

Homework 3 Theory

Question 1

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.

Let's first show a relationship between the eigenvalues of matrices X and Y :

We will start by stating the following:

$$Y = A^T A$$

We will start with the linear equation $Av = \lambda v$, where A is a matrix, v is a vector and λ is a scalar:

$$Yv = \lambda v$$

Given $Y = A^T A$, we have:

$$A^T Av = \lambda v$$

Multiply each side by matrix A :

$$AA^T Av = \lambda Av$$

We will call vector u as $u = Av$ and substitute appropriately:

$$AA^T u = \lambda u$$

Substitute X :

$$Xu = \lambda u$$

Therefore we know that given v is an eigenvector of $Y = A^T A$ with eigen value λ_i , then we know that $u = Av$ will be an eigenvector of $X = AA^T$ with the same eigen values.

Knowing this the algorithm with using the eigendecomposition “black box” subroutine we have:

- start with $Y = A^T A$
- find the eigenvalues and eigen vectors of Y using the eigendecomposition as seen below:

$$A = QDQ^{-1}$$

where D is a $d \times d$ diagonal matrix containing the eigenvalues (λ_i) of A , and Q is an orthogonal matrix containing the corresponding eigenvectors.

- compute the eigen vectors of X using $u = Av$ and multiplying the eigenvectors of Y by A

This algorithm takes the problem of finding the eigendecomposition of X which is an $m \times m$ matrix and reduces it to more efficiently finding the eigendecomposition of Y which is a $d \times d$ matrix, since $m \gg d$

Question 2

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.

True. Linear regression is a supervised learning technique as it tried to fit a line function that minimizes the error from the true distance of labeled data from the line. This could not be used if the data was unlabeled in unsupervised learning

(b) True or false: PCA is primarily a technique of unsupervised learning, i.e. where we are trying to find structure in unlabeled data.

False. PCA is not actually building a model that is trying to “fit” according to labeled data, but rather is learning the structure of the data in order to reduce its complexity/dimensionality

(c) True or false: SVD is primarily an operation on a dataset whereas PCA is primarily an operation on a matrix.

False. SVD is used to find the singular value of a matrix and is more generally used in Linear Algebra applications, whereas PCA is more specifically used often to reduce the dimensionality of a large dataset but retain much of the variance of the data in order to reduce the complexity

(d) A common problem in linear regression is multicollinearity, where the input variables are themselves linearly dependent. For example, imagine a healthcare data set where height is measured both in inches and centimetres. 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.

There could be a few things that one could do to solve the multicollinearity problem. The first thing to do would be to standardize the data and analyze the results potentially by using a heatmap technique. This may highlight data that is collinear. Once you know which items may not be independent of each other, then you could eliminate the redundant variables in the data. This is easy to see when looking at something such as different measurements of the same property such as different measurements of distance, or speed, or times. PCA is also a common technique for eliminating redundancy by identify principal components and reducing the dimensionality.