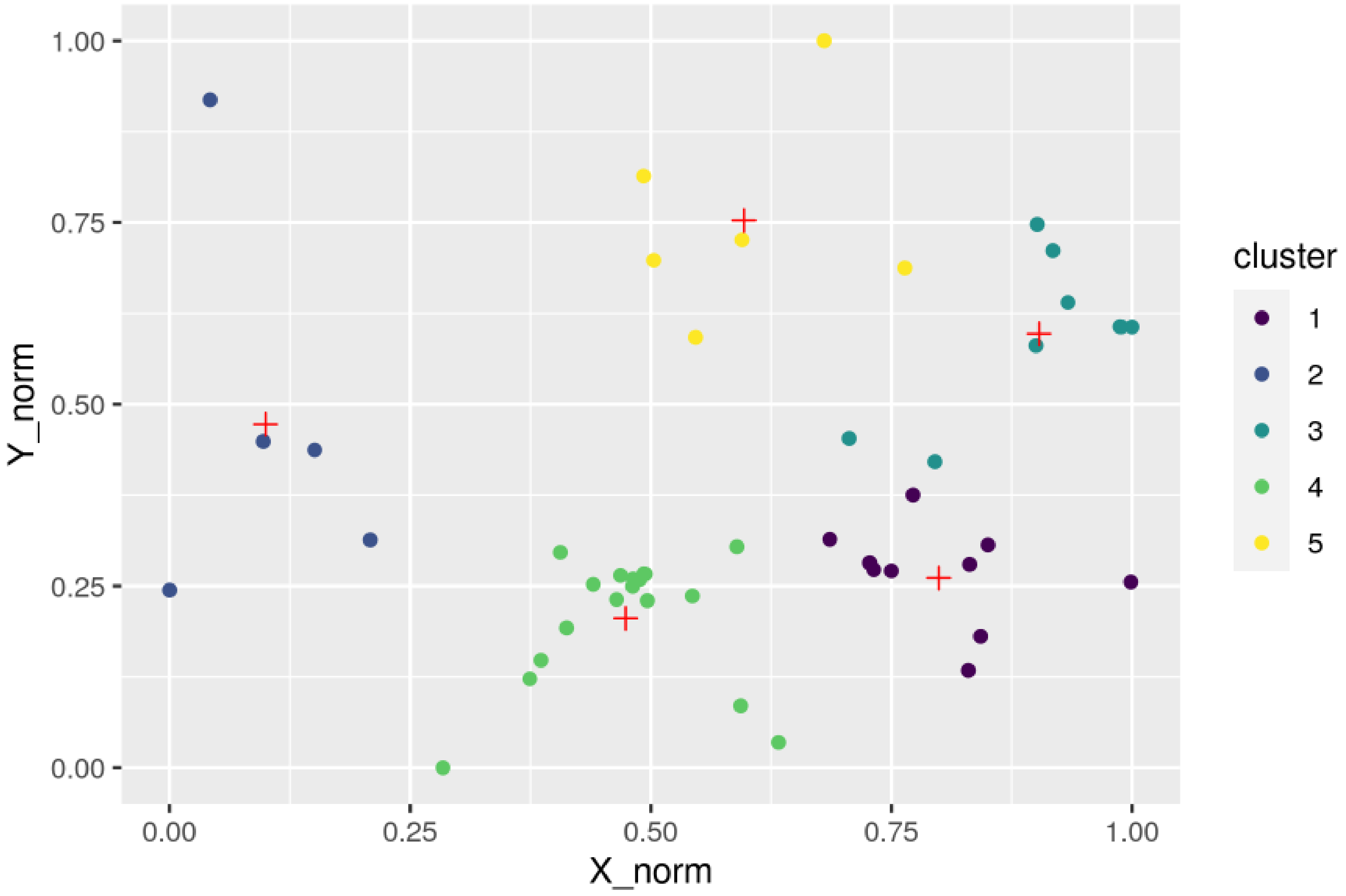
k-means clustering



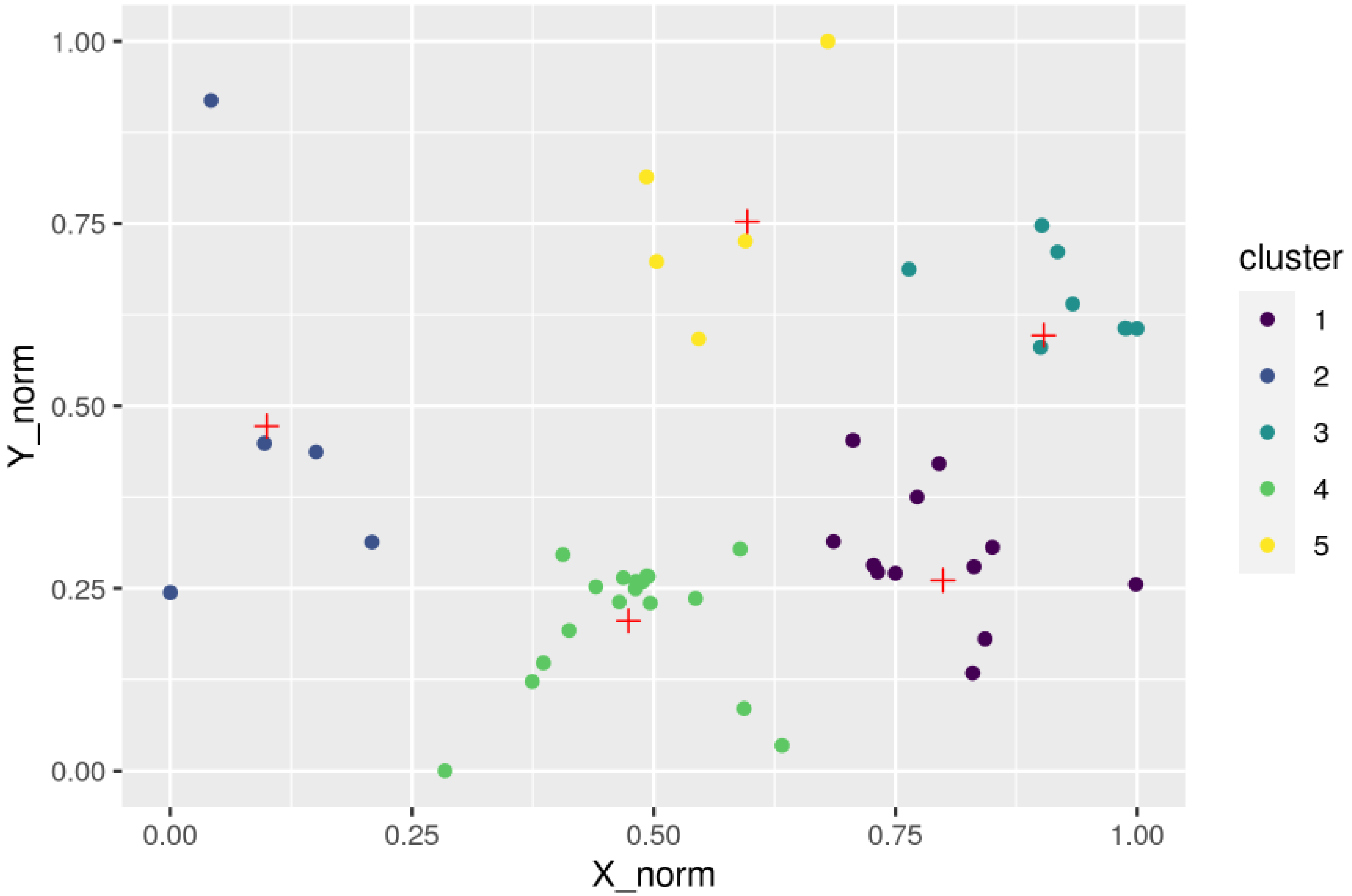
k-means clustering



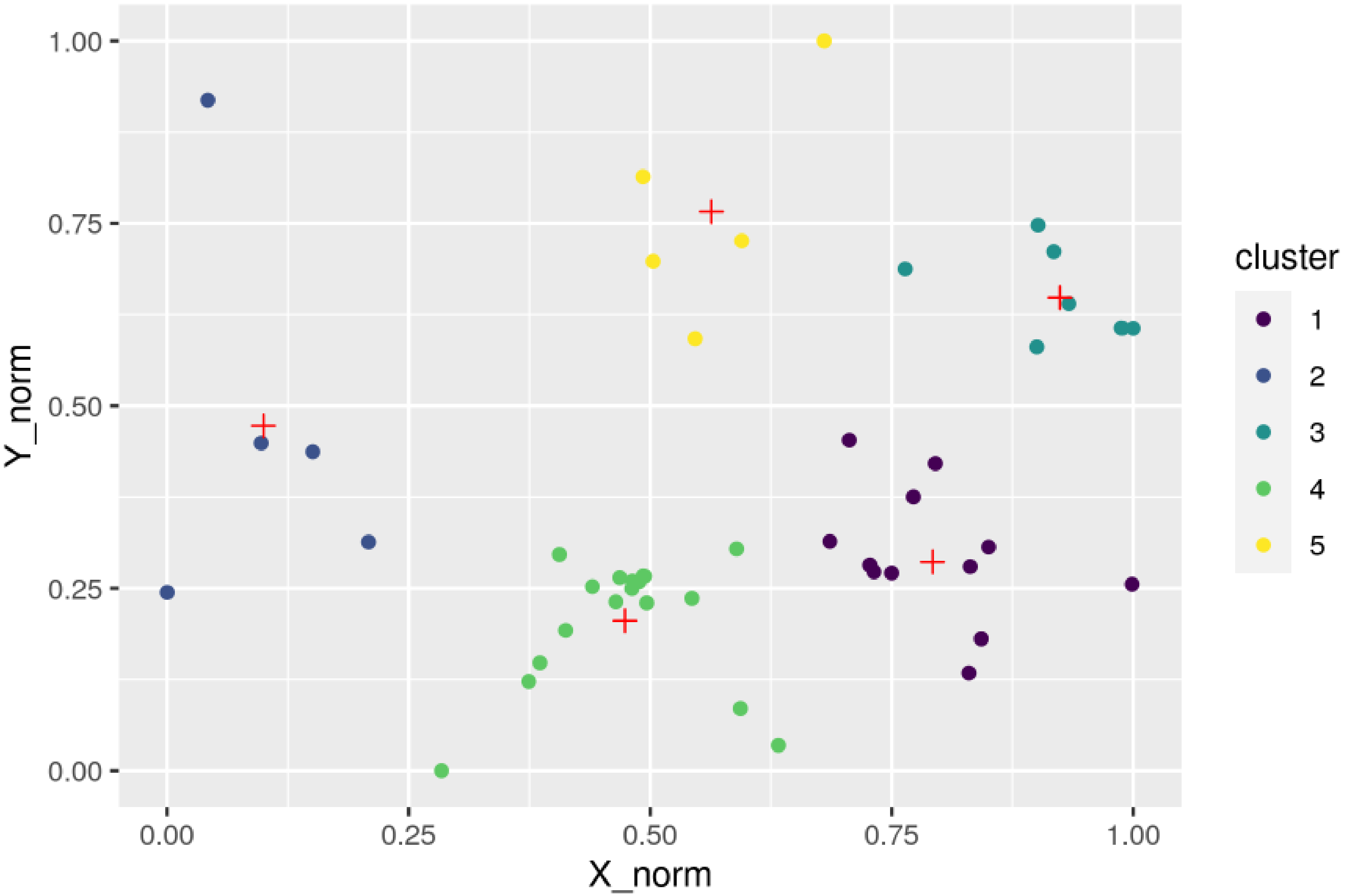
k-means clustering



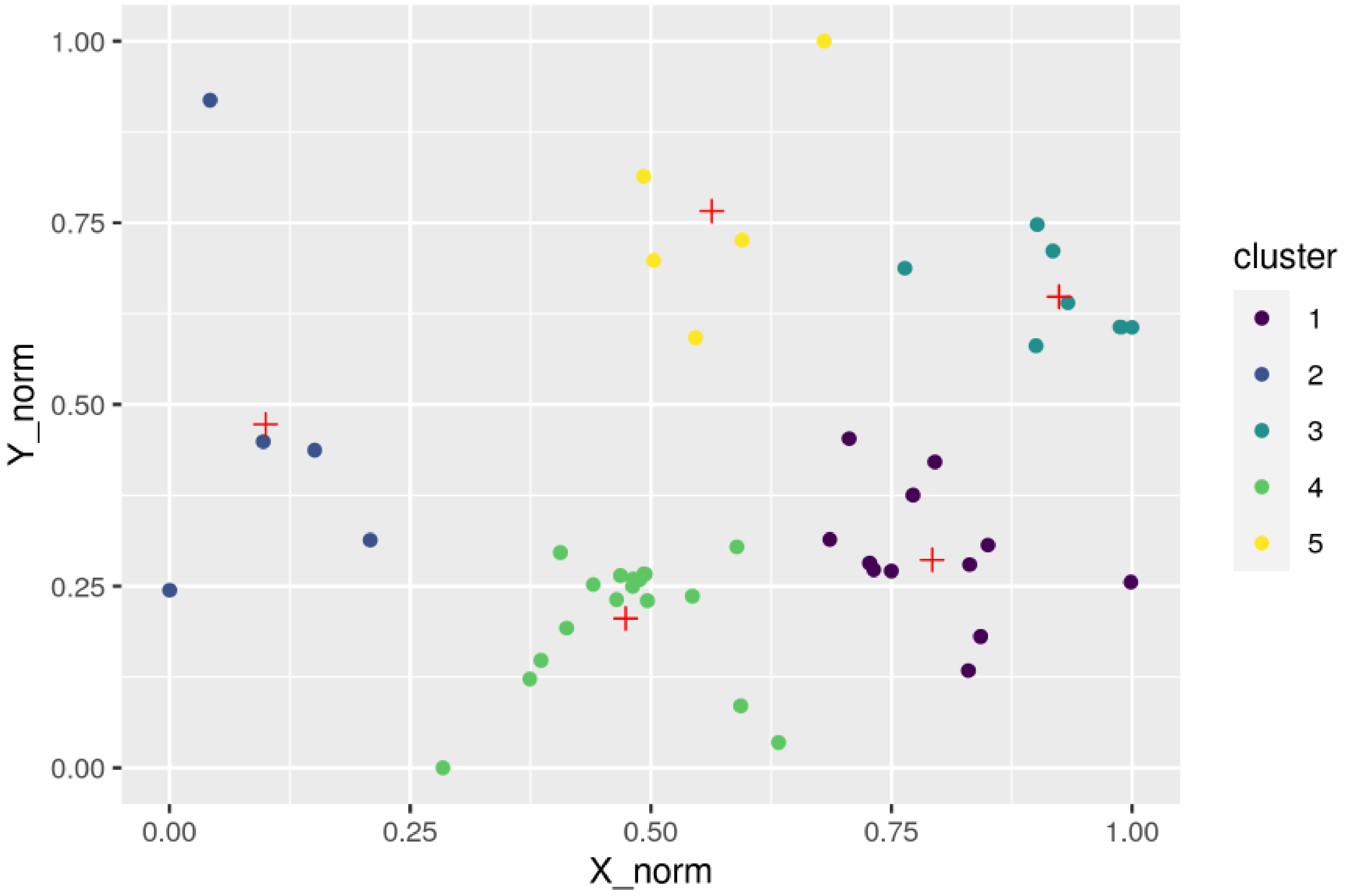
k-means clustering



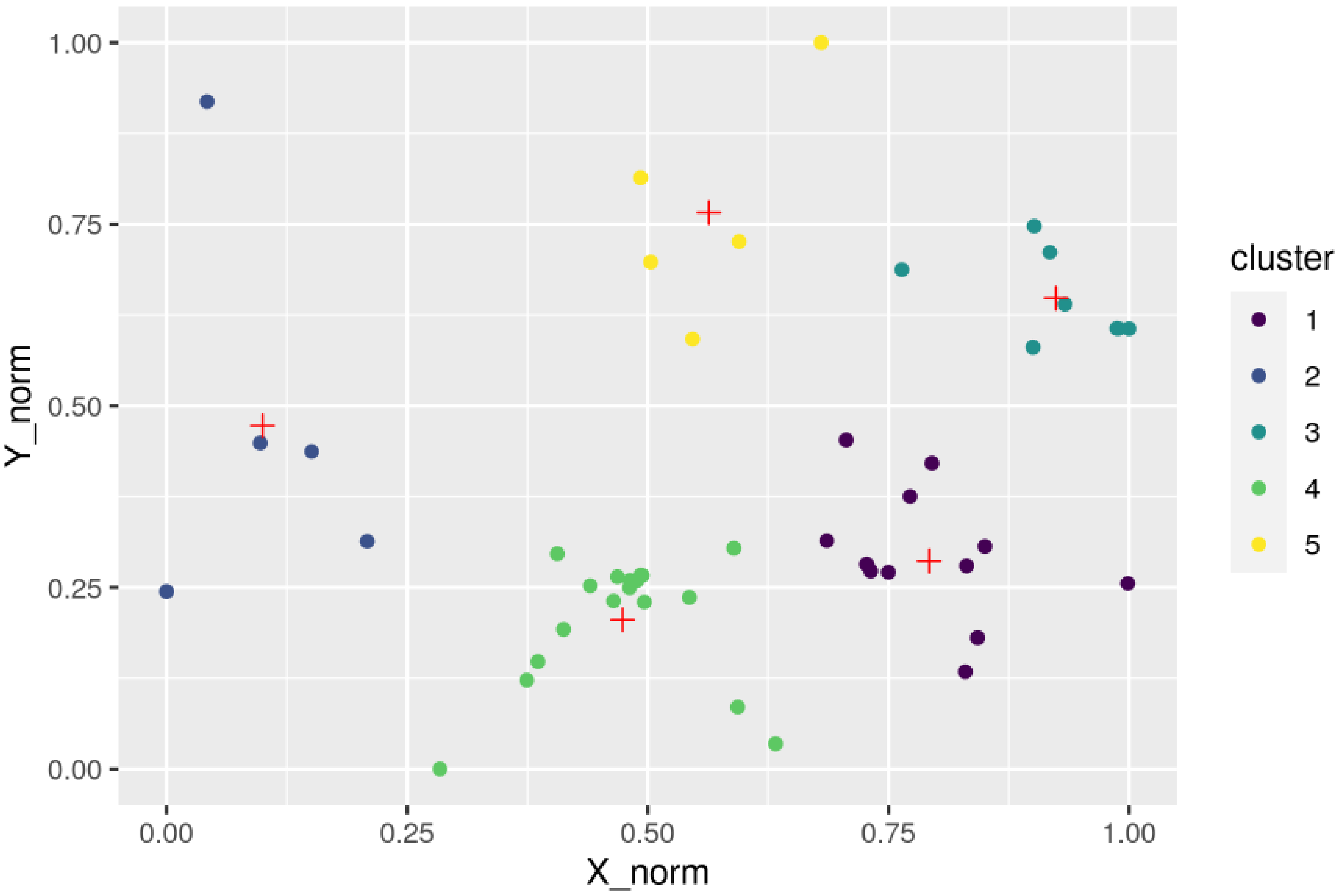
k-means clustering



k-means clustering



k-means clustering



k-means clustering

What questions do you have? (Or what problems may there be?)

k-means clustering

What questions do you have? (Or what problems may there be?)

- What do we mean by ‘centre’?
- Will this always converge?
