

1. NMF learns interpretable parts

In this video, you'll learn that the components of NMF represent patterns that frequently occur in the samples.

2. Example: NMF learns interpretable parts

Let's consider a concrete example, where scientific articles are represented by their word frequencies. There are 20000 articles, and 800 words. So the array has 800 columns.

3. Applying NMF to the articles

Let's fit an NMF model with 10 components to the articles. The 10 components are stored as the 10 rows of a 2-dimensional numpy array.

4. NMF components are topics

The rows, or components, live in an 800-dimensional space - there is one dimension for each of the words. Aligning the words of our vocabulary with the columns of the NMF components allows them to be interpreted.

5. NMF components are topics

Choosing a component, such as this one, and looking at which words have the highest values,

6. NMF components are topics

we see that they fit a theme: the words are 'species', 'plant', 'plants', 'genetic', 'evolution' and 'life'.

7. NMF components are topics

The same happens if any other component is considered.

8. NMF components

So if NMF is applied to documents, then the components correspond to topics, and the NMF features reconstruct the documents from the topics. If NMF is applied to a collection of images, on the other hand, then the NMF components represent patterns that frequently occur in the images. In this example, for instance, NMF decomposes images from an LCD display into the individual cells of the display. This example you'll investigate for yourself in the exercises. To do this, you'll need to know how to represent a collection of images as a non-negative array.

9. Grayscale images

An image in which all the pixels are shades of gray ranging from black to white is called a "grayscale image". Since there are only shades of grey, a grayscale image can be encoded by the brightness of every pixel. Representing the brightness as a number between 0 and 1, where 0 is totally black and 1 is totally white, the image can be represented as 2-dimensional array of numbers.

10. Grayscale image example

Here, for example, is a grayscale photo of the moon!

11. Grayscale images as flat arrays

These 2-dimensional arrays of numbers can then be flattened by enumerating the entries. For instance, we could read-off the values row-by-row, from left-to-right and top to bottom.

12. Grayscale images as flat arrays

The grayscale image is now represented by a flat array of non-negative numbers.

13. Encoding a collection of images

A collection of grayscale images of the same size can thus be encoded as a 2-dimensional array, in which each row represents an image as a flattened array, and each column represents a pixel. Viewing the images as samples, and the pixels as features, we see that the data is arranged similarly to the word frequency array. Indeed, the entries of this array are non-negative, so NMF can be used to learn the parts of the images.

14. Visualizing samples

It's difficult to visualize an image by just looking at the flattened array. To recover the image, use the reshape method of the sample, specifying the dimensions of the original image as a tuple. This yields the 2-dimensional array of pixel brightnesses. To display the corresponding image, import pyplot, and pass the 2-dimensional array to the plt dot imshow function.

15. Let's practice!

In this video, you've seen how NMF components can be interpreted as patterns that frequently occur in the samples. It's time to get some practice and investigate this phenomenon for yourself.