Homework 3 - Theory

(1) $A : m \times d$ $\qquad m >> d$

$X = AA^T$ $\qquad A$ match A^T

$A : m \times d$ $\qquad m \times d \quad d \times m$

$A^T : d \times m$

$A \cdot A^T$ $\qquad m \times m$

X

$d \times d \Rightarrow \lambda, v$

$\cancel{A \cdot A^T}$ $\quad (A^T \cdot A)$ $\quad d \times m \quad m \times d$

$= d \times d$

Given a matrix A, we can compute the matrix $X = AA^T$. However, as you noted, finding the eigendecomposition of an $m \times m$ matrix can be computationally expensive, especially when $m >> d$.

Instead, we will use the eigendecomposition of $d \times d$ matrix $A^T \cdot A$, which is cheaper to compute.

To accomplish this, we proceed with the following steps:

- Compute the matrix $B = A^T A$. The matrix B is of dimension $d \times d$.
- Compute the eigendecomposition of B:

  $B = VDV^{-1}$, where V is the matrix whose columns are the eigenvectors of B and D is the diagonal matrix whose entries are the eigenvalues of B.
- Compute the eigenvectors of X. The eigenvectors of X are given by $E = A \cdot V$, where V is the matrix of eigenvectors of B. In other words, we can obtain the eigenvectors of X by multiplying A with the eigenvectors of B.

- Compute the eigenvalues of X.

The eigenvalues of X are the same as the eigenvalues of B, which is given by the diagonal elements of D.

To summarize the key observation here is that the eigenvalues of $X = AA^T$ and $B = A^TA$ are the same, but their eigenvectors are related by a simple matrix multiplication: the eigenvectors of X are obtained by multiplying the matrix A by the eigenvectors of B.

This method allows you to find the eigendecomposition of X without ever having to compute the eigendecomposition of an mxm matrix, and instead only computing the eigendecomposition of dxd matrix.

(2)

(a) True. Linear regression is indeed a technique of supervised learning. It is used to establish a linear relationship between a dependent variable which we are tryiy to predict and 1 or more independent variables, which we're usiy to make prediction given a labeled dataset. The labels in this case are the actual values of the dependent variable for each instance in the dataset.

(b) True. Pricipal Component Analysis (PCA) is indeed an unsupervised learniy technique. It is used to find the directions (principal components) in which the data varies the most, and does not require any labels or class information. It's goal is to reduce the dimensionality of the data while preserviy as much variance as possible, which can reveal underlyiy structure in the dataset.

(c) False. Both Singular Value Decomposition (SVD) and Principal Component Analysis (PCA) are operations on matrices, not specifically on datasets. They are matrix factorization techniques used in a wide variety of applications including dimensionality reduction, data compression and noise reduction.

In the context of data analysis, the input matrix to both SVD and PCA is often a data matrix, where each row corresponds to a different observation (different costomer or a different experiment) and each column corresponds to a diffrent variable (a different product or different measurement).

For PCA, the matrix is typically a covariance or correlation matrix of the centered data, whereas for SVD, the matrix can be the data matrix itself. But these are both operations on matrices, regardless of the source of the data.

(d) Multicollinearity is a common issue in linear regression that can lead to unstable and sensitive estimates of the model parameters. One method to mitigate the effects of multicollinearity involves preprocessing the data using Principal Component Analysis (PCA).

PCA is a technique that transforms the original data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate called the first pricipal component, the second greatest variance on the second coordinate and so on.
This transformation is done in such a way that the new variables the principal components are uncorrelated with each other, eliminating multicollinearity.

Here is how we can use PCA as a preprocessing step to solve multicollinearity problem:

Center the data: Subtract the mean of each variable (each column in your data matrix) from their respective values. This centers the data around the origin.

Apply PCA: Apply PCA on the centered data to obtain the pricipal components. This yields a set of new variables, which are the principal components.

Transform the data: Transform the original data to the PCA space. Now, each data point is expressed as a linear combination of the pricipal components.

Perform Linear Regression: Perform linear regression using the new transformed variables. The resulting model parameters will be in the PCA space.

This approach effectively decorrelates the variables, preventing multicollinearity. However, interpreting the model can be more challenging because the regression coefficients are now associated with principal components, which are combinations of the original variables, instead of the original variables themselves.