- Doesn’t it depend on what points are initially chosen?
- How do I know what *k* I want?

k-means clustering

- What do we mean by ‘centre’?
  Most often: *centroid* (mean coordinate along every axis)
  or *medoid* (‘median’ data point).

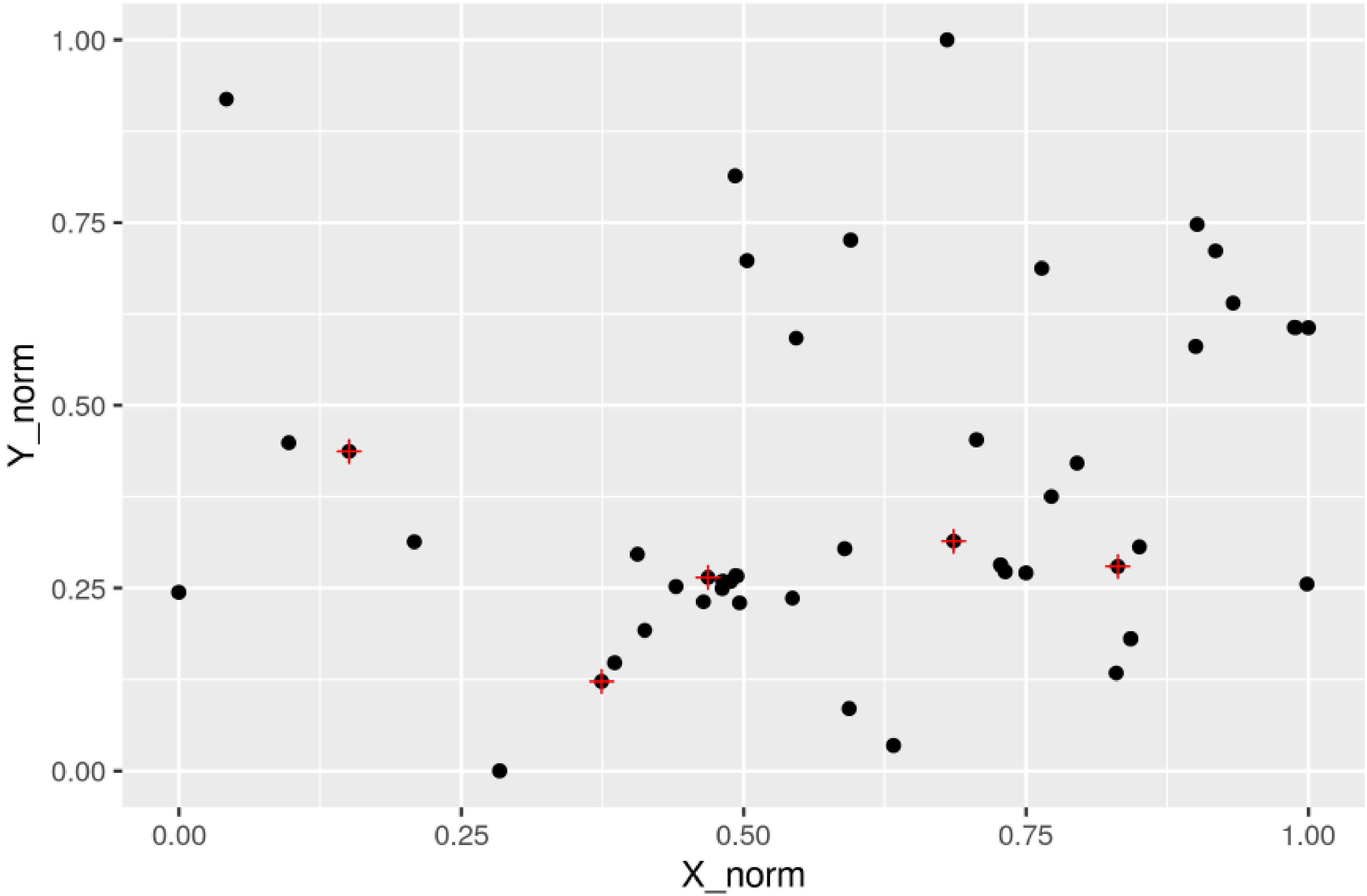
k-means clustering

- What do we mean by ‘centre’?
  Most often: *centroid* (mean coordinate along every axis)
  or *medoid* (‘median’ data point).
- Will this always converge?
  Ish. Most of the time it does, and quite quickly. You may need a way to break ties if a data point is equally far away from two centroids.

