# Dimensionality Reduction

Jiangzekun Wang

November 17, 2022

## 1 Introduction

Dimensionality reduction is the process of reducing the number of features in a dataset. For machine learning, tasks like regression or classification, there are often too many features to work with. And sometimes, most of these features are correlated. The curse of dimensionality states that the more features there are, the harder it is to model them. Hence it's necessary to remove redundant features from the training data in the beginning.

Dimensionality reduction is commonly used in data visualization to comprehend and interpret data, as well as in machine learning or deep learning techniques to simplify the task at hand.

## 2 Techniques for Dimensionality Reduction

### 2.1 Feature Selection

Feature selection involves identifying a subset of the input features. There are basically three strategies, which are listed as follows:

- *Filter*
- *Wrapper*
- *Embedded*

#### 2.1.1 Filter method

Variables are selected using filter methods regardless of the model. They are only based on general characteristics such as the correlation with the variable to predict. The least interesting variables are suppressed by filter methods. The remaining variables will be used in a classification or regression model to classify or predict data. These methods are particularly efficient in terms of computation time and are resistant to overfitting.

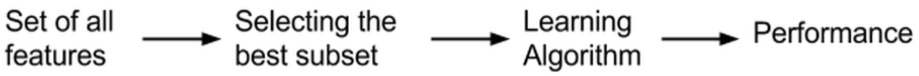

Figure 1: Filter Method for feature selection

#### 2.1.2 Wrapper method

Wrapper methods evaluate subsets of variables, allowing them to detect potential interactions between variables, unlike filter approaches. These methods have two major drawbacks:

- When the number of observations is insufficient, the risk of overfitting increases.
- When the number of variables is large, the computation time becomes significant.

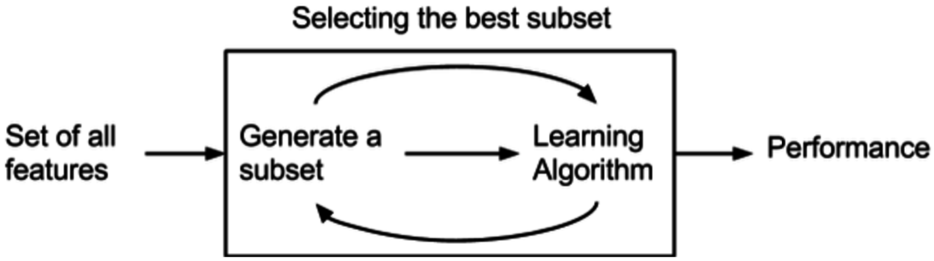

Figure 2: Wrapper Method for Feature selection

### 2.1.3 Embedded method

Embedded methods attempt to combine the benefits of both previous methods have been proposed. The FRMT algorithm, for example, uses its own variable selection process to perform feature selection and classification simultaneously.

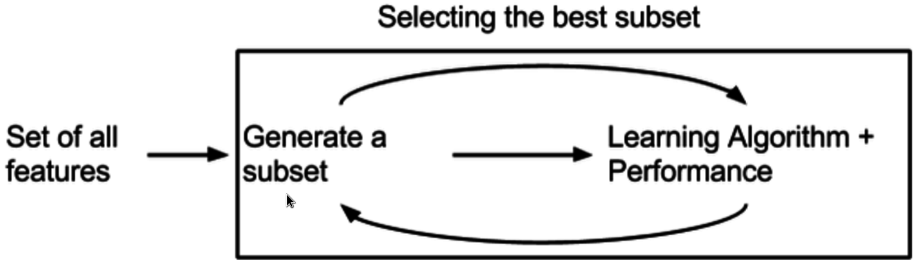

Figure 3: Embedded method for Feature selection

## 2.2 Feature projection

Feature projection converts data from a high-dimensional space to a lower-dimensional space. The data transformation may be linear, as in principal component analysis (PCA), but many nonlinear dimensionality reduction techniques exist. Tensor representation can be used in dimensionality reduction through multi-linear subspace learning for multidimensional data.

