Department of AI&DS

MACHINE LEARNING
22AD2203R

Topic:

DIMENSIONALITY REDUCTION

Session - 07

Dr. NARENDRA BABU TATINI
Associate Professor
Department of IoT

AIM

To familiarize students with the concepts of dimensionality reduction, its types, Covariance matrix, eigenvalues and eigen vectors

INSTRUCTIONAL OBJECTIVES

This session is designed to:

1. Learn about Dimensionality Reduction
2. Know different types in Dimensionality Reduction
3. Calculate Covariance Matrix for an n-dimensional data
4. Calculate eigenvalues and eigenvectors

LEARNING OUTCOMES

At the end of this session, you should be able to:

1. Define Dimensionality Reduction
2. Understand the Dimensionality Reduction
3. Calculate Covariance Matrix for an n-dimensional data
4. Calculate eigenvalues and eigenvectors

INTRODUCTION

- Often in machine learning problems, input vectors have high dimensionality D.
  - for an 1800 x 1200 colour image, D≅6.5 million
  - for a 1-second acoustic voice signal sampled at 5kHz, D = 5,000
  - Handling high dimensional data and applying ML models on such a huge dataset is a complex and time-consuming task.

Most of such datasets may contain noise and un-useful dimensions which need to be removed.

CURSE OF DIMENSIONALITY

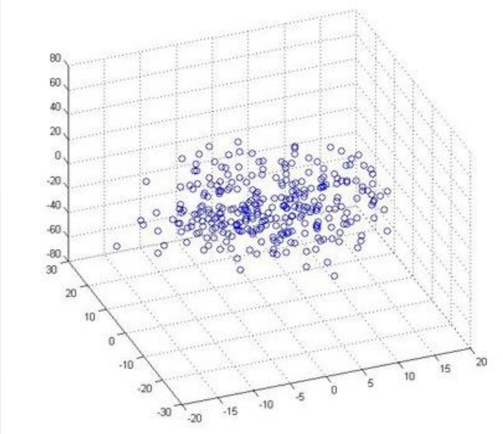
We have already seen the curse of dimensionality in CO-1.

- The higher the number of dimensions we have, the more training data we need.
- The dimensionality is an explicit factor for the computational cost of many algorithms.
- High dimensional data is harder for interpretation and visualization.
- It may be prone to noise and significantly effect the results of the learning algorithm

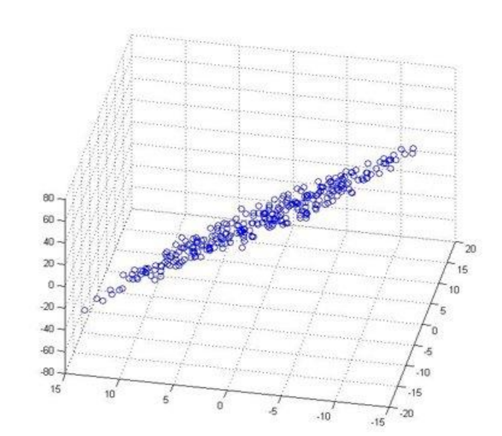
Dimensionality reduction is helpful in downsizing the data.

DEFINITION

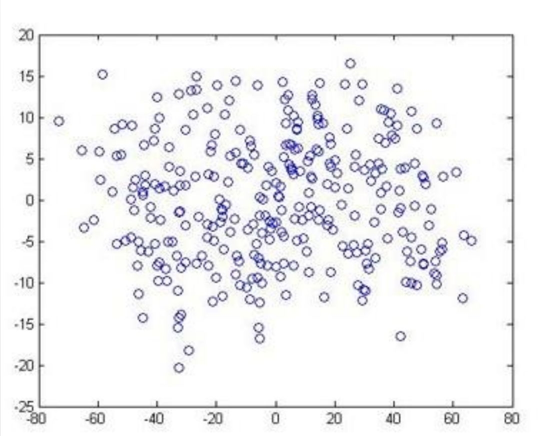
Dimensionality Reduction is the transformation of data from a high dimensional space (D-dimensions) into a low dimensional space (K- dimensions where K<D) so that the low-dimensional representation retains some meaningful properties of the original data.



