CS663 Assignment-4

Saksham Rathi, Kavya Gupta, Shravan Srinivasa Raghavan

Department of Computer Science,
Indian Institute of Technology Bombay

Question 1

Part (a)

We need to prove that the covariance matrix in PCA is symmetric and positive semi-definite.
   Symmetric: The covariance matrix is given by:

$$C = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T \tag{1}$$

where $\bar{x}$ is the mean of the data points. Consider the transpose of C, we know that $(AB)^T = B^T A^T$. Hence,

$$C^T = \left(\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T \right)^T = \frac{1}{n-1} \sum_{i=1}^{n} ((x_i - \bar{x})^T)^T (x_i - \bar{x})^T = \tag{2}$$

$$C^T = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T = C \tag{3}$$

Since, $C^T = C$, the covariance matrix is symmetric.
   Positive Semi-Definite: A matrix C is positive semi-definite if $w^t C w \geq 0$ for any vector w (definition of positive semi-definite). Let v be a vector of size $d \times 1$. Then, $v^T C v$ is given by:

$$v^T C v = v^T \left(\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T \right) v = \frac{1}{n-1} \sum_{i=1}^{n} v^T (x_i - \bar{x})(x_i - \bar{x})^T v \tag{4}$$

Let $y = (x_i - \bar{x})^T v$. Then, $v^T (x_i - \bar{x})(x_i - \bar{x})^T v = y^T y = ||y||^2 \geq 0$. Hence, $v^T C v \geq 0$ for any vector v. Therefore, the covariance matrix is positive semi-definite.

Part (b)

We need to prove that the eigen vectors of a symmetric matrix are orthonormal. (This is essentially the spectral theorem.)
   Let $\bar{v}$ and $\bar{w}$ be any two vectors. Since, A is symmetric, we have:

$$\bar{v}^T A \bar{w} = \bar{v}^T A^T \bar{w} = (A\bar{v})^T \bar{w} \tag{5}$$

In other words, $\bar{v}.(A\bar{w}) = (A\bar{v}).\bar{w}$.

Let λ_1 and λ_2 be two distinct eigen values of A and v_1 and v_2 be the corresponding eigen vectors. Then, we have:

$$Av_1 = \lambda_1 v_1 \quad \text{and} \quad Av_2 = \lambda_2 v_2 \tag{6}$$

From the previous result:

$$v_1.(Av_2) = (Av_1).v_2 \implies v_1.(\lambda_2 v_2) = (\lambda_1 v_1).v_2 \implies \lambda_2(v_1.v_2) = \lambda_1(v_1.v_2) \tag{7}$$

Since, $\lambda_1 \neq \lambda_2$, we have $v_1.v_2 = 0$. Hence, the eigen vectors of a symmetric matrix are orthonormal.

It is possible that an eigenvalue may have a larger multiplicity. However for a fixed eigen value λ, the set of vectors v for which $Av = \lambda v$ is a subspace, of full dimension, and we may simply choose an orthonormal basis for this subspace.

Therefore, eigen vectors of a symmetric matrix are orthonormal.

Part (c)

We are given a dataset of some N vectors in d dimensions given by $\{x_i\}_{i=1}^N$ with mean vector $\bar{x}$. The entire matrix of eigen vectors (denoted by V) is orthonormal (as shown in the previous part). We already know from lecture slides that:

$$x_i = \bar{x} + V\alpha_i \tag{8}$$

where α_i is the projection of x_i on the eigen vectors. We can write this as:

$$\alpha_i = V^T(x_i - \bar{x}) \tag{9}$$

We approximate x_i by the largest k eigen vectors. This is given by:

$$\tilde{x} = \bar{x} + V_k \alpha_{ik} \tag{10}$$

where $V_k = V(:, 1:k)$ (the largest k eigen vectors) and $\alpha_{ik} = (V_k)^T(x_i - \bar{x})$.
We need to find $||x_i - \tilde{x}_i||$. We have:

$$x_i - \tilde{x}_i = \bar{x} + V\alpha_i - \bar{x} - V_k\alpha_{ik} = V\alpha_i - V_k\alpha_{ik} \tag{11}$$

