Principal Component Analysis

The dimension of a dataset is defined as the number of input variables or features. Dimensionality reduction is the process of reducing the number of variables in a dataset. The curse of dimensionality exacerbates the difficulty of modelling predictive modelling tasks.

High-dimensional statistics and dimensionality reduction are commonly used in data visualisation. These methodologies can be used in applied machine learning to simplify categorization or regression datasets in order to improve predictive model fit.

This article will discuss dimensionality reduction in the context of machine learning.

After reading this, you will understand:

- Algorithms for machine learning frequently struggle with big feature sets as input.
- Dimensionality reduction is the process of reducing the number of features in an input.
- Dimensionality reduction techniques include feature selection, linear algebra, projection, and autoencoders.

Numerous Input Variables Problem

An excessive number of input variables can impair machine learning performance.

The input variables in a spreadsheet refer to the rows that are entered into a model to forecast the target variable. Features are considered to be input variables.

Columns represent dimensions and rows represent points in an n-dimensional feature space. This is a somewhat accurate geometric representation.

For example, the presence of multiple dimensions in the feature space results in a vast volume, and hence the points (rows of data) reflect a small and non-representative sample.

When working with huge datasets, the "curse of dimensionality" can significantly hinder the performance of machine learning systems.

Thus, it is often desirable to reduce the amount of input features.

This results in a reduction in the dimension of the feature space.

Dimensional Reduction

Reduce the number of variables in training data using the following techniques.

It is advantageous to reduce the dimension of high-dimensional data by projecting it into a lower-level subspace that encapsulates its "essence." This is known as reduced dimensionality.

Hundreds of thousands, if not millions, of input variables are possible.

Reduced input dimensions frequently equate to fewer parameters or a more simplified machine learning model structure, referred to as degrees of freedom. A model with an excessive number of degrees of freedom may overfit the training data, resulting in poor performance on new data.

Simple models with a high degree of generalisation, as well as input data with few variables, are recommended. In general, the number of inputs and degrees of freedom are tightly related in linear models.

High-dimensional functions can be far more challenging than low-dimensional functions, and the resulting difficulties are more difficult to recognise. The only way to defeat the curse is to possess the necessary facts.

Dimensionality reduction is a technique for pre-modeling data preparation. Prior to training a predictive model, it is possible to do data cleansing and scaling.

As a result, the user's attention is drawn to the most critical factors.

As a result, any dimension reduction performed on training data must also be performed on fresh data, including test, validation, and prediction data.

Methods for Dimensionality Reduction

There are numerous ways for lowering dimensionality.

Now, let's review the critical techniques.

Feature Selection Methods

The most common is feature selection, which uses scoring or statistical methods to determine which features should be retained and which should be discarded.

Two distinct types of feature selection approaches are wrapper and filter methods.

They fit and evaluate the model using a variety of subsets of the input features and then choose the subset that produces the greatest model performance. A technique similar to RFE for selecting wrapper features.

Filter methods employ scoring techniques such as feature-target correlation to choose the most accurately predicted input features. Examples include correlation and Chi-Squared testing.

Matrix Factorization

Reduce dimensionality using linear algebra techniques. Algorithms for matrix factorization can be used to deconstruct the matrix of a dataset. There are two types of decomposition: eigendecomposition and singular value decomposition.

After prioritising the segments, a subset can be chosen that best captures the matrix's obvious structure.

Principal component analysis (PCA) is the most often used technique for component ranking. and dimensionality reduction.

Manifold Learning

Additionally, techniques from high-dimensional statistics can be used to minimise dimensionality. A projection is a function or mapping in mathematics that transforms data.

Learnvista Pvt Ltd.
2nd Floor, 147, 5th Main Rd, Rajiv Gandhi Nagar HSR Sector 7,Near Salarpuria Serenity, Bengaluru, Karnataka 560102
Mob:- +91 779568798, Email:- contacts@learnbay.co

These techniques are used to create a two-dimensional projection of three-dimensional data, which is frequently used for data visualisation. While the projection produces a two-dimensional picture of the data, it also keeps its critical structure and relationships.

Various modes of instruction include the following:

- Kohonen Self-Organizing Map (SOM).
- Sammons Mapping
- Multidimensional Scaling (MDS)
- t-distributed Stochastic Neighbor Embedding (t-SNE).

Typically, the projection's characteristics bear no resemblance to the original columns, which may be perplexing to novices.

Autoencoder Methods

Deep learning neural networks are capable of reducing the dimension.Autoencoders are a common technique. In a self-supervised learning problem, a model must accurately recreate the input.

By reducing the dimensions of the data flow to a bottleneck layer that is significantly smaller than the original input data. The encoder reads the bottleneck output and reconstructs the input, whereas the decoder reads and reconstructs the bottleneck output.