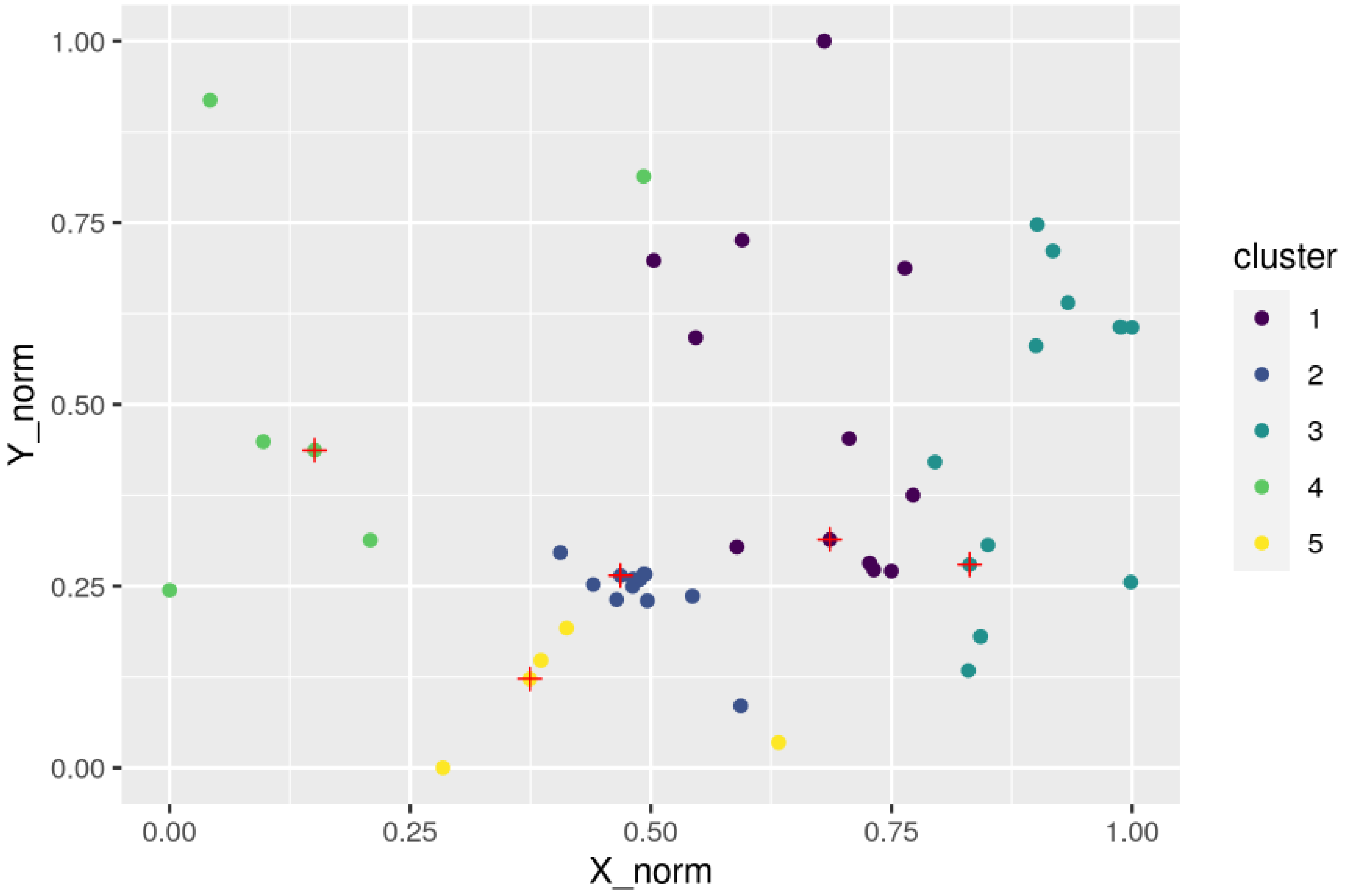
k-means clustering

- What do we mean by ‘centre’?
  Most often: *centroid* (mean coordinate along every axis)
  or *medoid* (‘median’ data point).
- Will this always converge?
  Ish. Most of the time it does, and quite quickly. You may need a way to break ties if a data point is equally far away from two centroids.
- Doesn’t it depend on what points are initially chosen?
  Yes!