Data in a 3-D space



Transformation/Rotation



Projecting into a 2-D Space

TYPES OF DIMENSIONALITY REDUCTION

- There are three ways to do dimensionality reduction.
  - Feature Selection
  - Feature Extraction/Feature Derivation
  - Clustering

FEATURE SELECTION

Selecting through the features/dimensions that are available and seeing whether they are useful, i.e., correlated to the output variables.

- Choosing K<D important features and ignoring D-K features.
- Types of Feature Selection Methods
  - Filtering Methods : Filter methods pick up the intrinsic properties of the features measured via **univariate statistics**
    - Information gain
    - Variance Threshold
    - Chi-Square
    - Etc.

FEATURE SELECTION (CONT..)

- **Wrapper Methods:** It follows a greedy search approach by evaluating **all the possible combinations** of features against the evaluation criterion. The wrapper methods usually result in better predictive accuracy than filter methods.
  - **Forward Feature Selection**
  - **Backward Feature Elimination**
  - **Exhaustive Feature Selection**
  - Etc.

FEATURE SELECTION

- **Embedded Methods:** Embedded methods combine the qualities' of filter and wrapper methods.
- Lasso regression
- Ridge regression

These algorithms have inbuilt penalization functions to reduce overfitting

FEATURE EXTRACTION

- Deriving new features from the old ones, generally by applying transforms to the dataset that simply change the axes **(coordinate system)** of the graph by moving and rotating them.
- i.e., Project the original X_i , i=1,2,..D dimensions to new d_j, j=1,2,..K dimensions where K<D.
- Example:
  - Principal Component Analysis (PCA)
  - Linear Discriminants Analysis (LDA)

CLUSTERING

- **Clustering:** in order to group together similar datapoints, and to see whether this allows fewer features to be used.
  - K-Means Clustering
  - Hierarchical Clustering

DISADVANTAGES OF DIMENSIONALITY REDUCTION

- It may lead to some amount of data loss.
- PCA tends to find linear correlations between variables, which is sometimes undesirable.
- PCA fails in cases where mean and covariance are not enough to define datasets.
- We may not know how many principal components to keep- in practice, some thumb rules are applied.

VARIANCE AND COVARIANCE

Measure of the "spread" of a set of points around their center of mass(mean)

- Variance:
  - Measure of the deviation from the mean for points in one' dimension

