# Notes on Dimension Reduction

Tianjian Li

February 2022

## 1 Principle Component Analysis

**Principle Component Analysis(PCA)** is a common dimension reduction technique. We want to find a hyperplane that have the following properties:

1. The distance between the samples to the plane are as near as possible
2. The projection of the samples onto the plane are as far as possible

Suppose we are given a dataset $\{x^{(i);i=1,...,n}\}$ of n different samples, and let $x^{(i)} \in \mathbb{R}^d$. We first normalize each feature so that their mean is 0 and variance is 1.

$$x_j^{(i)} := \frac{x_j^{(i)} - \mu_j}{\sigma_j}$$

Where $\mu_j$ and $\sigma_j$ is the mean and variance of feature $j$, respectively.

We want to maximize the variance of the projected samples. The variance is given by

$$\frac{1}{n}\sum_{i=1}^{n}(x^{(i)\top}u)^2 = \frac{1}{n}\sum_{i=1}^{n}(u^\top x^{(i)}x^{(i)\top}u) = u^\top(\frac{1}{n}\sum_{i=1}^{n}x^{(i)}x^{(i)\top})u$$

So our goal is:

$$\arg\max_u = u^\top(\frac{1}{n}\sum_{i=1}^{n}x^{(i)}x^{(i)\top})u$$

Under the constraint $||u||_2 = 1$.

The largest $u$ corresponds to the principle eigenvector(eigenvector which has the largest eigenvalue) of the matrix $\Sigma = \frac{1}{n}\sum_{i=1}^{n}x^{(i)}x^{(i)\top}$. If we want to reduce the input dimension to k, we need the k largest eigenvectors $u_1, u_2, ..., u_k$. The k largest eigenvectors are called the **principle components** of the data.

## 2 Singular Value Decomposition

In the previous section, we have shown that to perform PCA on a dataset is the same as performing the eigendecomposition of the feature covariance matrix $\frac{X^T X}{n}$. In this section we will show that performing PCA also means to perform **Singular Value Decomposition** on the feature matrix $X$. Note that SVD is performed on the feature matrix instead of the covariance matrix.

The eigendecomposition of the covariance matrix is

$$C = VLV^\top$$

Where $V$ is a matrix of eigenvectors and $L$ is a diagonal matrix of the eigenvalues sorted from largest to smallest. The first k Rows in $V$ corresponds to the principle components and $XV$ is the projected dataset.

The Singular Value Decomposition of $X$ is given by

$$X = USV^\top$$

Where $U$ is a unitary matrix and $S$ is the diagonal matrix with singular values. Plugging $X = USV^\top$ to $C = \frac{X^T X}{n}$, we obtain

$$C = \frac{VSU^\top USV^\top}{n} = \frac{VS^2V}{n}$$

Meaning that $\frac{S^2}{n}$ is the diagonal eigenvalue matrix $L$. The projected dataset $XV = USV^\top V = US$ Therefore we have shown that performing SVD on the feature matrix could get us the principle components of the dataset.

Numerically it is better to perform SVD than calculating the covariance matrix. The formation of the covariance matrix can lead to loss of precision.