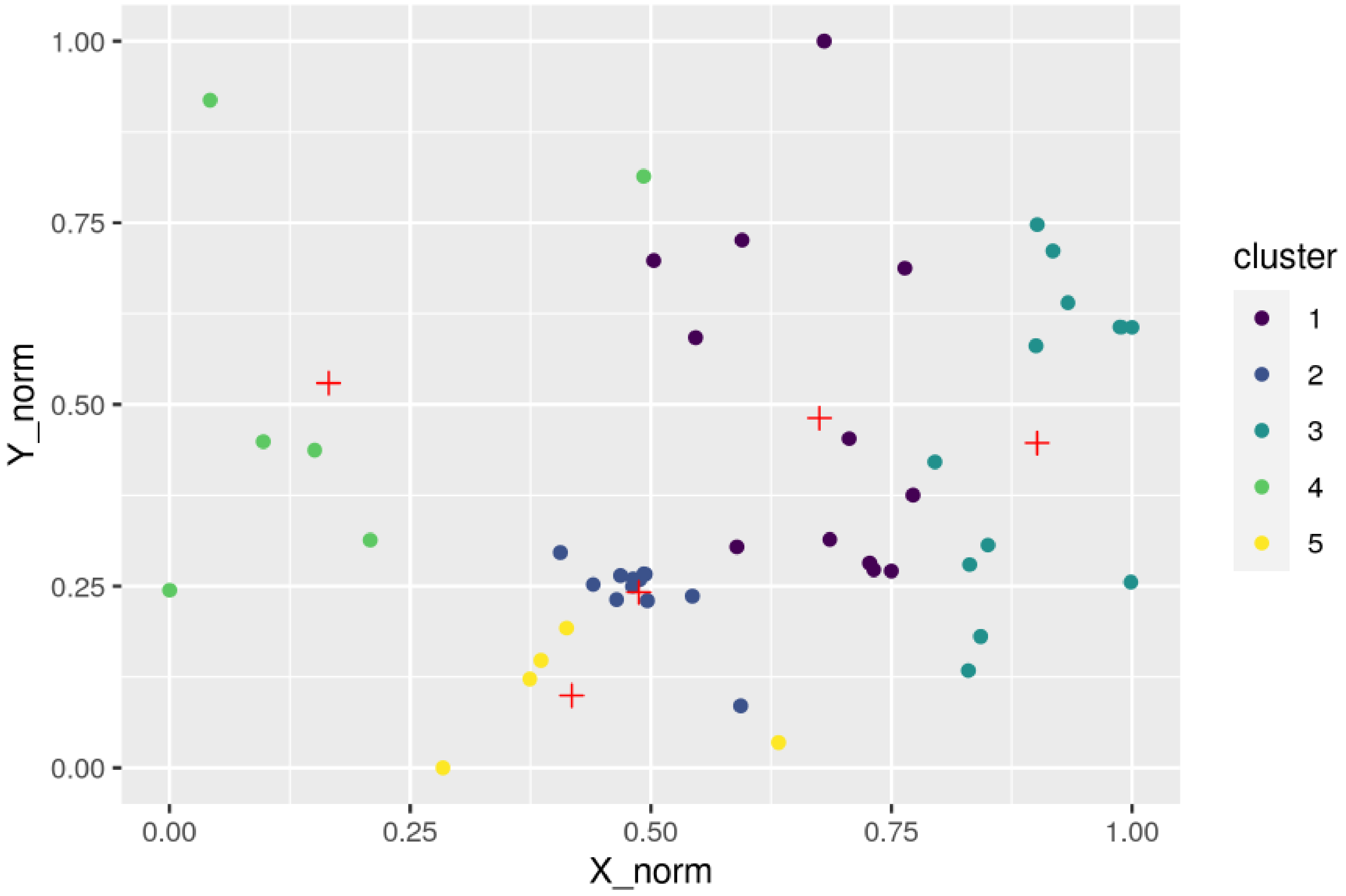
k-means clustering



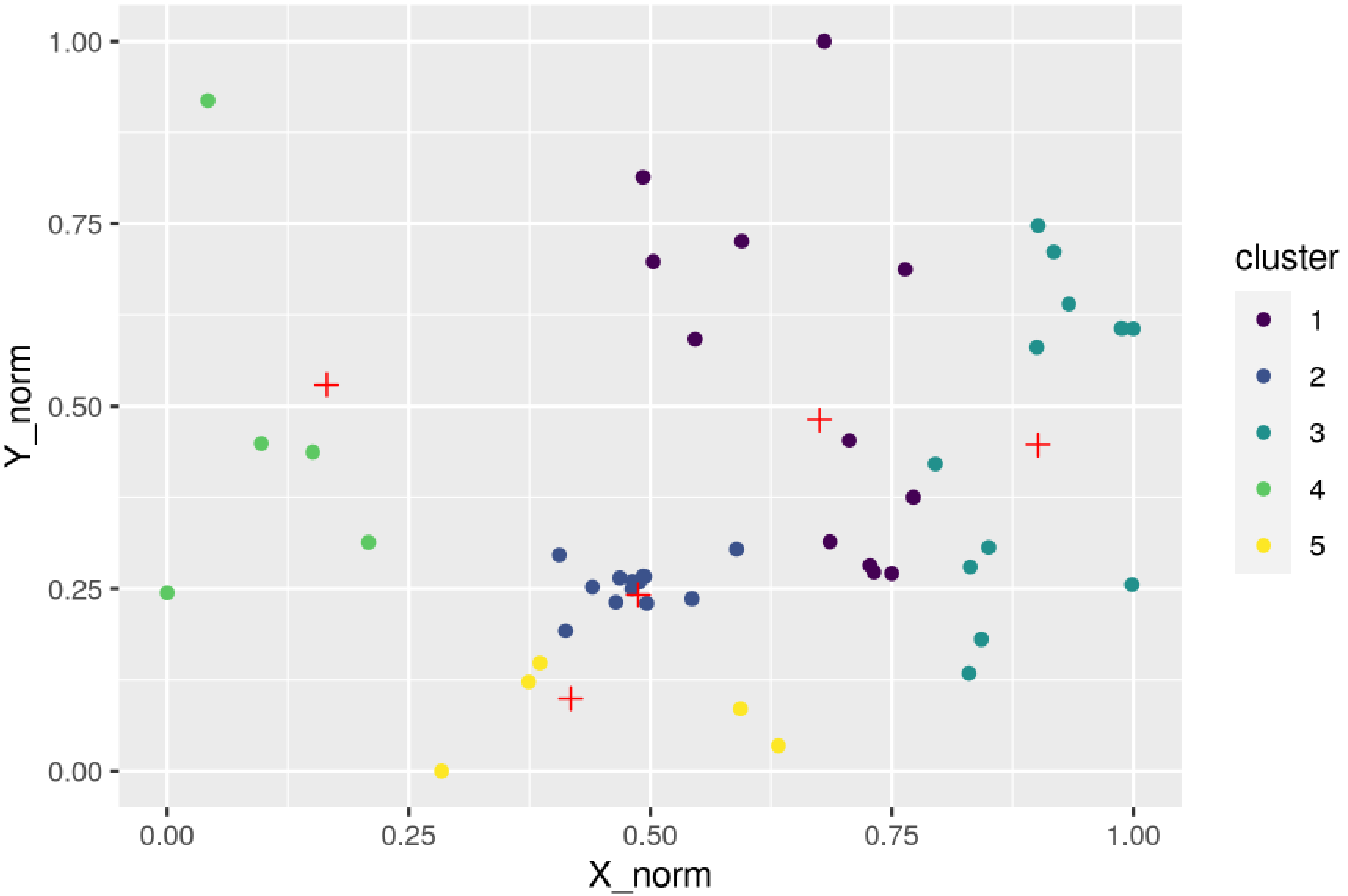
k-means clustering



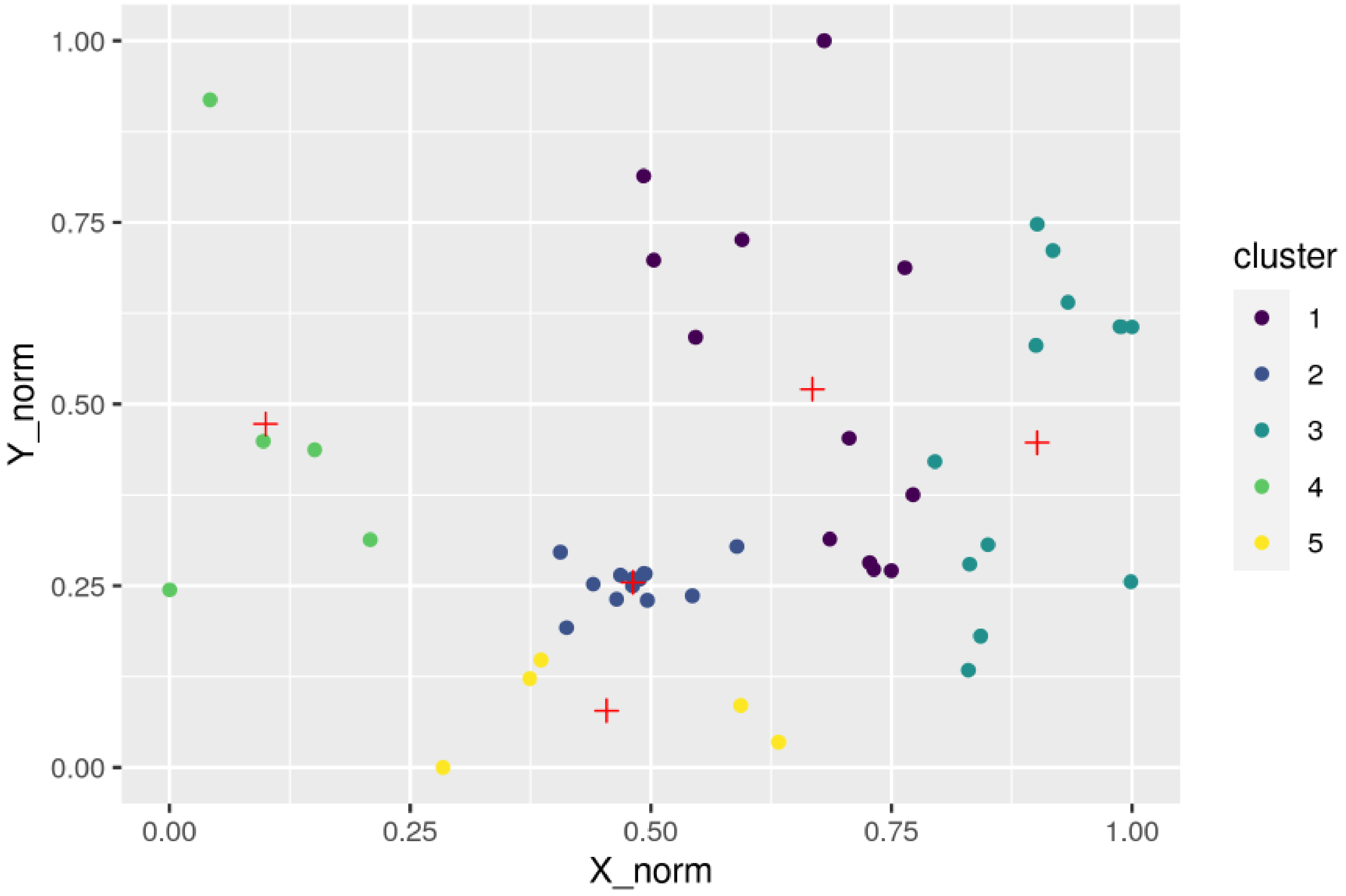
k-means clustering



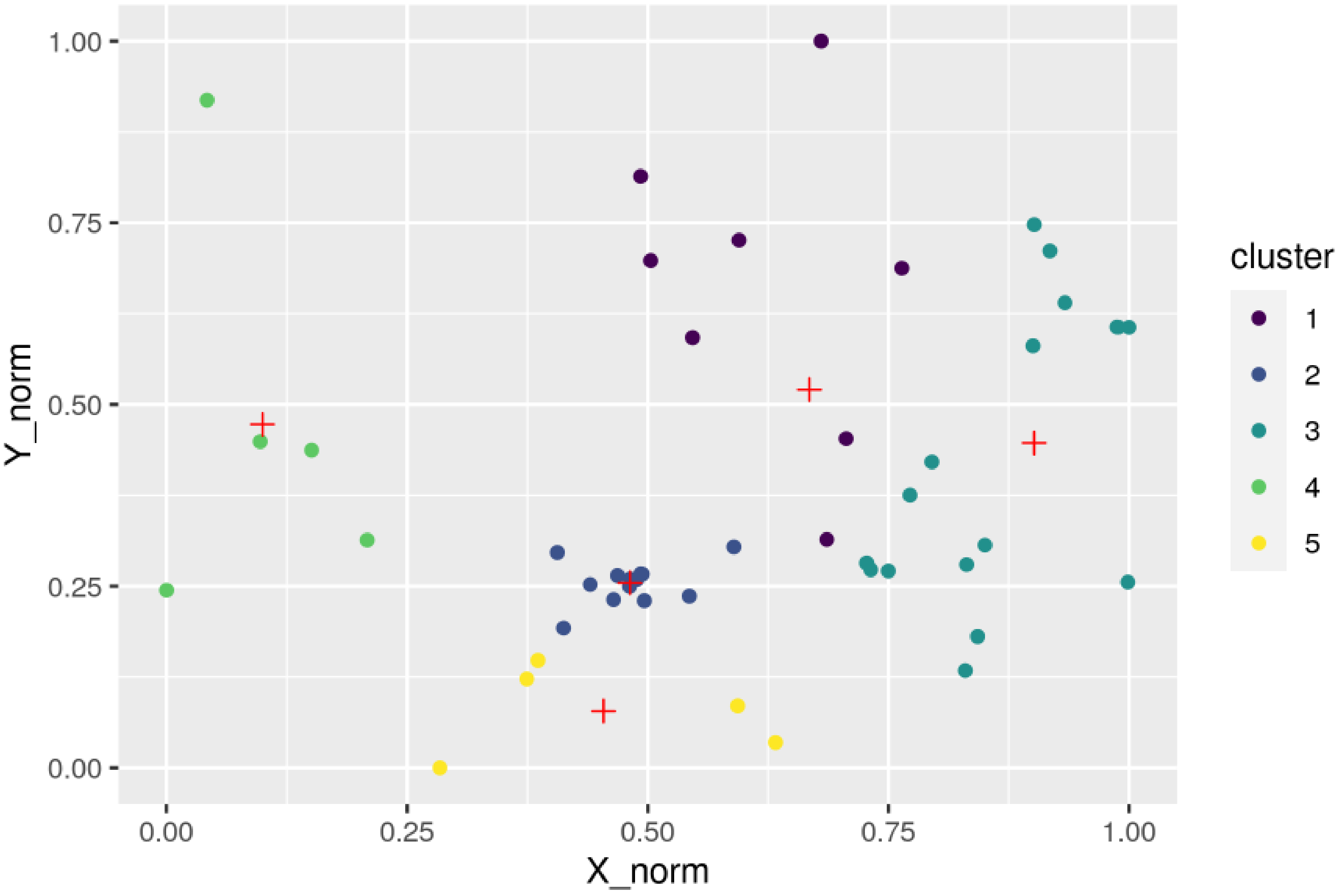
k-means clustering



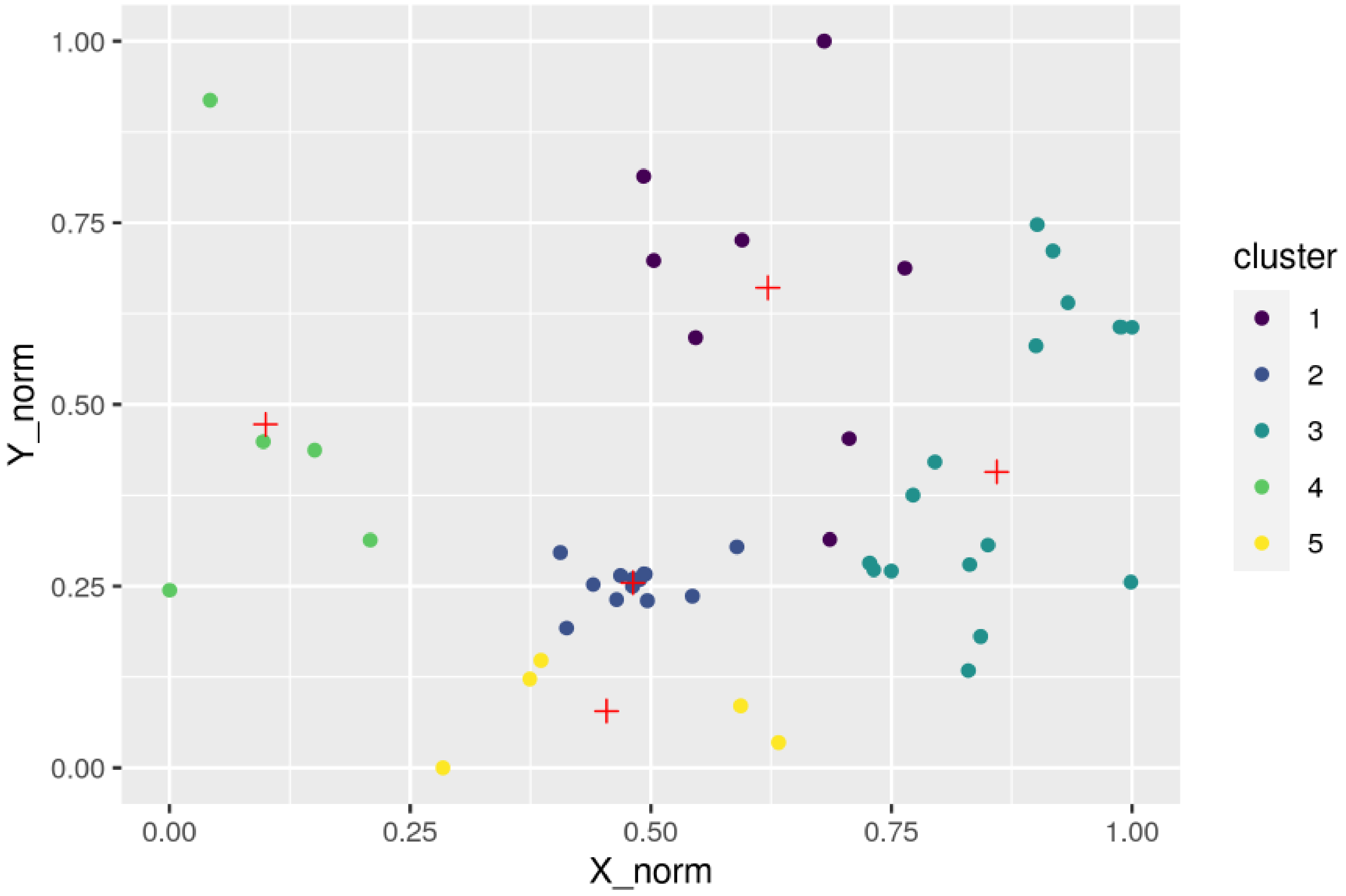
k-means clustering



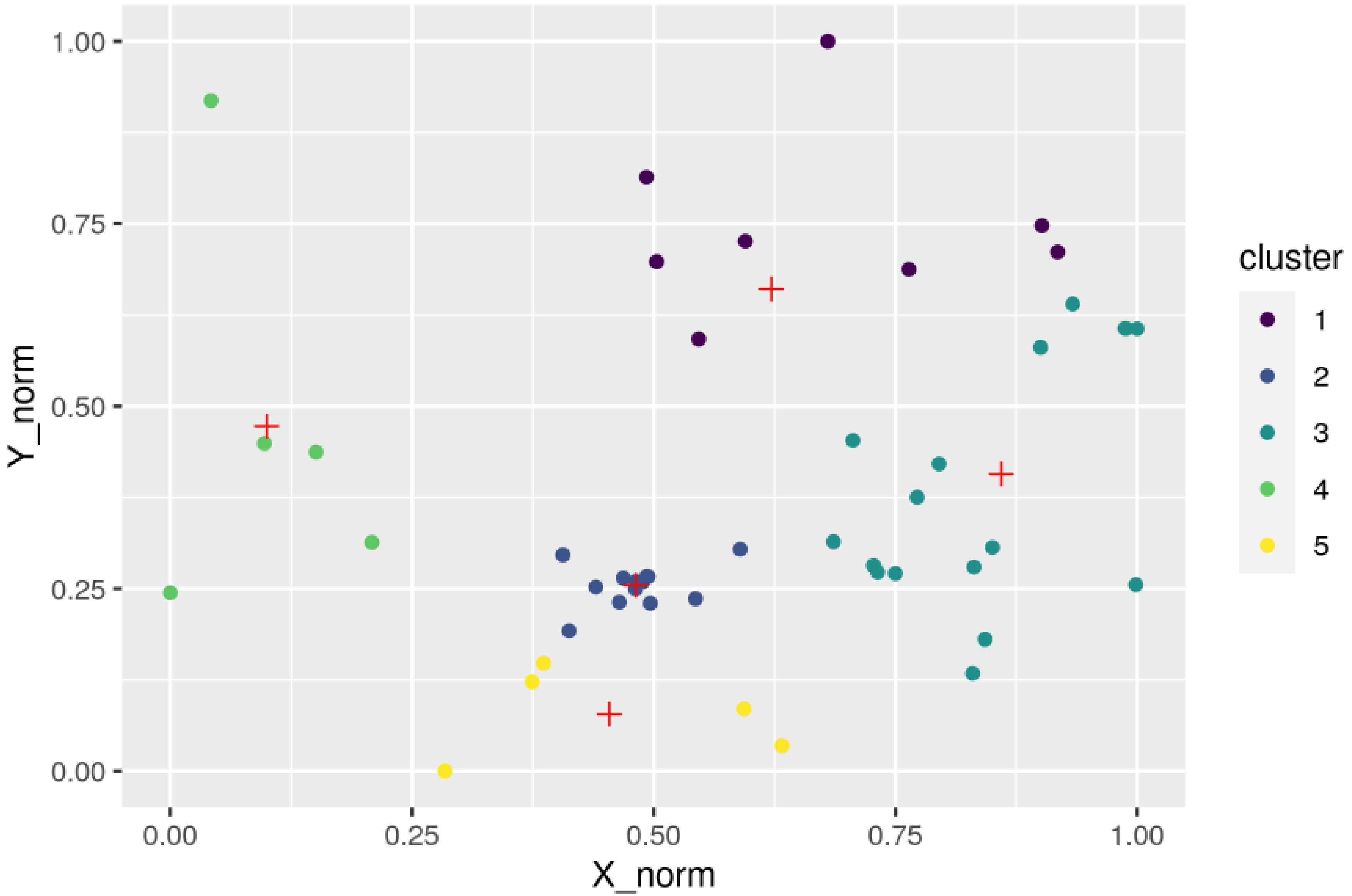
k-means clustering



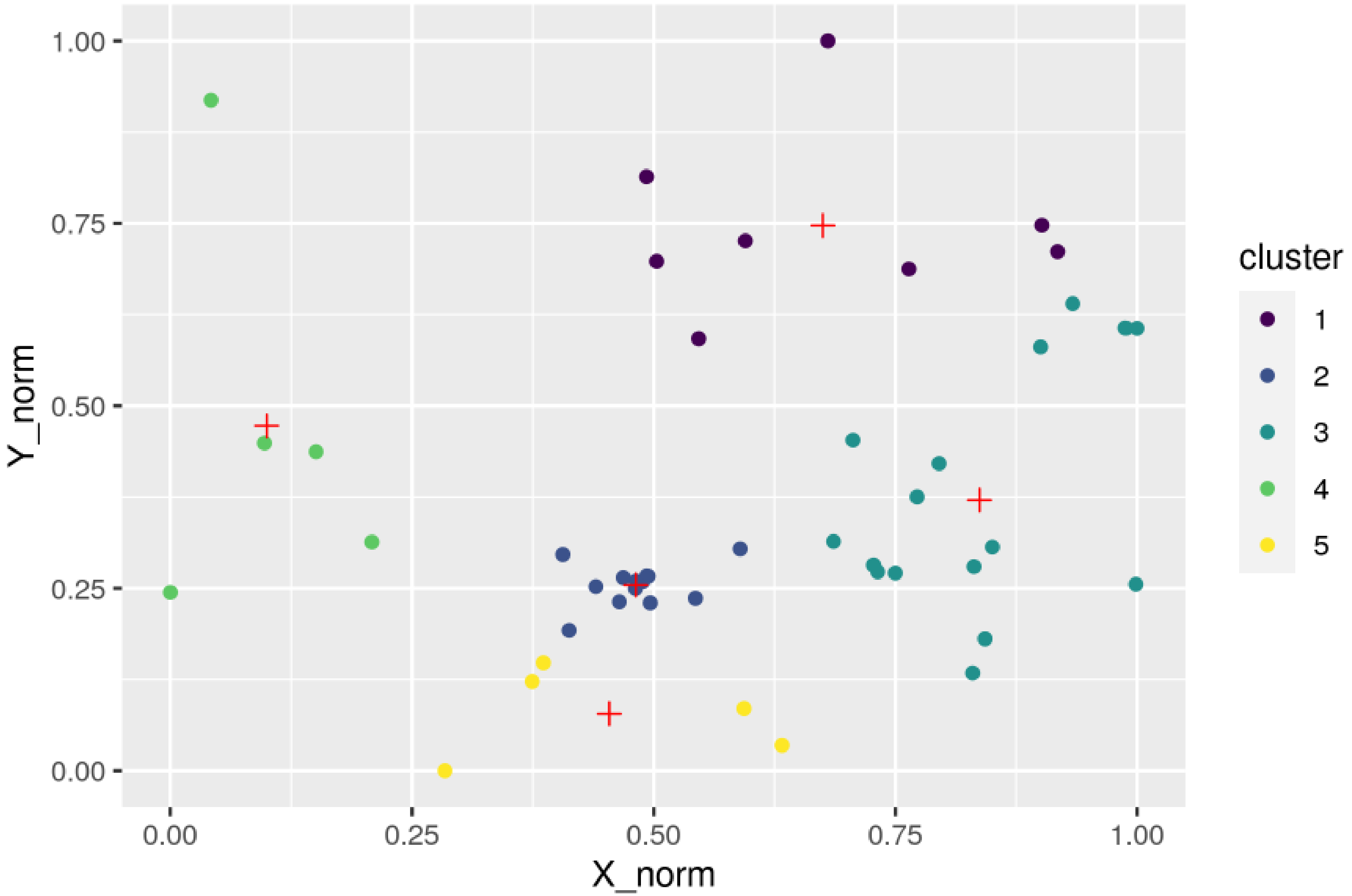
k-means clustering



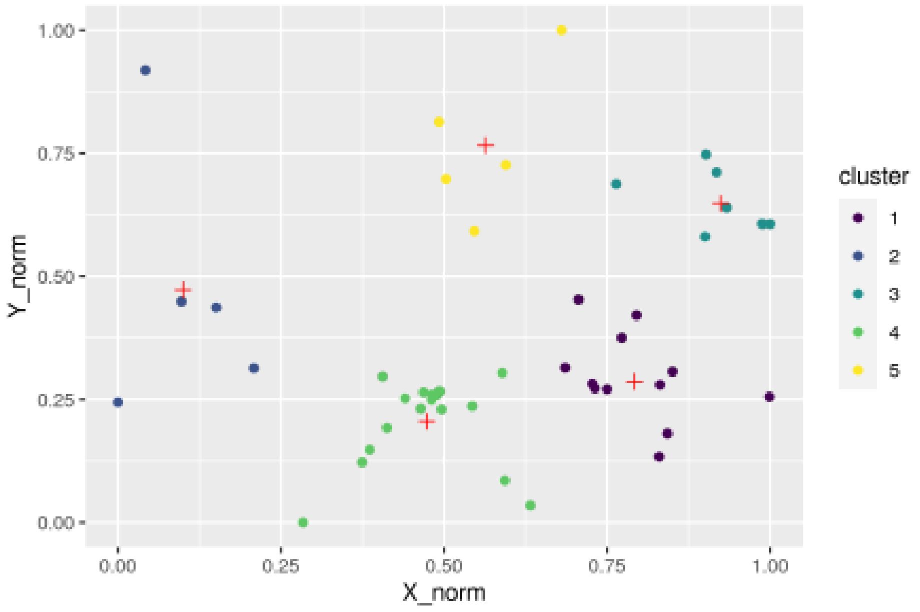
k-means clustering



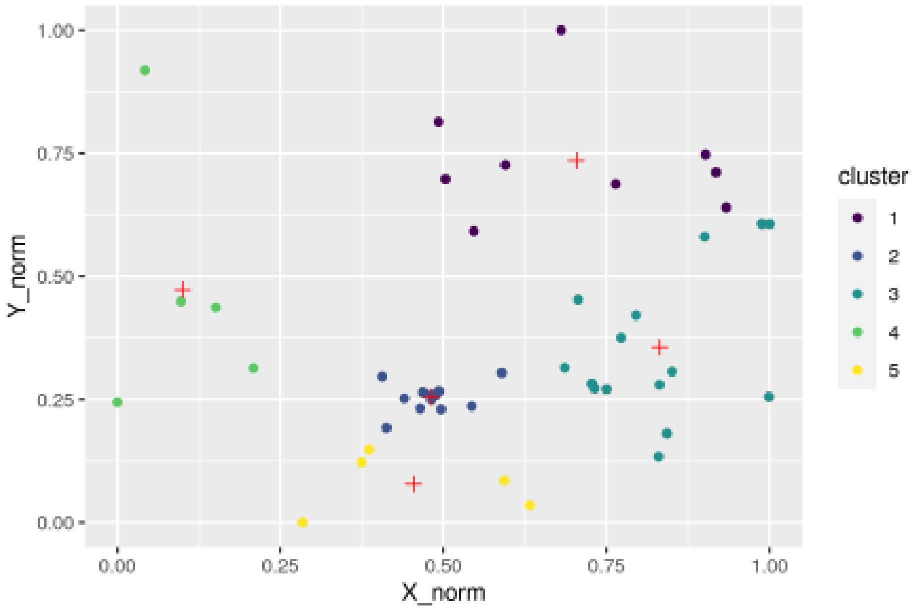
k-means clustering



k-means clustering
Outcome depends on initial conditions



previous initial centers



new initial centers

k-means clustering

- What do we mean by ‘centre’?
  Most often: *centroid* (mean coordinate along every axis)
  or *medoid* (‘median’ data point).
- Will this always converge?
  Ish. Most of the time it does, and quite quickly. You may need a way to break ties if a data point is equally far away from two centroids.
- Doesn’t it depend on what points are initially chosen?
  Yes! So, run several iterations and pick the best.

k-means clustering

- What do we mean by ‘centre’?
  Most often: *centroid* (mean coordinate along every axis)
  or *medoid* (‘median’ data point).
- Will this always converge?
  Ish. Most of the time it does, and quite quickly. You may need a way to break ties if a data point is equally far away from two centroids.
- Doesn’t it depend on what points are initially chosen?
  Yes! So, run several iterations and pick the best.
- But how do I know what ‘the best’ is?
