CS464 Introduction to Machine Learning

Homework 2 Report

Ece Kahraman

21801879

1. PCA & Cats

1.1

The output of this part is given below as it is.

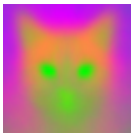
```
For the red channel:

-with 1 principal components PVE is 23.507%
-with 2 principal components PVE is 39.158%
-with 3 principal components PVE is 48.163%
-with 4 principal components PVE is 54.993%
-with 5 principal components PVE is 58.746%
-with 6 principal components PVE is 61.141%
-with 7 principal components PVE is 63.417%
-with 8 principal components PVE is 65.530%
-with 9 principal components PVE is 67.324%
-with 10 principal components PVE is 68.673%
We need to choose the first 12 principal components to achieve at
least 70% PVE.

For the green channel:

-with 1 principal components PVE is 20.874%
-with 2 principal components PVE is 36.758%
-with 3 principal components PVE is 46.017%
-with 4 principal components PVE is 52.828%
-with 5 principal components PVE is 56.627%
-with 6 principal components PVE is 59.073%
-with 7 principal components PVE is 61.501%
-with 8 principal components PVE is 63.650%
-with 9 principal components PVE is 65.537%
-with 10 principal components PVE is 66.959%
We need to choose the first 13 principal components to achieve at
least 70% PVE.

For the blue channel:

-with 1 principal components PVE is 22.859%
-with 2 principal components PVE is 38.508%
-with 3 principal components PVE is 47.299%
-with 4 principal components PVE is 53.502%
-with 5 principal components PVE is 57.243%
-with 6 principal components PVE is 59.659%
```

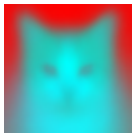
```
-with 7 principal components PVE is 62.064%
-with 8 principal components PVE is 64.124%
-with 9 principal components PVE is 65.969%
-with 10 principal components PVE is 67.398%
We need to choose the first 13 principal components to achieve at
least 70% PVE.
```

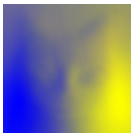
As we choose more values as the principal components, the PVE of the channel increases. Red channel has the highest PVEs for every PCs. Blue channel comes second, and the green channel comes last with lowest PVEs. The difference between the last two channels is at most 0.5. Because the red channel already has higher PVEs, it needs less principal components to reach 70%, while the other two need more.

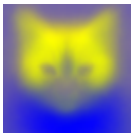
1.2

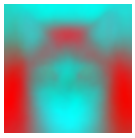
The outputs of the reconstructed pictures of each of the first 10 principal values are given below as they are.

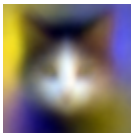

PC 1


PC 2

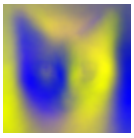

PC 3


PC 4

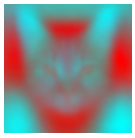

PC 5


PC 6


PC 7


PC 8


PC 9


PC 10

The reconstructed images are still able to convey the shapes of a cat. However, the details are blurry, and the colors are separated to different images and concentrated on different images. These principal components are the core features that are representing 5653 cat images. They will bring the computation costs down.

1.3

While reconstructing the “flickr_cat_000003.jpg”, first the color channels were resized to 64x64 arrays. They were centered by subtracting their means. Then, the dot product of these images and the eigenvectors of the first k values. And lastly, the means of the channels were added back to get the original images back. The outputs of this question are given below as they are.

k=1 k=50 k=250 k=500 k=1000 k=4069

When there is only one principal component, while some features of a cat are still visible, like its eyes, nose, and ears, the colors are completely off and the position of the cat tells nothing about the original image. As we increase the number of principal components, the colors blur together to match the colors of the original image, while the position of the cat turns clear. In the last image, every feature of the image is treated as a principal component. While the picture is very clear and understandable, it was the slowest one to produce. These images draw attention to the carefully chosen principal components number. Too many PCs have the danger of slow computation, while too little PCs have the danger of losing information.