Lab Assignment 2
Steven Truong

Task 1

Since the covariance matrix is symmetric, the eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_d$ are orthogonal. We can also assume that they are of unit norm. Thus, they are an orthonormal basis of $\mathbb{R}^d$. So, for any $\mathbf{x}_n \in \mathcal{D}$, there exist coefficients $c_1, \ldots, c_d$ such that

$$\mathbf{x}_n - \overline{\mathbf{x}} = \sum_{i=1}^{d} c_i \mathbf{u}_i.$$

Taking the inner product of this vector with $\mathbf{u}_j$, we get

$$\langle \mathbf{x}_n - \overline{\mathbf{x}}, \mathbf{u}_j \rangle = \sum_{i=1}^{d} c_i \langle \mathbf{u}_i, \mathbf{u}_j \rangle = c_j \| \mathbf{u}_j \|^2 = c_j.$$

So, we get

$$\mathbf{x}_n = \overline{\mathbf{x}} + \sum_{i=1}^{d} \langle \mathbf{x}_n - \overline{\mathbf{x}}, \mathbf{u}_i \rangle \mathbf{u}_i.$$

Task 2

We can extend $\mathbf{v}_1, \ldots, \mathbf{v}_k$ into an orthonormal basis of $\mathbb{R}^d$.

Using the same argument as in task 1, we can write any vector $\mathbf{x} \in \mathbb{R}^d$ as

$$\mathbf{x} = \sum_{i=1}^{d} \langle \mathbf{x}, \mathbf{v}_i \rangle \mathbf{v}_i = \sum_{i=1}^{k} \langle \mathbf{x}, \mathbf{v}_i \rangle \mathbf{v}_i + \sum_{i=k+1}^{d} \langle \mathbf{x}, \mathbf{v}_i \rangle \mathbf{v}_i \coloneqq \mathbf{x}_{\|} + \mathbf{x}_{\perp},$$

where $\mathbf{x}_{\|}$ is the part of $\mathbf{x}$ parallel to V, and $\mathbf{x}_{\perp}$ is the part of $\mathbf{x}$ orthogonal to V. This is easy to verify by taking inner products with $\mathbf{v}_1, \ldots, \mathbf{v}_k$.

The orthogonal projection removes the orthogonal part, so we get

$$p_V(\mathbf{x}) = \mathbf{x}_{\|} = \sum_{i=1}^{k} \langle \mathbf{x}, \mathbf{v}_i \rangle \mathbf{v}_i.$$

Task 3

Let

$$J_m = \frac{1}{N} \sum_{n=1}^{N} \| \hat{\mathbf{x}}_n - \mathbf{x}_n \|^2,$$

where $\mathbf{u}_n$ is a unit vector of the data covariance matrix, and

$$\hat{\mathbf{x}}_n = \overline{\mathbf{x}} + \sum_{i=1}^{m} \langle \mathbf{x}_n - \overline{\mathbf{x}}, \mathbf{u}_i \rangle \mathbf{u}_i = \overline{\mathbf{x}} + \sum_{i=1}^{m} (\mathbf{x}_n - \overline{\mathbf{x}})^\top \mathbf{u}_i \mathbf{u}_i.$$

If we write $\mathbf{x} = \sum c_i \mathbf{u}_i$, then by the Pythagorean theorem, $\|\mathbf{x}\|^2 = \sum c_i^2$. Then

$$\|\hat{\mathbf{x}}_n - \mathbf{x}_n\|^2 = \|\sum_{i=1}^{m} \langle \mathbf{x}_n - \overline{\mathbf{x}}, \mathbf{u}_i \rangle \mathbf{u}_i - \sum_{i=1}^{d} \langle \mathbf{x}_n - \overline{\mathbf{x}}, \mathbf{u}_i \rangle \mathbf{u}_i\|^2$$

$$= \|\sum_{i=m+1}^{d} \langle \mathbf{x}_n - \overline{\mathbf{x}}, \mathbf{u}_i \rangle \mathbf{u}_i\|^2$$

$$= \sum_{i=m+1}^{d} \langle \mathbf{x}_n - \overline{\mathbf{x}}, \mathbf{u}_i \rangle^2$$

$$= \sum_{i=m+1}^{d} \mathbf{u}_i^\top (\mathbf{x}_n - \overline{\mathbf{x}})(\mathbf{x}_n - \overline{\mathbf{x}})^\top \mathbf{u}_i.$$