  Various measures/objectives exists. For instance minimise within-cluster variance.

k-means clustering

- What do we mean by ‘centre’?
  Most often: *centroid* (mean coordinate along every axis)
  or *medoid* (‘median’ data point).
- Will this always converge?
  Ish. Most of the time it does, and quite quickly. You may need a way to break ties if a data point is equally far away from two centroids.
- Doesn’t it depend on what points are initially chosen?
  Yes! So, run several iterations and pick the best.
- But how do I know what ‘the best’ is?
  Various measures/objectives exists. For instance minimise within-cluster variance.
- How do I know what *k* I want?
  It depends. Practical considerations (what is the clustering for? How many clusters do you need/want?). Or try different *k* values, and use the best.

k-means clustering

- What do we mean by 'centre'?
  Most often: *centroid* (mean coordinate along every axis)
  or *medoid* ('median' data point).
- Will this always converge?
  Ish. Most of the time it does, and quite quickly. You may need a way to break ties if a data point is equally far away from two centroids.
- Doesn't it depend on what points are initially chosen?
  Yes! So, run several iterations and pick the best.
- But how do I know what 'the best' is?
  Various measures/objectives exists. For instance minimise within-cluster variance.
- How do I know what *k* I want?
  It depends. Practical considerations (what is the clustering for? How many clusters do you need/want?). Or try different *k* values, and use the best.
- But how do I know what 'the best' is?

