COMP 7180 Quantitative Methods for Data Analytics and Artificial Intelligence
Exercise 4 Answer

1. Consider 3 data points in a 2-dimensional space: (-1, 2), (0, 0), (-1, -2).

   a) What is the first principal component of the given dataset?

   b) If we project the original data points onto the 1-dimensional subspace spanned by the first principal component, what are their coordinates in this subspace?

   c) Consider a point: (2,3). What is the coordinate in this subspace?

   Answer:

   a) Construct data matrix $\mathbf{X} = \begin{bmatrix} -1 & 0 & -1 \\ 2 & 0 & -2 \end{bmatrix}$

   Centralize data matrix $\mathbf{X} = \begin{bmatrix} -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ 2 & 0 & -2 \end{bmatrix}$

   Calculate the covariance matrix $\mathbf{S} = \frac{1}{n}\mathbf{X}\mathbf{X}^T = \begin{bmatrix} \frac{2}{9} & 0 \\ 0 & \frac{8}{3} \end{bmatrix}$.

   Conduct an eigen decomposition of the covariance matrix $\mathbf{S}$, we have eigenvalues $\frac{2}{9}$ and $\frac{8}{3}$, with corresponding eigenvectors $[1\ 0], [0\ 1]$.

   b) We take the largest eigenvalue $\frac{8}{3}$, and corresponding eigenvector $u = [0\ 1]$.

   Their coordinates are $\mathbf{Z} = u^T\mathbf{X}$, which are 2, 0, -2, respectively.

   c) The coordinates: $\mathbf{Z} = u^T\mathbf{x} = 3$.

2. Suppose we perform PCA on a two-dimensional dataset and the resulting two eigenvalues are equal. What does it mean?

   Answer:

   It means that this two dimensions play the same important role in preserving the variance. Performing dimensionality reduction would lose the information of the dimension that we reduced.

   Note: It can not be considered as that the two dimensions of this dataset are the same.

3. Consider we obtain 5 eigenvalues from the covariance matrix $S = \frac{1}{n}\mathbf{X}_c\mathbf{X}_c^T$, which are $10, 8, 5, 0.5, 0.2, 0.01$. How many Principal Components should we use?

   Answer:

   We should use 3 Principal Components. The Principal Components we use are the eigenvectors with corresponding eigenvalue $10, 8, 5$.

4. If $\mathbf{A}$ has singular values $v_1, v_2, \cdots, v_n$,what are the singular values of $k\mathbf{A}$? ($k > 0$)

   Answer:

   $kv_1, kv_2, \cdots, kv_n$

5. Suppose $\mathbf{A} = \mathbf{U\Sigma V}^T$ is the SVD. Suppose $\mathbf{A}$ is square and invertible. Find the SVD of the inverse of $\mathbf{A}$, which is $\mathbf{A}^{-1}$.

   Answer:

   $\mathbf{A}^{-1} = (\mathbf{U\Sigma V}^T)^{-1} = (\mathbf{V}^T)^{-1}\mathbf{\Sigma}^{-1}\mathbf{U}^{-1}$

   As we know about SVD, the matrix of $\mathbf{U}$ and $\mathbf{V}$ are orthogonal, which means that their transposes are their inverses $\mathbf{U}^T\mathbf{U} = \mathbf{U}\mathbf{U}^T = I, \mathbf{V}^T\mathbf{V} = \mathbf{V}\mathbf{V}^T = I$. Therefore, $(\mathbf{V}^T)^{-1} = (\mathbf{V}^{-1})^{-1} = \mathbf{V}$ and $\mathbf{U}^{-1} = \mathbf{U}^T$

   As a result, we have $\mathbf{A}^{-1} = \mathbf{V\Sigma}^{-1}\mathbf{U}^T$.

6. Consider a dataset with k classes. The data is $[(x_1, y_1), (x_2, y_2), \cdots, (x_m, y_m)]$. $x_i, i = 1, \cdots, m$ are the n dimensional vectors. $y_i \in [1, 2, \cdots, k], i = 1, \cdots, m$ are the labels of data, which represents the class of data x_i. Develop an LDA method for dataset with multiple classes and reduce the dimension of dataset from n to d.

Answer:

In here, we are mapping data into a d-dimension space, which corresponding base vectors are $w_1, w_2, \cdots, w_d$. Matrix $\mathbf{W}$ is constructed by base vectors, which is a $n \times d$ matrix.

In here, we want the center of each class are far from that of other classes. And we could consider that the center of each class is as far as possible from the center of all data. Therefore, we consider the between-class scatter matrix $\mathbf{S}_b = \sum_{j=1}^{k} N_j(\mu_j - \mu)(\mu_j - \mu)^T$, where N_j is the number of data in j^{th} class, and μ_j is the center of j^{th} class, and μ is the center of all data.

And the within-class matrix would be the sum of within-clas matrix of each class: $\mathbf{S}_w = \sum_{j=1}^{k} \mathbf{S}_{wj} = \sum_{j=1}^{k} \sum_{i=1}^{N_j} (x_i - \mu_j)(x_i - \mu_j)^T$.

Therefore, our objective function could be

$$\max_{\mathbf{W}} \frac{\mathbf{W}^T \mathbf{S}_b \mathbf{W}}{\mathbf{W}^T \mathbf{S}_w \mathbf{W}} \tag{1}$$

However, $\mathbf{W}^T\mathbf{S}_b\mathbf{W}$ and $\mathbf{W}^T\mathbf{S}_w\mathbf{W}$ are matrix instead of value. We can not maximize the matrix. Therefore, we consider the following objective function.

$$\max_{\mathbf{W}} \frac{\prod_{diag} \mathbf{W}^T \mathbf{S}_b \mathbf{W}}{\prod_{diag} \mathbf{W}^T \mathbf{S}_w \mathbf{W}} = \max_{\mathbf{W}} \prod_{i=1}^{d} \frac{\mathbf{w}_i^T \mathbf{S}_b \mathbf{w}_i}{\mathbf{w}_i^T \mathbf{S}_w \mathbf{w}_i} \tag{2}$$

where $\prod_{diag} A$ is the product of the main diagonal elements of matrix $\mathbf{A}$.

The maximum of above problem is the eigenvector corresponding to the largest eigenvalue of $\mathbf{S}_w^{-1}\mathbf{S}_b$. And the production of the largest d values is the production of the largest d eigenvalues. Therefore, the resulting matrix $\mathbf{W}$ is constructed by the d eigenvectors that corresponding to the largest d eigenvalues.

It may be worth to note that the maximum of the dimension d that we reduce is $k-1$. The rank of $(\mu_j - \mu)(\mu_j - \mu)^T$ is 1. Therefore, the rank of $\mathbf{S}_b$ can not be larger than k (The rank of the matrix is less than or equal to the sum of the ranks of the individual sum matrices). However, if we know the first $k-1$ μ_j, the last μ_k can be represented as the linear combination of the first $k-1$ μ_j. Therefore, the rank of $\mathbf{S}_b$ can not be larger than $k-1$, which indicates that the eigenvectors can not be more than $k-1$.