$$\mathrm{var}(x) = \frac{\sum_1^n (x_i - \mu)^2}{n}$$

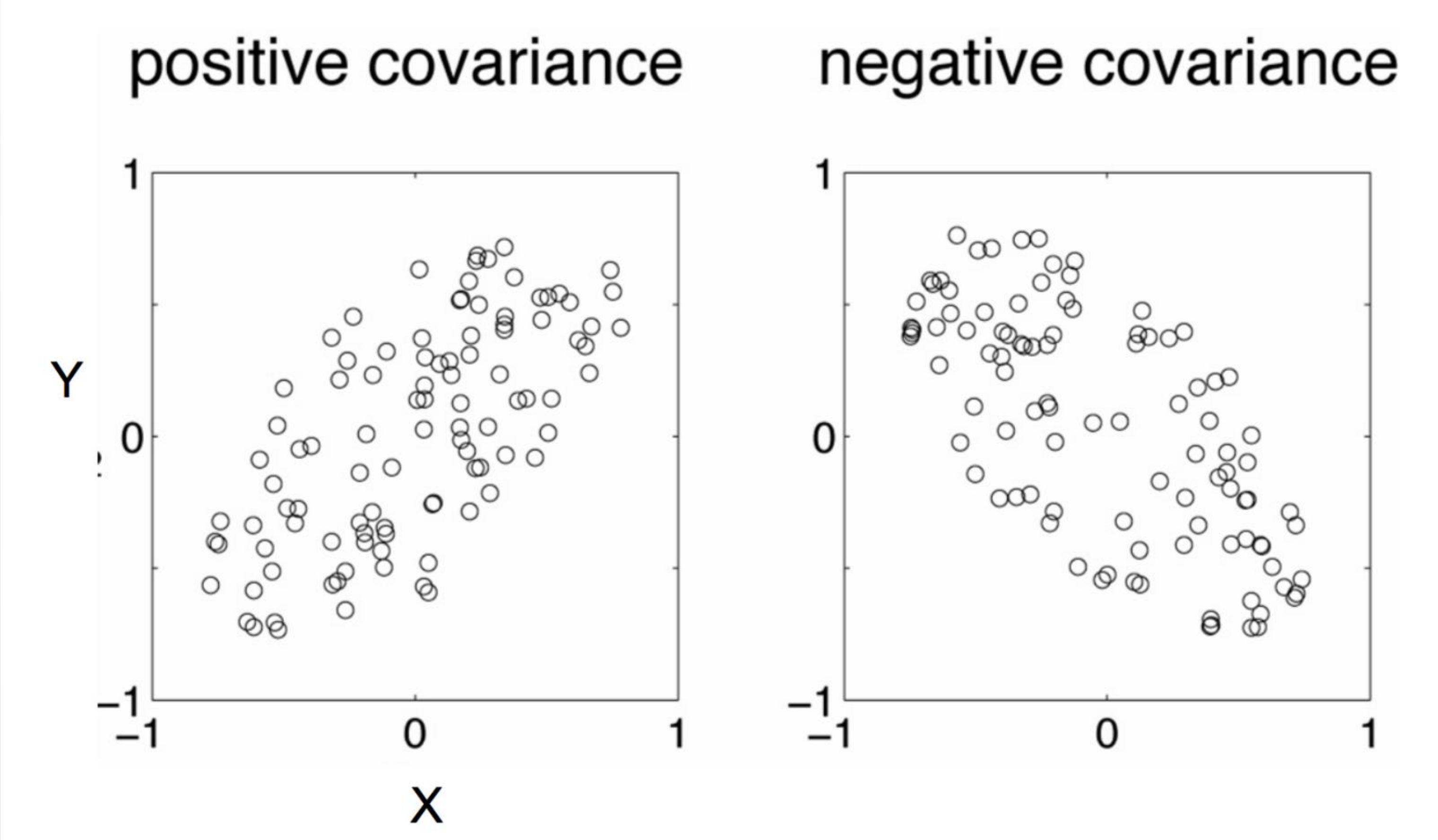
- Covariance:
  - Measure of how much each of the dimensions vary from the mean with **respect to each other**

Where μ = mean of population data.
n = number of observations in the dataset.

$$\mathrm{cov}(x, y) = \frac{\sum_1^n (x_i - \mu_x)(y_i - \mu_y)}{n}$$

- **Covariance is measured between two dimensions**
- **Covariance sees if there is a relation between two dimensions**
- **Covariance between one dimension is the variance**

COVARIANCE



Positive: Both dimensions increase or decrease together Negative: While one increase the other decrease

COVARIANCE MATRIX

- Used to find relationships between dimensions in high dimensional data sets

$$cor(x_j, x_k) = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \overline{x_j})(x_{ik} - \overline{x_k})$$

The Sample mean

COVARIANCE MATRIX

$$\begin{array}{c|cc} & X & Y \\ \hline X & \multicolumn{2}{c}{\multirow{2}{*}{}} \end{array}$$

$$\begin{array}{ccc} & X & Y \\ \begin{matrix} X \\ Y \end{matrix} & \multicolumn{2}{c}{\begin{bmatrix} var(x) & cov(x, y) \\ cov(x, y) & var(y) \end{bmatrix}} \end{array}$$

Two Dimensions

$$\begin{array}{cccc} & X & Y & Z \\ \begin{matrix} X \\ Y \\ Z \end{matrix} & \multicolumn{3}{c}{\begin{bmatrix} var(x) & cov(x, y) & cov(x, z) \\ cov(x, y) & var(y) & cov(y, z) \\ cov(x, z) & cov(y, z) & var(z) \end{bmatrix}} \end{array}$$