Then if S is the data covariance matrix, we get

$$J_m = \frac{1}{N} \sum_{n=1}^{N} \|\hat{\mathbf{x}}_n - \mathbf{x}_n\|^2$$

$$= \frac{1}{N} \sum_{n=1}^{N} \sum_{i=m+1}^{d} \mathbf{u}_i^\top (\mathbf{x}_n - \overline{\mathbf{x}})(\mathbf{x}_n - \overline{\mathbf{x}})^\top \mathbf{u}_i$$

$$= \sum_{i=m+1}^{d} \mathbf{u}_i^\top \left(\frac{1}{N} \sum_{n=1}^{N} (\mathbf{x}_n - \overline{\mathbf{x}})(\mathbf{x}_n - \overline{\mathbf{x}})^\top \right) \mathbf{u}_i$$

$$= \sum_{i=m+1}^{d} \mathbf{u}_i^\top S \mathbf{u}_i$$

$$= \sum_{i=m+1}^{d} \mathbf{u}_i^\top \lambda_i \mathbf{u}_i$$

$$= \sum_{i=m+1}^{d} \lambda_i \|\mathbf{u}_i\|^2$$

$$= \sum_{i=m+1}^{d} \lambda_i.$$

Task 4

There is a large drop off in the size of the eigenvalues after the first two. This tells us that the data set probably came from a subspace of $\mathbb{R}^{100}$ with dimension 2, and that the smaller eigenvalues were also probably created by noise.

Task 5
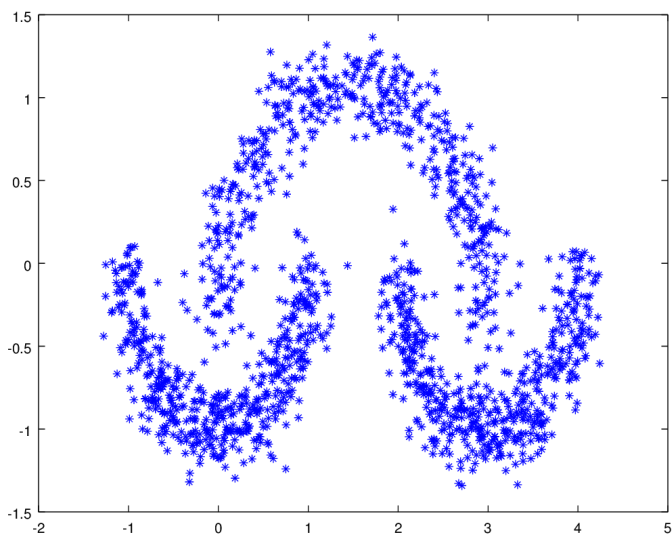
Yes, it looks like three crescent moons.

Task 6

Task 7

The inherent dimensionality is 1, since we only need the angle of rotation two distinguish any two images from the data set.

Task 8

The picture vectors lie in an at most N-dimensional subspace of $\mathbb{R}^{4096}$. Namely, they live in their own span, which can have at most N linearly independent vectors. So, we only need to calculate the N largest eigenvalues to figure out the principal components.

Task 9

They are circles with faint loop-looking shapes inside some of them. This makes sense since our data set was generated via rotations of the number 8, so circular principal components with loops are not surprising.

Task 10

It looks roughly like the original picture, but it's slightly rotated, the white pixels aren't as intense, and there are some extra faded loop. I don't think it's reasonable to project onto this subspace because the projected original image looks like one of the rotated images, which means we're unable to preserve the feature that differentiates each image from another.

Task 11

About 10 principal components are required before the projection looks very close to the original. This is not very surprising. Even though the inherent dimensionality of the data is 1, it's obvious that projecting onto 1 principal component will not give us the original image.