

# Homework 3 Theory

September 2024

## 1 Problem 1

### Problem Statement

Let  $A$  be an  $m \times d$  matrix, and let  $X = AA^T$ . Assume that  $X$  has  $d$  distinct, non-zero eigenvalues. Assume that  $m \gg d$ . In order to find the eigendecomposition of  $X$ , we will need to find the eigendecomposition of an  $m \times m$  matrix. Since  $m$  is much larger than  $d$ , this is slow. Give an algorithm for finding the eigenvectors and eigenvalues of  $X$  that only requires computing the eigendecomposition of a  $d \times d$  matrix. You can use simple matrix operations and assume that you have an eigendecomposition “black box” subroutine, but avoid using the SVD as a black box.

### Solution

#### Intuition

Since the rank of  $X$  is at most  $d$  (because  $A$  only has  $d$  columns), we only need to compute the eigendecomposition of a smaller  $d \times d$  matrix. This is because while  $AA^T$  is size  $m \times m$ ,  $A^T A$  is size  $d \times d$ , and the two matrices share the same eigenvalues because they are both derived from  $A$ . Thus, we can use the calculated eigenvalues for  $B = A^T A$  to construct the eigenvectors of  $X$ .

#### Algorithm

##### Compute $B$

Given  $A$ , an  $m \times d$  matrix, a  $d \times d$  matrix  $B$  will be computed as:

$$B = A^T A$$

##### Compute Eigendecomposition of $B$

The eigendecomposition of  $B$  will be computed as:

$$B = QDQ^T$$

Where  $D$  is a diagonal matrix containing the eigenvalues of  $B$ , and  $Q$  is an orthogonal matrix whose columns are the eigenvectors of  $B$ .

## Derive the Eigenvectors and Eigenvalues of $X$ from the Eigendecomposition of $B$

### 1. Define an eigenvector of $B$

Let's fix  $\mathbf{v}_i$  which is an eigenvector of  $B$  in  $d$ -dimensional space corresponding to the eigenvalue  $\lambda_i$ , using the eigenvector and eigenvalue definition  $B\mathbf{v}_i = \lambda_i\mathbf{v}_i$ :

$$B\mathbf{v}_i = A^T A \mathbf{v}_i = \lambda_i \mathbf{v}_i$$

### 2. Define a new vector $\mathbf{u}_i$ to map $\mathbf{v}_i$ into higher dimensional space

Let's define a vector  $\mathbf{u}_i$  in the larger  $m$ -dimensional space as:

$$\mathbf{u}_i = A \mathbf{v}_i$$

Where multiplying  $\mathbf{u}_i$  by  $A$  maps the vector  $\mathbf{v}_i$  from  $d$ -dimensional space to the  $m$ -dimensional space.

### 3. Show that $\mathbf{u}_i$ is an eigenvector of $X$

To prove that  $\mathbf{u}_i$  is an eigenvector of  $X$ , we must show:

$$X\mathbf{u}_i = \lambda_i \mathbf{u}_i$$

Substituting  $\mathbf{u}_i = A\mathbf{v}_i$  and  $X = AA^T$  into the equation:

$$AA^T(A\mathbf{v}_i) = \lambda_i(A\mathbf{v}_i)$$

Since  $B = A^T A$ :

$$A(B\mathbf{v}_i) = \lambda_i(A\mathbf{v}_i)$$

And since  $B\mathbf{v}_i = \lambda_i\mathbf{v}_i$ :

$$A(\lambda_i\mathbf{v}_i) = \lambda_i(A\mathbf{v}_i)$$

Which is the same as:

$$\lambda_i(A\mathbf{v}_i) = \lambda_i(A\mathbf{v}_i)$$

Proving that:

$$X\mathbf{u}_i = \lambda_i \mathbf{u}_i$$

## 2 Problem 2

### Problem Statement

In this problem we explore some relationships between SVD, PCA and linear regression.

- a) True or false: linear regression is primarily a technique of supervised learning, i.e., where we are trying to fit a function to labeled data.
- b) True or false: PCA is primarily a technique of unsupervised learning, i.e., where we are trying to find structure in unlabeled data.
- c) True or false: SVD is primarily an operation on a dataset, whereas PCA is primarily an operation on a matrix.
- d) A common problem in linear regression is multicollinearity, where the input variables are themselves linearly dependent. For example, imagine a healthcare dataset where height is measured both in inches and centimeters. This is a problem because there may now be multiple  $w$  satisfying  $y = w \cdot x$ . Explain how you could use a preprocessing step to solve this problem.

## Solution

### Part A

**True.**

Linear regression is a supervised learning technique, since we have a training set  $\mathbf{S}$  which has data points and associated labels. The goal is to find the best fitting line to the labeled training set so we can use the model to predict on unseen data.

### Part B

**True.**

Since PCA is a technique to find patterns and information in a matrix of unlabeled data, it is a technique of unsupervised learning. The goal is to find the directions upon which the data has the most variance (information).

### Part C

**False.**

SVD and PCA are both matrix operations. SVD is a technique for decomposing a matrix into a simpler structure (and also performing the eigendecomposition at the same time).

### Part D

To address the multicollinearity issue, PCA can be performed as a preprocessing step. The PCA will find the directions for which the data has the most variance, or provides the most information about the dataset, and reduces the redundancy of the dataset. Furthermore, **the resulting principal components will be linearly independent of one another, eliminating the multicollinearity.**