DSC 190 - Homework 05

Due: Wednesday, May 4

Write your solutions to the following problems by either typing them up or handwriting them on another piece of paper. Unless otherwise noted by the problem's instructions, show your work or provide some justification for your answer. Homeworks are due via Gradescope at 11:59 PM.

Problem 1.

Let $x^{(1)}, \ldots, x^{(n)}$ be n vectors in $\mathbb{R}^d$, and define the matrix X so that its ith row is the vector $x^{(i)}$. Let $\vec{v}$ be a vector in $\mathbb{R}^d$. Show that:

$$\frac{1}{n} \sum_{i=1}^{n} \left(x^{(i)} \cdot \vec{v} \right)^2 = \frac{1}{n} \vec{v}^T X^T X \vec{v}$$

Hint: this is an exercise in vector and matrix algebra. There are many ways of showing this, some involving much more writing than others. One of the easier ways is to define $\vec{u} = X\vec{v}$, and to recognize that the ith entry of $\vec{u}$ is $x^{(i)} \cdot \vec{v}$. It is also helpful to remember that, for any vector $\vec{a}$, $\vec{a}^T \vec{a} = \|\vec{a}\|^2$, and that for any matrices/vectors A and B, $(AB)^T = B^T A^T$.

The significance of this result is that the left hand side is exactly the variance in the direction of $\vec{v}$ (if we assume that $\vec{v}$ is a unit vector), and the right hand side is equal to $\vec{v}^T C \vec{v}$, where C is the covariance matrix. Therefore, this result shows that the variance in a direction can be computed using the covariance matrix.

Problem 2.

The Olivetti faces data set contains a collection of images of peoples' faces. It can be downloaded at the following URL:

`https://f000.backblazeb2.com/file/dsc-data/faces.csv`

Each row in that file represents one face. It is a vector with 4096 entries, each entry recording the intensity of a pixel in a 64×64 image. The row can be reshaped and plotted using `matplotlib` to display it as an image. For example, the code below plots the first image in the data set.

```python
import numpy as np
import matplotlib.pyplot as plt

faces = np.loadtxt("faces.csv", delimiter=',')
example_image = faces[0]
plt.imshow(example_image.reshape((64, 64)))
```

You should see this image:
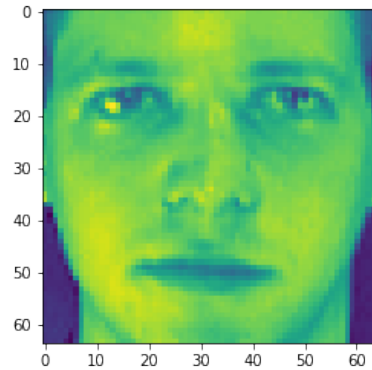
Some of the people in this dataset are wearing glasses – but who? We can use PCA to find eyeglass wearers without using any labels whatsoever. PCA can "discover" this automatically as one of the directions of greatest variance.

For this problem, you can use whatever programming language or package you'd like. But, as always, the tradeoff is that if you choose to use a package, you're expected to read the documentation to figure out how the package works.

a) Compute the **seventh** eigenvector $u^{(7)}$ of the data set's sample covariance matrix (that is, the eigenvector with seventh largest eigenvalue). It, too, should be a vector in $\mathbb{R}^{4096}$. Reshape this vector using code like the above, and visualize it. You should see something like a face with eyeglasses.

In your submission, show your code and the resulting image.

b) Take the dot product of eigenvector $u^{(7)}$ with every image in the data set. Plot the 20 images whose dot product with the eigenvector is the largest in absolute value. We will grade your answer to this problem manually by inspecting your plots.

If you did everything right, you should see a lot of eyeglasses.

Again, for this problem you should show the resulting images and your code.