# 3 Linear Dimensionality Reduction

Linear dimensionality reduction methods are a cornerstone of analyzing high dimensional data, due to their simple geometric interpretations and typically attractive computational properties. These methods capture many data features of interest, such as covariance, dynamical structure, correlation between data sets, input-output relationships, and margin between data classes.

## 3.1 Principal Component Analysis (PCA)

PCA is a linear transformation. transformation. It involves the process of finding the principal components, which is the decomposition of the feature matrix into eigenvectors. The implementation of PCA is really straightforward. The entire procedure may be broken down into only four steps:

- **Standardization**: The data has to be transformed to a common scale by taking the difference between the original dataset with the mean of the whole dataset. This will make the distribution 0 centered

- **Finding covariance**: Covariance will help us to understand the relationship between the mean and original data.

- **Determining the principal components**: The eigenvectors and eigenvalues can be used to calculate the principal components. A unique collection of vectors called eigenvectors aids in our comprehension of the principle component-like structure and characteristics of the data. On the other hand, the eigenvalues assist us in identifying the primary components. The most significant main components are those with the highest eigenvalues and matching eigenvectors.
- **Final output**: It is the dot product of the standardized matrix and the eigenvector. Note that the number of columns or features will be changed.

## 3.2 PCA Case Study

The case study employs Iris dataset to demonstrate the PCA technique. The original data has 4 features (sepal length, sepal width, petal length, and petal width). Then the original data is reduced into 2 dimensions. The new components are just the two main dimensions of variation.

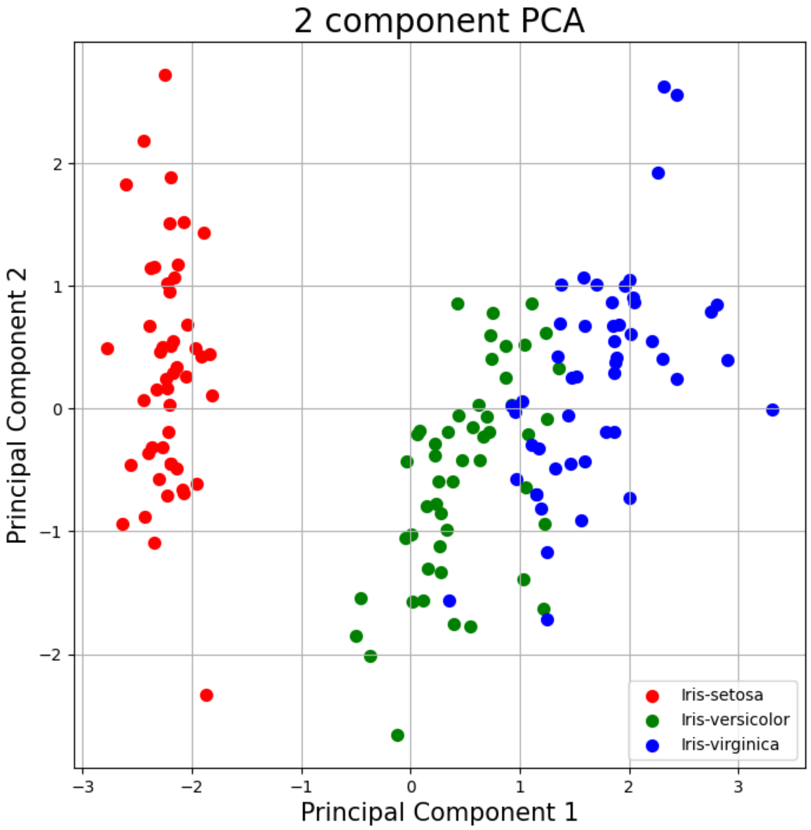

Figure 4: 2 Component PCA

When we convert 4 dimensional space to 2 dimensional space, some of the variance (information) is lost. In this simulation, first principal component contains 72.77% of the variance and the second principal component contains 23.03% of the variance. Together, the two components contain 95.80% of the information. Hence, this is a trade-off between the accuracy and complication.