An auto-encoder is a type of unsupervised neural network that is used to reduce the dimension of a dataset and to discover features. A feedforward neural network trained to anticipate its own input is called an auto-encoder.

After training, the decoder is removed and the output of the bottleneck is utilised to approximate the reduced dimensionality of the input. The output of this encoder can be fed into any model, not simply a neural network.

Deep autoencoders are quite successful at reducing nonlinear dimension. Once constructed, the encoder's top layer, code layer hc, can be used for supervised classification.

As is the case with other projection methods, the encoder's output bears no direct link to the original input variables, making interpretation challenging.

There is no apparent winner when it comes to dimensionality reduction.

Rather than that, conduct controlled experiments to determine which solutions for dimensionality reduction perform best with your model.

Typically, linear algebra and approaches for manifold learning assume that all input characteristics have the same scale or distribution. This implies that it is prudent to normalise or standardise data before applying these procedures if the input variables include a range of scales or units.

Finding Eigenvalues And Eigenvectors

EXAMPLE 1: Find the eigenvalues and eigenvectors of the matrix

$$A = \begin{pmatrix} 1 & -3 & 3 \\ 3 & -5 & 3 \\ 6 & -6 & 4 \end{pmatrix}$$

Solution:

• In such problems, we first find the eigenvalues of the matrix.

Finding eigenvalues

• To do this, we find the values of λ which satisfy the characteristic equation of the matrix A, namely those values of λ for which
det(A − λI) = 0,

where I is the 3×3 identity matrix.
• Form the matrix A − λI:
A − λI=

$$\begin{pmatrix} 1 & -3 & 3 \\ 3 & -5 & 3 \\ 6 & -6 & 4 \end{pmatrix} - \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix} = \begin{pmatrix} 1-\lambda & -3 & 3 \\ 3 & -5-\lambda & 3 \\ 6 & -6 & 4-\lambda \end{pmatrix}$$

Notice that this matrix is just equal to A with λ subtracted from each entry on the main diagonal.

• Calculate det(A − λI):

det(A − λI) = $(1-\lambda)\begin{vmatrix} -5-\lambda & 3 \\ -6 & 4-\lambda \end{vmatrix} - (-3)\begin{vmatrix} 3 & 3 \\ 6 & 4-\lambda \end{vmatrix} + 3\begin{vmatrix} 3 & -5-\lambda \\ 6 & -6 \end{vmatrix}$

= (1 − λ) ((−5 − λ)(4 − λ) − (3)(−6)) + 3(3(4 − λ) − 3 × 6) + 3(3 × (−6) − (−5 − λ)6)

= (1 − λ)(−20 + 5 λ − 4 λ +λ^2+ 18) + 3(12 − 3λ − 18) + 3(−18 + 30 + 6λ)

= (1 − λ)(−2 + λ + λ^2) + 3(−6 − 3λ) + 3(12 + 6λ)

= −2 + λ +λ^2 + 2λ − λ^2− λ^3 − 18 − 9λ + 36 + 18λ

= 16 + 12λ − λ^3

• Therefore

det(A − λI) = −λ^3+ 12λ + 16.

REQUIRED: To find solutions to det(A − λI) = 0 i.e., to solve

$$\lambda^3 - 12\lambda - 16 = 0 \qquad (1)$$

* Look for integer valued solutions.

* Such solutions divide the constant term (-16). The list of possible integer solutions is

±1, ±2, ±4, ±8, ±16.

* Taking λ = 4, we find that 4^3 − 12.4 − 16 = 0.
* Now factor out λ − 4:

(λ − 4)(λ^2 + 4λ + 4) = λ^3 − 12λ^2 + 16.

* Solving λ^2 + 4λ + 4 by formula1 gives

$$\lambda = \frac{-4 \pm \sqrt{4^2 - 4.1.4}}{2} = \frac{-4 \pm 0}{2},$$

and so λ = −2 (a repeated root).

• Therefore, the eigenvalues of A are λ = 4, −2. (λ = −2 is a repeated root of the characteristic equation.)

Finding eigenvectors

• Once the eigenvalues of a matrix (A) have been found, we can find the eigenvectors by Gaussian Elimination.

Learnvista Pvt Ltd.

2nd Floor, 147, 5th Main Rd, Rajiv Gandhi Nagar HSR Sector 7,Near Salarpuria Serenity, Bengaluru, Karnataka 560102

Mob:- +91 779568798, Email:- contacts@learnbay.co

• STEP 1: For each eigenvalue λ, we have

(A − λI)x = 0,

where x is the eigenvector associated with the eigenvalue λ.