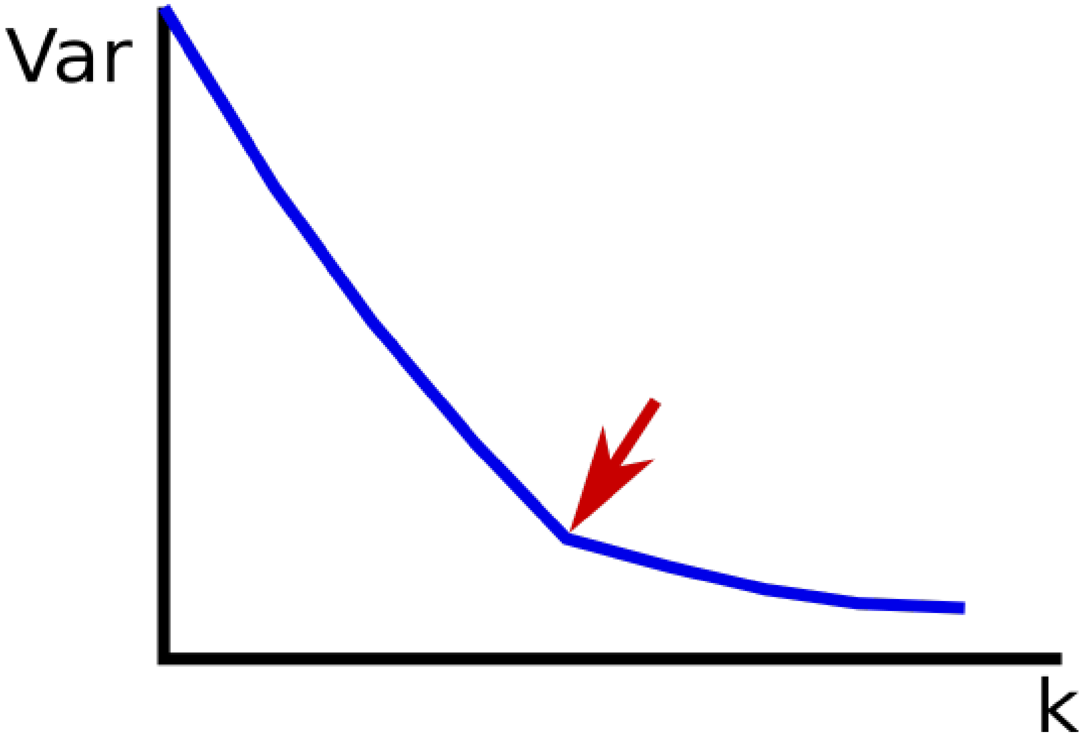
k-means clustering

Can we just minimise within-cluster variance to see what *k* is best?

k-means clustering

Can we just minimise within-cluster variance to see what *k* is best?

No! Then you would always just end up with *n* clusters. But what you can do is look at the *k* that yield the biggest improvement ('elbow plot').



k-means clustering

But what if I really, really don’t want to choose a *k*?

Outline

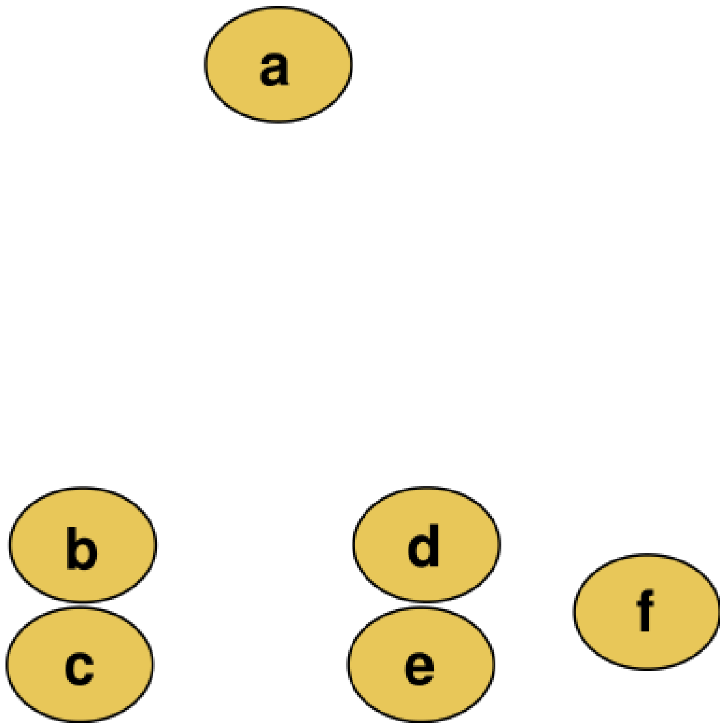
1. What is clustering?
2. k-means clustering
3. **Hierarchical clustering**
4. Machine learning

Hierarchical clustering

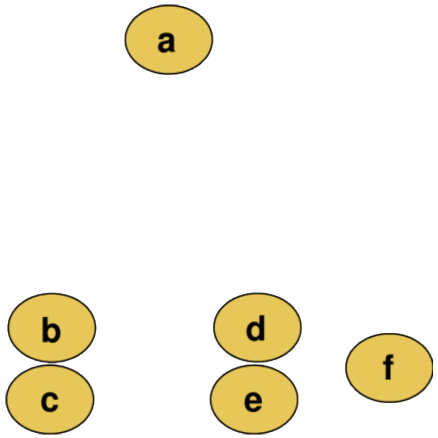
- Basic idea:

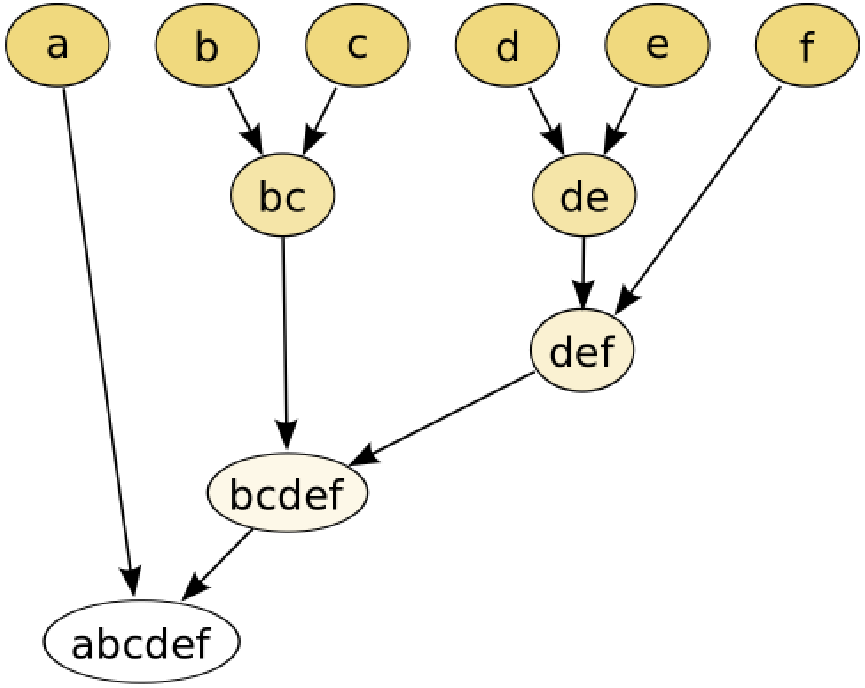
- No prior ideas about cluster sizes – we will just to *every* cluster size!
- Start by putting every individual data point in its own cluster
- **Group the two closest clusters** together in a common cluster
- Repeat until you have only one big cluster

Hierarchical clustering



Hierarchical clustering





Hierarchical clustering

What questions do you have?

Hierarchical clustering

What questions do you have?

- How do we define ‘closest together’?
- Does this outcome depend on initial choices, like for k-means?
- Why are we doing this?

Hierarchical clustering

- How do we define ‘closest together’?

Hierarchical clustering

- How do we define ‘closest together’?

- Several methods exist:
  - Shortest distance between points of two clusters (single-linkage clustering)
  - Longest distance between points of two clusters (complete-linkage clustering)
  - Mean distance between points of two clusters (average linkage cluster)
  - …

Hierarchical clustering

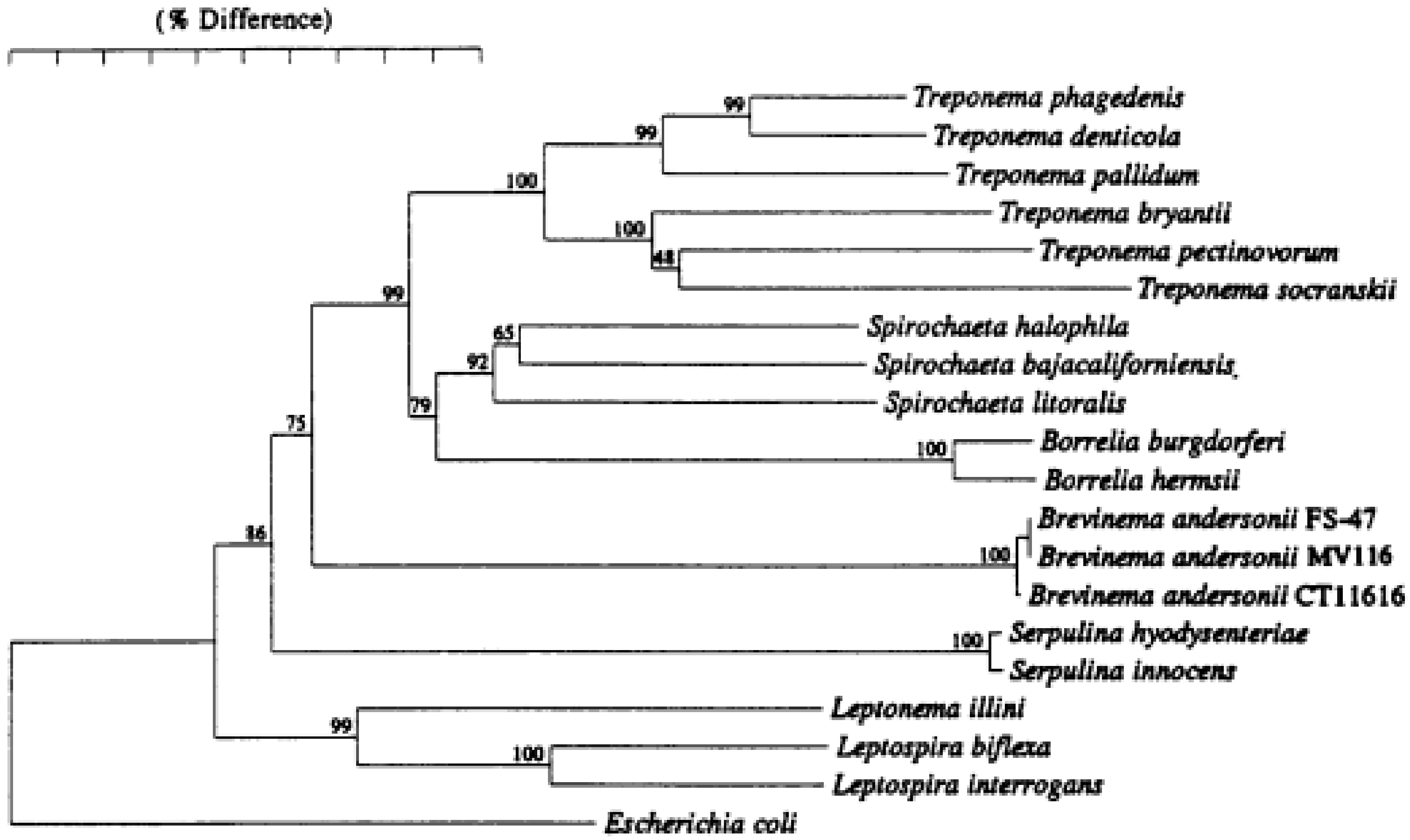
- Does the outcome depend on initial choices, like for k-means?
  No (apart from the distance measure chosen)

Hierarchical clustering

- Does the outcome depend on initial choices, like for k-means?
  No (apart from the distance measure chosen)

- Why are we doing this?
  It produces a phylogeny that is useful to choose possible cluster numbers, cluster sizes, etc. The phylogeny is also closely related to evolutionary analysis.

Hierarchical clustering



Hierarchical vs k-means clustering

Which one would you pick?

Outline

1. What is clustering?
2. k-means clustering
3. Hierarchical clustering
4. **Machine learning**

Machine learning: what is it?

- To learn something from the data given

Machine learning: what is it?