Three Dimensions

$$\begin{bmatrix} Var(x_1) & \dots & Cov(x_1, x_n) \\ \vdots & . & \vdots \\ \vdots & . & \vdots \\ Cov(x_n, x_1) & \dots & Var(x_n) \end{bmatrix}$$

n Dimensions

EIGEN VECTOR AND EIGEN VALUE

Eigenvector of a square matrix is defined as a non-zero vector in which when a given matrix is multiplied, it is equal to a scalar multiple of that vector. We have to find eigenvalues always before finding the eigenvectors.

$$Av = \lambda v$$

A: Square Matrix
v: Eigenvector or characteristic vector of A
λ : Eigenvalue or characteristic value of A

$$|A - \lambda I| = 0$$

- ***The zero vector can not be an eigenvector***
- ***The value zero can be eigenvalue***

COMPUTING EIGEN VECTOR AND EIGEN VALUES

- Find the eigenvalues and eigenvectors of matrix A $=\begin{bmatrix} 5 & 4 \\ 1 & 2 \end{bmatrix}$
- Let λ represent the eigenvalue(s) and **v** $= \begin{bmatrix} x \\ y \end{bmatrix}$ represent the eigenvector
- $|A - \lambda I| = 0$
- $\begin{vmatrix} 5-\lambda & 4 \\ 1 & 2-\lambda \end{vmatrix} = 0$
- $(5 - \lambda)(2 - \lambda) - (4)(1) = 0$
- $10 - 5\lambda - 2\lambda + \lambda^2 - 4 = 0$
- $\lambda^2 - 7\lambda + 6 = 0$
- $(\lambda - 6)(\lambda - 1) = 0$
- $\lambda = 6, \lambda = 1.$

- Thus, the eigenvalues are 1 and 6. Let us find the corresponding eigenvector to each eigenvalue in each case.
- **When λ = 1 and 6 in the equation** (A - λI) **v** = O
- the eigenvectors of the given 2 × 2 matrix are
- $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 4 \\ 1 \end{bmatrix}$

SUMMARY

- Dimensionality Reduction is a process of mapping a high dimensional space into a low dimensional space
- There are three different types of Dimensionality reduction techniques
  - Feature selection
  - Feature Extraction
  - Clustering
- Dimensionality reduction can be useful for data compression and simplification of the learning process

SELF-ASSESSMENT QUESTIONS

1. Given a covariance matrix $\begin{array}{c|cc} & X & Y \\ \hline X & 4 & 3 \\ Y & 3 & 8 \end{array}$. What is the variance of data set Y?

(a) 4
(b) 3
(c) (3,4)
(d) 8

Answer: d

2. Which of the following is not a filtering method

(a) Chi-Square
(b) Variance Threshold
(c) Principal Component Analysis
(d) None of the above

Answer: c

ALM

- Construct the correlation matrix for the following three-dimensional dataset

X	Y	Z
15	12.5	50
35	15.8	55
20	9.3	70
14	20.1	65
28	5.2	80

ALM

- Find the eigen values and eigenvectors of 3 × 3 matrix $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

TERMINAL QUESTIONS

1. Differentiate between Feature selection methods and Feature Extraction methods?
2. Differentiate between Wrapper model and Filtering model?
3. Define Dimensionality reduction and list its advantages.
4. List different Dimensionality reduction techniques.

REFERENCES

Reference Books:
1. Stephen Marsland "Machine Learning – An Algorithmic Perspective" - 2nd edition
2. Tom M Mitchel "Machine Learning" A.S. Tanenbaum, David J. Wetheral "Computer Networks" Pearson, 5th –Edition.

Sites and Web links:
1. https://www.upgrad.com/blog/top-dimensionality-reduction-techniques-for-machine-learning/)
2. https://www.analyticsvidhya.com/blog/2018/08/dimensionality-reduction-techniques-python/

THANK YOU

TEAM ML