We can expand V and V_k and write this as:

$$x_i - \tilde{x}_i = \sum_{j=1}^d v_j\alpha_i(j) - \sum_{j=1}^k v_j\alpha_{ik}(j) = \sum_{j=1}^k v_j(\alpha_i(j) - \alpha_{ik}(j)) + \sum_{j=k+1}^d v_j\alpha_i(j) \tag{12}$$

It is clear that the first term is 0. Because, $\alpha_i(j) = V^T(x_i - \bar{x})$ and $\alpha_{ik}(j) = (V_k)^T(x_i - \bar{x})$. Since V_k is a subset of V, the first k eigen vectors, the first term is 0. Hence, we have:

$$x_i - \tilde{x}_i = \sum_{j=k+1}^d v_j\alpha_i(j) \tag{13}$$

Therefore, $||x_i - \tilde{x}_i||_2^2 = \sum_{j=k+1}^d (\alpha_i(j))^2$, this is because the eigen vectors are orthonormal. Hence, the error is given by the sum of the squares of the projections of the data points on the eigen vectors from $k+1$ to d.

The sum of this error over all data points is given by:

$$\sum_{i=1}^{N} ||x_i - \tilde{x}_i||_2^2 = \sum_{i=1}^{N} \sum_{j=k+1}^{d} (\alpha_i(j))^2 = \sum_{i=1}^{N} \sum_{j=k+1}^{d} (v_j^T(x_i - \bar{x}))(v_j^T(x_i - \bar{x})) \tag{14}$$

where v_j is the jth eigen vector.

$$\sum_{i=1}^{N} ||x_i - \tilde{x}_i||_2^2 = \sum_{i=1}^{N} \sum_{j=k+1}^{d} (v_j^T(x_i - \bar{x}))((x_i - \bar{x})^T v_j) = \sum_{j=k+1}^{d} (N-1)v_j^T C v_j = \sum_{j=k+1}^{d} (N-1)\lambda_j \tag{15}$$

where λ_j is the jth eigen value.
Therefore:

$$\frac{1}{N} \sum_{i=1}^{N} ||x_i - \tilde{x}_i||_2^2 = \frac{N-1}{N} \sum_{j=k+1}^{d} \lambda_j \tag{16}$$

It is already given to us that the last $d - k$ eigen values are very small. Thus, the error is also small. (We have also expressed the error in terms of the eigen values of the covariance matrix.)

Part (d)

We have two uncorrelated zero-mean random variables $X_1 \sim \mathcal{N}(0, 100)$ and $X_1 \sim \mathcal{N}(0, 1)$. We need to find the covariance matrix of the random vector $X = [X_1, X_2]^T$.

The covariance matrix is given by:

$$C = \begin{bmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) \\ \text{Cov}(X_1, X_2) & \text{Var}(X_2) \end{bmatrix} \tag{17}$$

Since, X_1 and X_2 are uncorrelated, $\text{Cov}(X_1, X_2) = 0$. Hence, the covariance matrix is given by:

$$C = \begin{bmatrix} 100 & 0 \\ 0 & 1 \end{bmatrix} \tag{18}$$

The eigen values of the matrix can be calculated by equating the determinant of the following matrix to 0:

$$\begin{vmatrix} 100 - \lambda & 0 \\ 0 & 1 - \lambda \end{vmatrix} = 0 \tag{19}$$

Therefore, the eigen values are $\lambda_1 = 100$ and $\lambda_2 = 1$. The eigen vectors are given by:
The corresponding eigen vectors are given by $[1, 0]^T$ and $[0, 1]^T$. Clearly one of the eigen values is much larger than the other. Therefore, the principal component of the data is given by projection of the data on the first eigen vector which is X_1.
If the variance of X_1 is reduced to 1, the covariance matrix becomes:

$$C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{20}$$

The eigen values of the matrix can be calculated by equating the determinant of the following matrix to 0:

$$\begin{vmatrix} 1 - \lambda & 0 \\ 0 & 1 - \lambda \end{vmatrix} = 0 \tag{21}$$

Therefore, the eigen values are $\lambda_1 = 1$ and $\lambda_2 = 1$. The corresponding eigen vectors are $[1, 0]^T$ and $[0, 1]^T$. Since, in this case the eigen values are the same, we can't reduce the dimensionality of the data. The principal components are the same as the original data.

4