



[Unit 4 Unsupervised Learning \(2](#)

[Course](#) > [weeks](#))

> [Lecture 14. Clustering 2](#) >

5. Determining the Number of  
Clusters

**Audit Access Expires May 11, 2020**

You lose all access to this course, including your progress, on May 11, 2020.

## 5. Determining the Number of Clusters

## Determining the Number of Clusters



### Video

[Download video file](#)

### Transcripts

[Download SubRip \(.srt\) file](#)

[Download Text \(.txt\) file](#)

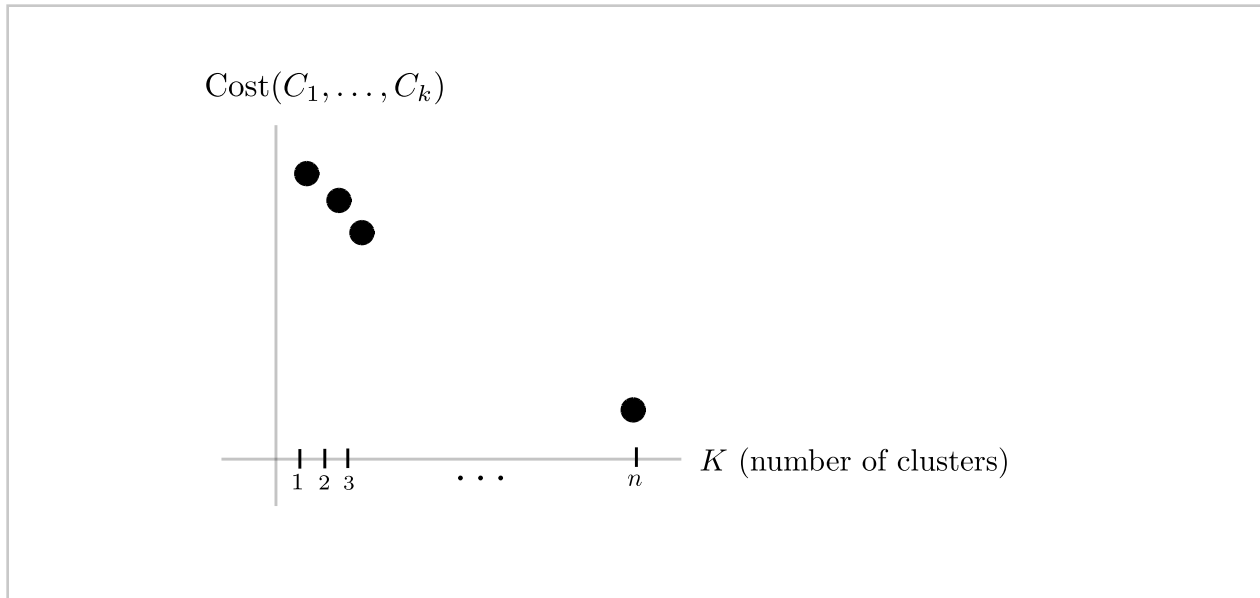
---

## Image Quantization Example

1/1 point (graded)

Remember that in our the first clustering lecture, the professor discussed how clustering can be used in image quantization. In short, by clustering many different colors into a few clusters, we can save(compress) the number of bits used to denote different colors.

The picture below is a general trend between the number of clusters  $K$  and the total cost of clustering.



We expect the clustering cost to decrease as we increase the number of clusters  $K$ . As mentioned in the lecture, the case of  $K = n$  is when every point is its own cluster.

In the context of image quantization, what is the problem with every point being its own cluster?

- ☐ Cost of clustering is too low
- ☐ The time complexity is too large when  $K = n$
- ☒ There is no compression at all



### Solution:

When  $K = n$ , every point is its own cluster, which means that there is no real clustering happening. (In other words, there is no point of doing this clustering) The number of representatives is equal to the number of original colors, so no bits can be saved/compressed.

**Submit**

You have used 1 of 2 attempts

 Answers are displayed within the problem

## Supervised Elements of Unsupervised Learning

1/1 point (graded)

Remember that clustering is an example of unsupervised learning. However, unlike its name suggests, there are some elements that we can "supervise" in clustering. In other words, there are some parts clustering that needs to be determined or "tuned" by us, depending on the application. Which of the following are elements of clustering that we can and should tune, depending on the application?

☒ Number of clusters  $K$ ☐  $C_1, \dots, C_K$ , the partition of  $x_1, \dots, x_n$ ☐  $z_1, \dots, z_K$ , the representatives of each of the  $K$  clusters☒ The cost measure for distance between  $x^{(i)} \in C_j$  and  $z_j$  ( $\text{dist}(x^{(i)}, z_j)$ )

### Solution:

The number of clusters  $K$  and The cost measure for distance between  $x^{(i)} \in C_j$  and  $z_j$  ( $\text{dist}(x^{(i)}, z_j)$ ) are what we can control/ supervise. As mentioned in the lecture, different dataset/ application context require different cost measures( $\text{dist}(x^{(i)}, z_j)$ ). Also, different contexts require different optimum number of clusters. These are where you apply your domain knowledge on the application.

**Submit**

You have used 1 of 2 attempts


 Answers are displayed within the problem

## Discussion

[Hide Discussion](#)

**Topic:** Unit 4 Unsupervised Learning (2 weeks) :Lecture 14.  
Clustering 2 / 5. Determining the Number of Clusters

[Add a Post](#)[Show all posts](#)[by recent activity](#)

 [Using classification performance to choose the value of K](#) 3  
[After appending the distances to the centroids, each data point representation is changed, so...](#)

© All Rights Reserved