- To learn something from the data given
- Clustering is a king of machine learning – ‘unsupervised learning’
  - No *supervisors* needed
  - i.e. all we get is a bag of data points without any labels
  - e.g. 50 hospitals, a bunch of pictures

Machine learning: what is it?

- To learn something from the data given
- Clustering is a king of machine learning – ‘unsupervised learning’
  - No *supervisors* needed
  - i.e. all we get is a bag of data points without any labels
  - e.g. 50 hospitals, a bunch of pictures
- Another kind of machine learning – ‘supervised learning’
  - Do need a *supervisor* (training data)
  - i.e. we need to get some data points with given labels
  - e.g. 50 hospitals labelled with grades (‘Edinburgh’, ‘Glasgow’, …), pictures labelled as dog or cat

Remember Week 2.9 ‘Supervised and Unsupervised learning’ !

Machine learning: what is it?

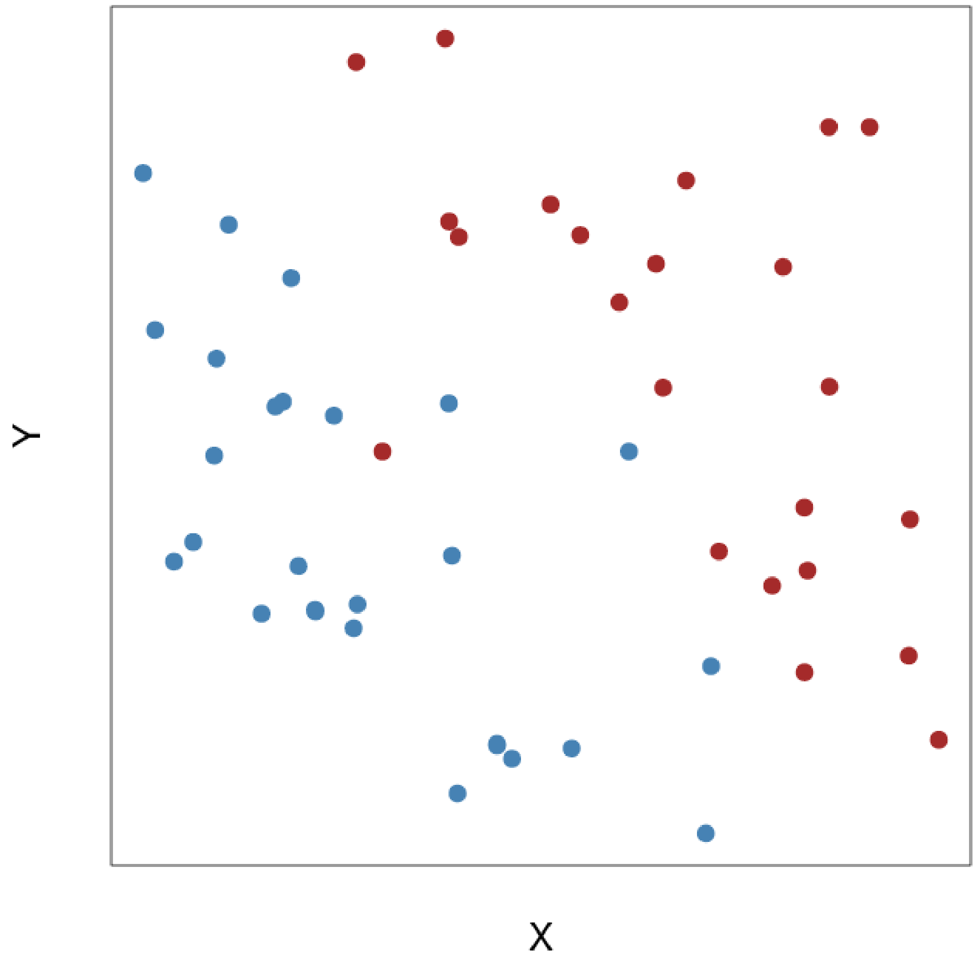
- To learn something from the data given
- Clustering is a king of machine learning – ‘unsupervised learning’
  - No *supervisors* needed
  - i.e. all we get is a bag of data points without any labels
  - e.g. 50 hospitals, a bunch of pictures
- Another kind of machine learning – ‘supervised learning’
  - Do need a *supervisor* (training data)
  - i.e. we need to get some data points with given labels
  - e.g. 50 hospitals labelled with grades (‘Edinburgh’, ‘Glasgow’, …), pictures labelled as dog or cat
- A third kind of machine learning – ‘reinforcement learning’
  - Also need some interactive *supervisor*
  - i.e. we get some feedback from the environment to adjust our actions
  - e.g. robot control in catching things, self-driving cars

Two types of supervised learning

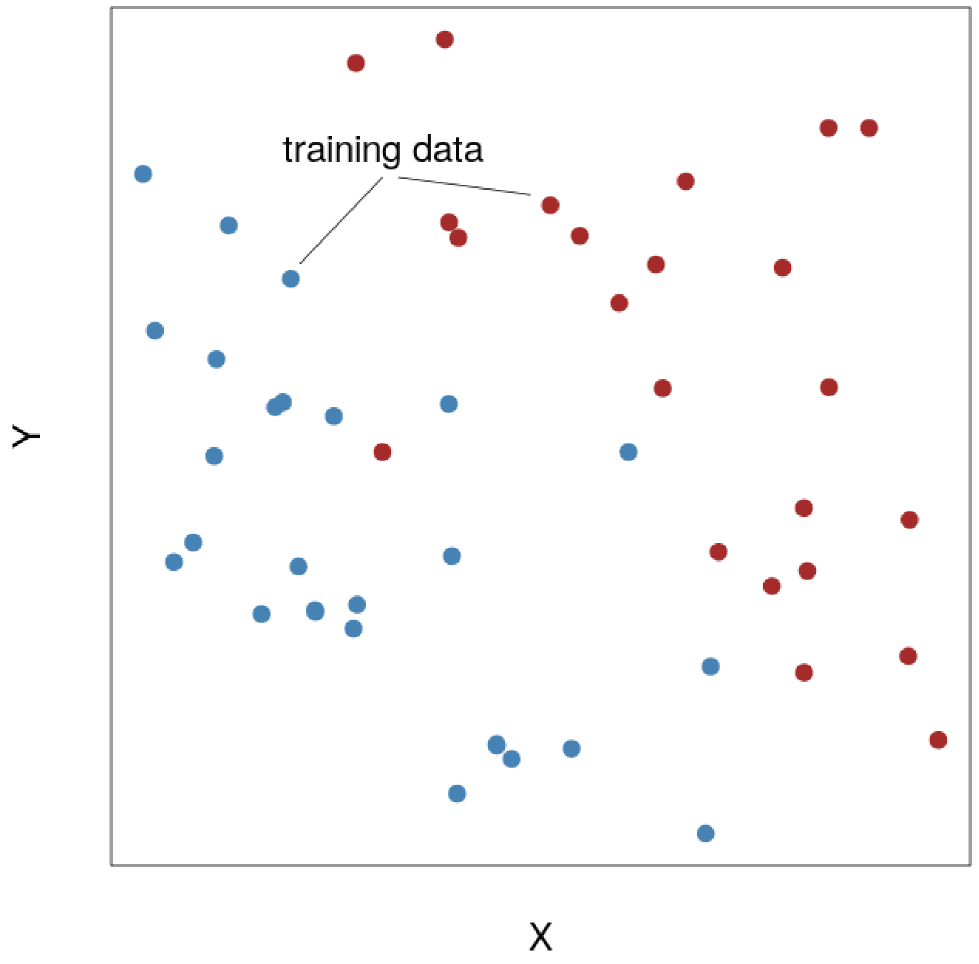
- Classification
  - The data are with discrete labels
  - e.g. 50 hospitals labelled with grades (‘Edinburgh’, ‘Glasgow’, …), pictures labelled as dog or cat
- Regression
  - The data are with continuous labels
  - e.g. 50 hospitals labelled with some score range from 0 to 100, patients’ blood pressure in some mmHg, patients’ temperature in some degrees Celsius

Remember Week 2.9 ‘Supervised and Unsupervised learning’ !

