

COMP 7180 Quantitative Methods for Data Analytics and Artificial Intelligence

Exercise 4

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?
2. Suppose we perform PCA on a two-dimensional dataset and the resulting two eigenvalues are equal. What does it mean?
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?
4. If \mathbf{A} has singular values v_1, v_2, \dots, v_n , what are the singular values of $k\mathbf{A}$? ($k > 0$)
5. Suppose $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ is the SVD. Suppose \mathbf{A} is square and invertible. Find the SVD of the inverse of \mathbf{A} , which is \mathbf{A}^{-1} .
6. Consider a dataset with k classes. The data is $[(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)]$. $x_i, i = 1, \dots, m$ are the n dimensional vectors. $y_i \in [1, 2, \dots, k], i = 1, \dots, m$ 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 .