2022-2023 Spring Semester

Homework 2

Course: CS464

Section: 1

Name: Alper

Surname: Mumcular

Student ID: 21902740

Date: 29 April 2023

1) PCA & Cats

Question 1.1 - PCA

```
Red PVE   (1) = 0.2350696993627993      Cumulative Red PVE   ----> 0.2350696993627993
Green PVE (1) = 0.20873714854025432     Cumulative Green PVE ----> 0.20873714854025432
Blue PVE  (1) = 0.22859035906545536     Cumulative Blue PVE  ----> 0.22859035906545536

Red PVE   (2) = 0.1565111520673864      Cumulative Red PVE   ----> 0.39158085143018573
Green PVE (2) = 0.15884565962402095     Cumulative Green PVE ----> 0.36758280816427524
Blue PVE  (2) = 0.15649257925344162     Cumulative Blue PVE  ----> 0.38508293831889695

Red PVE   (3) = 0.09005253857044315     Cumulative Red PVE   ----> 0.4816333900006289
Green PVE (3) = 0.09258856862586815     Cumulative Green PVE ----> 0.4601713767901434
Blue PVE  (3) = 0.08790595575693261     Cumulative Blue PVE  ----> 0.47298889407582956

Red PVE   (4) = 0.06829954682854636     Cumulative Red PVE   ----> 0.5499329368291752
Green PVE (4) = 0.06811111746109606     Cumulative Green PVE ----> 0.5282824942512395
Blue PVE  (4) = 0.06203548174652188     Cumulative Blue PVE  ----> 0.5350243758223514

Red PVE   (5) = 0.037527339511068306    Cumulative Red PVE   ----> 0.5874602763402436
Green PVE (5) = 0.0379850527572437      Cumulative Green PVE ----> 0.5662675470084833
Blue PVE  (5) = 0.03740134203265638     Cumulative Blue PVE  ----> 0.5724257178550078

Red PVE   (6) = 0.02394753913490536     Cumulative Red PVE   ----> 0.6114078154751489
Green PVE (6) = 0.024467317448625132    Cumulative Green PVE ----> 0.5907348644571084
Blue PVE  (6) = 0.024165873862426374    Cumulative Blue PVE  ----> 0.5965915917174343

Red PVE   (7) = 0.022764658780062664    Cumulative Red PVE   ----> 0.6341724742552115
Green PVE (7) = 0.024279163413091297    Cumulative Green PVE ----> 0.6150140278701998
Blue PVE  (7) = 0.024047333972009102    Cumulative Blue PVE  ----> 0.6206389256894433

Red PVE   (8) = 0.021128209465633447    Cumulative Red PVE   ----> 0.6553006837208449
Green PVE (8) = 0.02149052826450393     Cumulative Green PVE ----> 0.6365045561347037
Blue PVE  (8) = 0.020596134585636246    Cumulative Blue PVE  ----> 0.6412350602750796

Red PVE   (9) = 0.017935920584406427    Cumulative Red PVE   ----> 0.6732366043052513
Green PVE (9) = 0.018870002898966152    Cumulative Green PVE ----> 0.6553745590336698
Blue PVE  (9) = 0.01845899436619506     Cumulative Blue PVE  ----> 0.6596940546412746

Red PVE   (10) = 0.01349360904861359    Cumulative Red PVE   ----> 0.686730213353865
Green PVE (10) = 0.014211335203650746   Cumulative Green PVE ----> 0.6695858942373205
Blue PVE  (10) = 0.01428572007428644    Cumulative Blue PVE  ----> 0.673979774715561

PVE sums for the first 10 PC
Red   PVE = 0.686730213353865
Green PVE = 0.6695858942373205
Blue  PVE = 0.673979774715561

the minimum number of principal components that are required to obtain at least 70% PVE for RED channel   = 12
the minimum number of principal components that are required to obtain at least 70% PVE for GREEN channel = 13
the minimum number of principal components that are required to obtain at least 70% PVE for BLUE channel  = 13
```

Figure 1 : PVE for each of the principal components and their sum for each channel

The PVE values obtained for the first 10 components are as follows:

Red PVE = 0.686730213353865
Green PVE = 0.6695858942373205
Blue PVE = 0.673979774715561

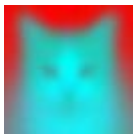
As you can see in Figure 1, as the number of principal components increases, the corresponding PVE value decreases. Because the first components explain more variance than the others. When we look at the cumulative for each channel, it increases gradually, but at a slower rate each time. If we had 4096 PCs, the cumulative PVE would be 1.

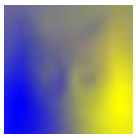
If we want the PVE value to be at least 70% for each channel, the number of PCs required for each channel is as follows.

For Red Channel: 12
For Green Channel: 13
For Blue Channel: 13

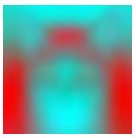
Question 1.2 - Eigencats

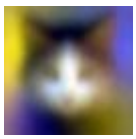

Eigencat-1

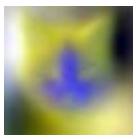

Eigencat-2

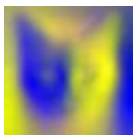

Eigencat-3


Eigencat-4

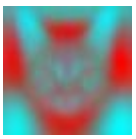

Eigencat-5


Eigencat-6


Eigencat-7


Eigencat-8


Eigencat-9


Eigencat-10

All eigencats have a cat-like appearance. However, They don't seem like a real cat image. With these 10 eigencats, we can explain (approximately) 67% of the variance of cats in our dataset. Yellow and dark blue color tones predominate in four of the images. However, there are also pictures in red and turquoise. This is because each PC reflects a different dimension. The advantage of these images for us is that they allow us to preserve our dataset as much as possible without increasing the number of dimensions too much. Initially, we had 5653 cat photos, and we reduced them to 64x64 pixels to create 4096 dimensions. While we can explain 100% of the variance with 4096 eigencats, we can explain about 67% of it with 10 eigencats. Therefore, there is a tradeoff between the PVE value and the dimension count.

Question 1.3 – Reconstruction

We can construct a cat image using the PCs obtained in Question 1.1 as follows:

1- I get the image as a numpy array.

2- I subtracted the mean values from the image array.

3- Dot product with principal components and eigencats

4- Since we subtracted the mean in Question 1.1, I added mean values.

5- Project the data back onto the original space using the first k eigencats.

The image (flickr_cat_000003.jpg) I reconstructed using the values of k ∈ {1, 50, 250, 500, 1000, 4096} is as follows:

k=1

k=50

k=250

k=500

k=1000

k=4096

As seen in the reconstruction images, when the k value is low, the image is blurry and does not look very much like the real picture. However, as the k value increases, the picture becomes clearer and closer to the original. If the k value is 4096, we get the original picture. This is due to the PVE value I explained in the previous sections. For example, when k = 1, the PVE values for each channel are as follows:

Red PVE = 0.2350696993627993

Green PVE = 0.20873714854025432

Blue PVE = 0.22859035906545536

This means that when the k value is 1, we can preserve the variance of approximately 22% of the image. When the k value is 4096, the PVE value will be 1 for all channels. This means that we can obtain the same image as the original by using reconstruction.