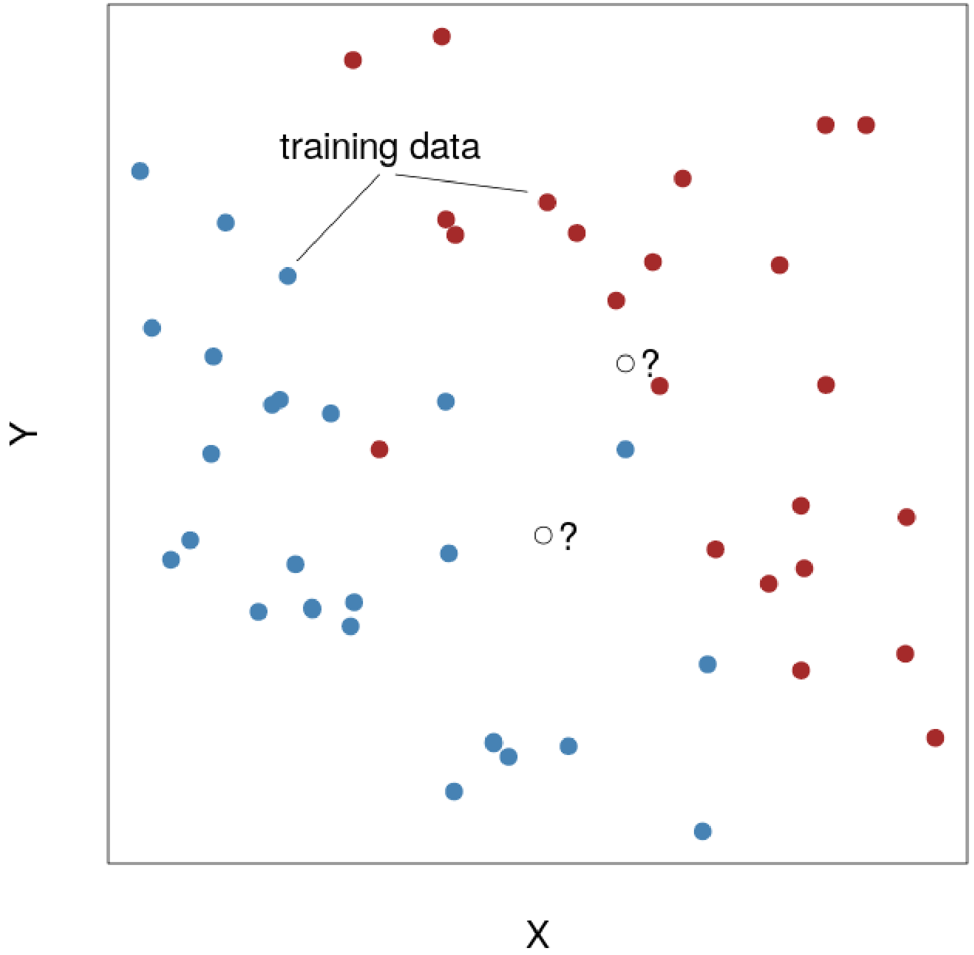
Example: classification



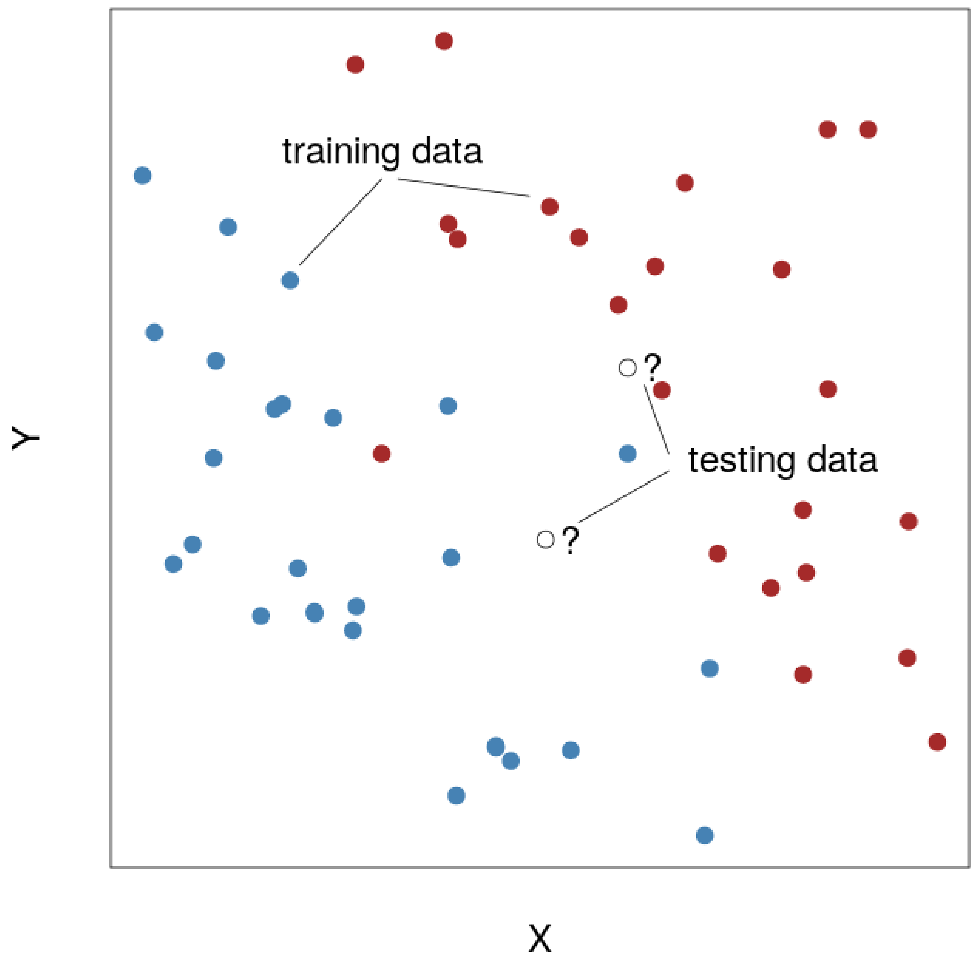
Example: classification



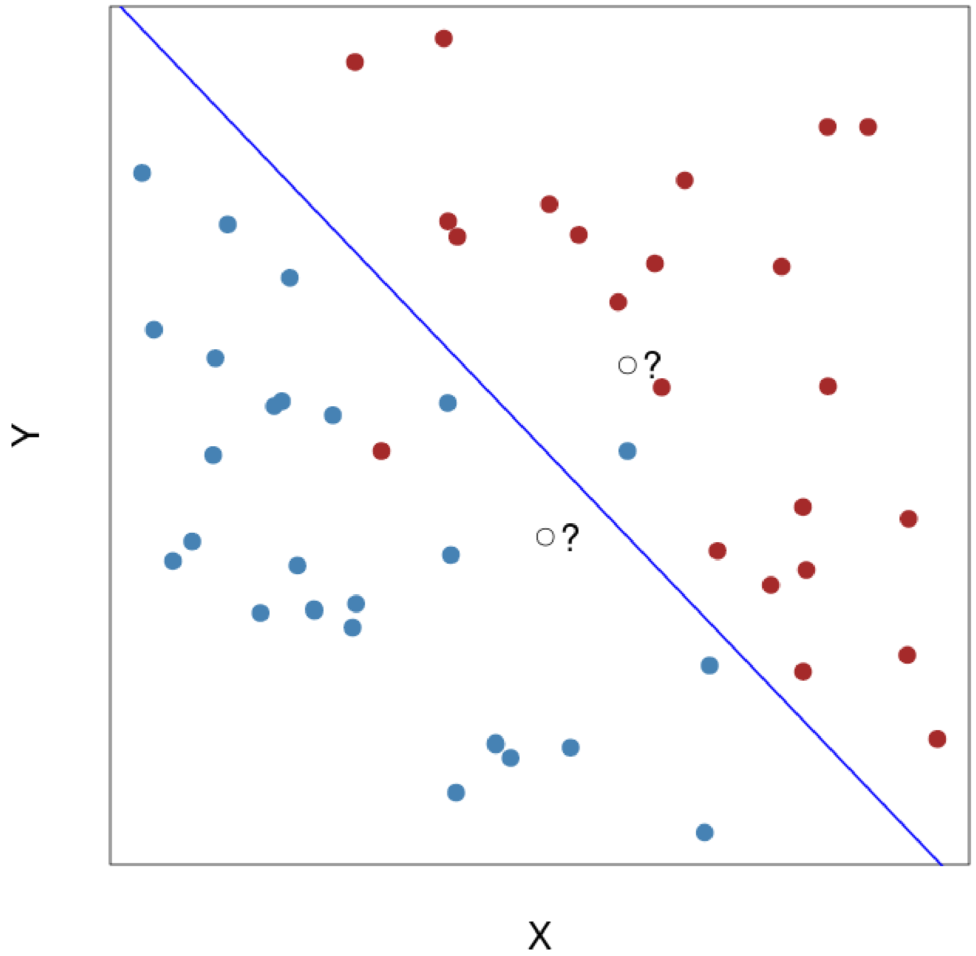
Example: classification



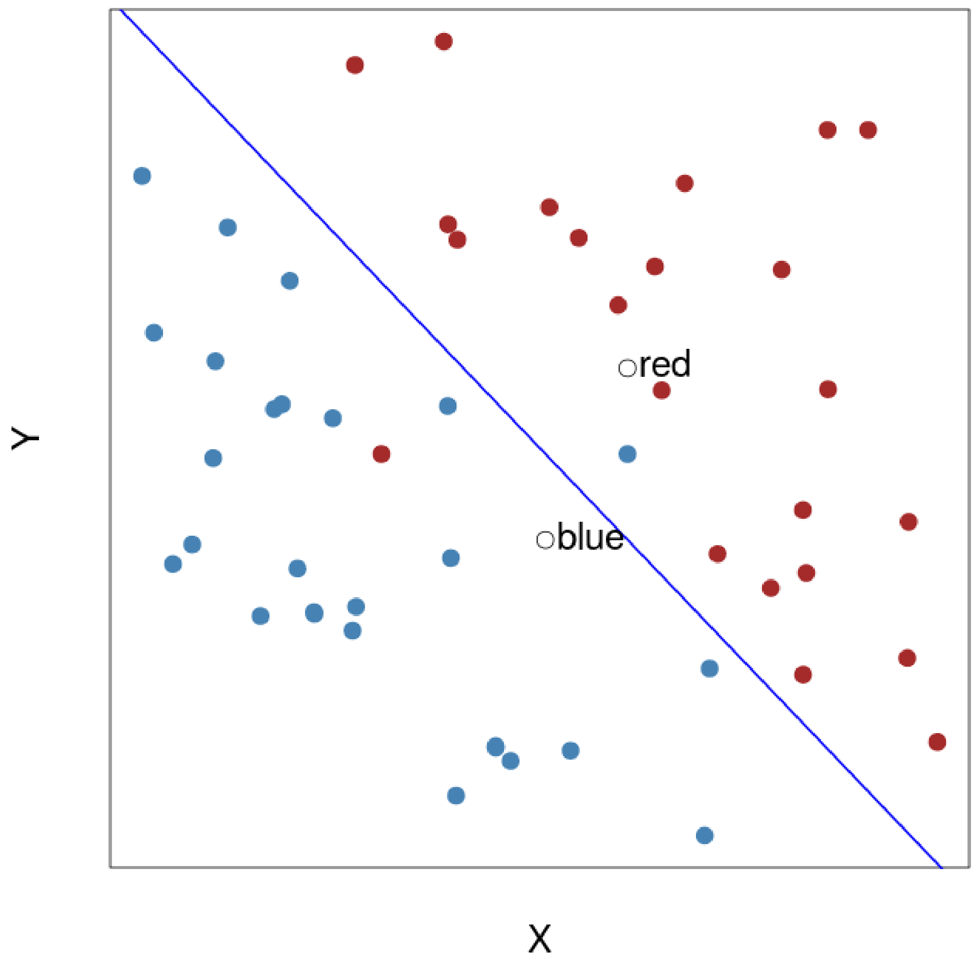
Example: classification



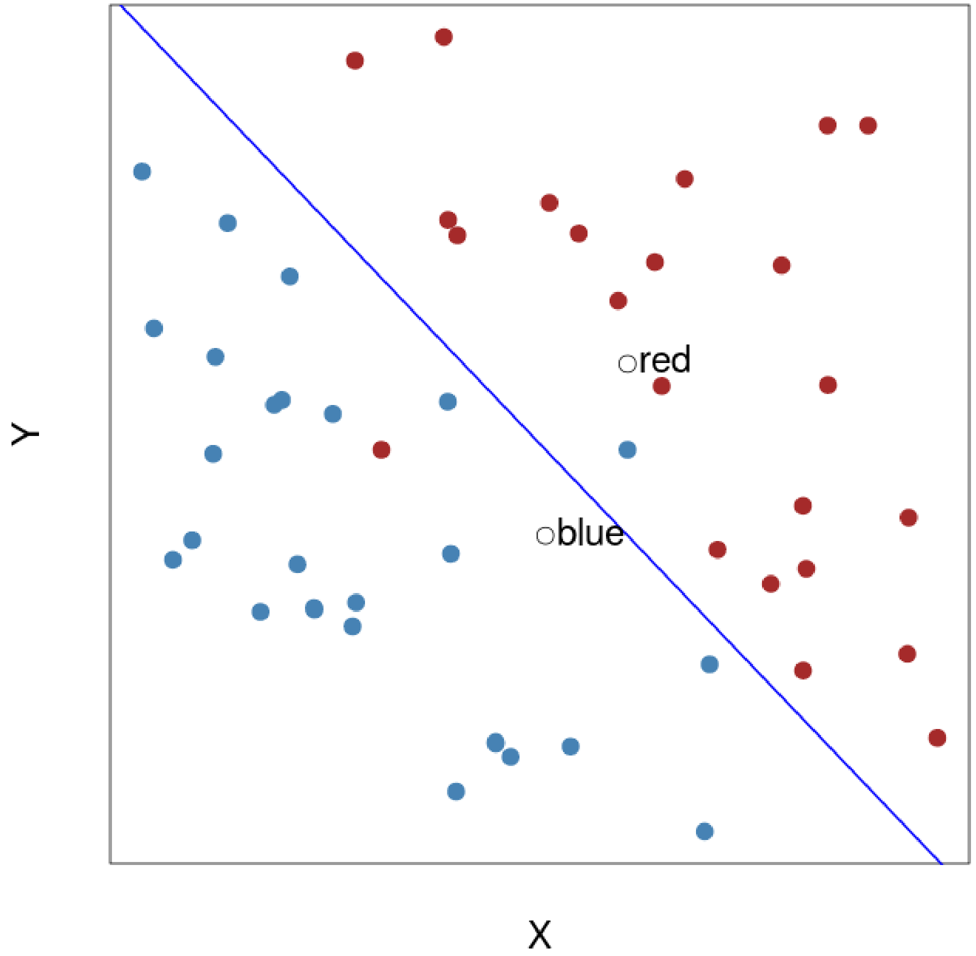
Example: classification



Example: classification



Example: classification



- Actually, this model is not perfect. One red point would be misclassified as blue and one blue point would be misclassified as red.

Measuring the quality of a classification model

- In a classification problem with two labels, we often refer to one label as ‘positive’ and the other ‘negative’.
- Suppose the red points are the ‘positive’ group, and the blue points the ‘negative’ group.

	Known red (positive)	Known blue (negative)
Classified as red (positive)	True positive (TP)	False positive (FP)
Classified as blue (negative)	False negative (FN)	True negative (TN)

- What are the TP, FP, FN, and TN in our classification example?

Measuring the quality of a classification model

- Proportion of data correctly classified:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

- Proportion of positive data correctly classified

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

- Proportion of negative data correctly classified

$$\text{Specificity} = \frac{TN}{TN + FP}$$

- Proportion of true positive in the data classified as ‘positive’

$$\text{Positive Predictive Value (PPV)} = \frac{TP}{TP + FP}$$

- What are these measures in our classification example?

Learning objectives

Now, you should be able to:

- Explain how clustering works (k-means, hierarchical clustering)
- Discuss the choices that need to be made when tackling a clustering problem
- Explain how supervised machine learning works (classification)
- How to simply evaluate a classification method

浙江大学爱丁堡大学联合学院

ZJU-UoE Institute

Clustering and Machine Learning
Any questions?

ADS2, Lecture 2.13

Dr Rob Young – robert.young@ed.ac.uk

Semester 2, 2023/24