• STEP 2: Find x by Gaussian elimination. That is, convert the augmented matrix

$$(A - \lambda I \,\vdots\, 0)$$

to row echelon form, and solve the resulting linear system by back substitution.

We find the eigenvectors associated with each of the eigenvalues

• Case 1: λ = 4

– We must find vectors x which satisfy (A − λI)x = 0.

– First, form the matrix A − 4I:

$$A - 4I = \begin{pmatrix} -3 & -3 & 3 \\ 3 & -9 & 3 \\ 6 & -6 & 0 \end{pmatrix}$$

– Construct the augmented matrix $(A - \lambda I \,\vdots\, 0)$ and convert it to row echelon form

$$\begin{array}{cc} \begin{pmatrix} -3 & -3 & 3 & 0 \\ 3 & -9 & 3 & 0 \\ 6 & -6 & 0 & 0 \end{pmatrix} & \begin{array}{c} R1 \\ R2 \\ R3 \end{array} \end{array} \xrightarrow{R1 \to -1/3 \times R3} \begin{array}{cc} \begin{pmatrix} 1 & 1 & -1 & 0 \\ 3 & -9 & 3 & 0 \\ 6 & -6 & 0 & 0 \end{pmatrix} & \begin{array}{c} R1 \\ R2 \\ R3 \end{array} \end{array}$$

$$\xrightarrow[R3 \to R3 - 6 \times R1]{R2 \to R2 - 3 \times R1} \begin{array}{cc} \begin{pmatrix} 1 & 1 & -1 & 0 \\ 0 & -12 & 6 & 0 \\ 0 & -12 & 6 & 0 \end{pmatrix} & \begin{array}{c} R1 \\ R2 \\ R3 \end{array} \end{array}$$

$$\xrightarrow{R2 \to -1/12 \times R2} \begin{array}{cc} \begin{pmatrix} 1 & 1 & -1 & 0 \\ 0 & 1 & -1/2 & 0 \\ 0 & -12 & 6 & 0 \end{pmatrix} & \begin{array}{c} R1 \\ R2 \\ R3 \end{array} \end{array}$$

Learnvista Pvt Ltd.
2nd Floor, 147, 5th Main Rd, Rajiv Gandhi Nagar HSR Sector 7,Near Salarpuria Serenity, Bengaluru, Karnataka 560102
Mob:- +91 779568798, Email:- contacts@learnbay.co

$$\xrightarrow{R3 \to R3 + 12 \times R2} \begin{array}{c} \begin{pmatrix} 1 & 1 & -1 & 0 \\ 0 & 1 & -1/2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{array} \begin{array}{c} R1 \\ R2 \\ R3 \end{array}$$

– Rewriting this augmented matrix as a linear system gives

$x_1 - 1/2x_3 = 0$

$x_2 - 1/2x_3 = 0$

So the eigenvector x is given by:

$$\begin{pmatrix} x_1 = \frac{x_3}{2} \\ x_2 = \frac{x_3}{2} \\ x_3 \end{pmatrix} = x_3 \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 1 \end{pmatrix}$$

For any real number $x_3 \neq 0$. Those are the eigenvectors of A associated with the eigenvalue λ = 4.

• Case 2: λ = −2

– We seek vectors x for which (A − λI)x = 0.

– Form the matrix A − (−2)I = A + 2I

$$A + 2I = \begin{pmatrix} 3 & -3 & 3 \\ 3 & -3 & 3 \\ 6 & -6 & 6 \end{pmatrix}.$$

– Now we construct the augmented matrix $(A - \lambda I : 0)$ and convert it to row echelon form

$$\begin{pmatrix} 3 & -3 & 3 & 0 \\ 3 & -3 & 3 & 0 \\ 6 & -6 & 6 & 0 \end{pmatrix} \begin{array}{c} R1 \\ R2 \\ R3 \end{array} \xrightarrow{R1 \to 1/3 \times R3} \begin{pmatrix} 1 & -1 & 1 & 0 \\ 3 & -3 & 3 & 0 \\ 6 & -6 & 6 & 0 \end{pmatrix} \begin{array}{c} R1 \\ R2 \\ R3 \end{array}$$

$$\xrightarrow[R3 \to R3 - 6 \times R1]{R2 \to R2 - 3 \times R1} \begin{pmatrix} 1 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{array}{c} R1 \\ R2 \\ R3 \end{array}$$

– When this augmented matrix is rewritten as a linear system, we obtain

$x_1 + x_2 - x_3 = 0,$

so the eigenvectors x associated with the eigenvalue λ = −2 are given by:

$$x = \begin{pmatrix} x_1 = x_3 - x_2 \\ x_2 \\ x_3 \end{pmatrix}$$

– Thus

$$x = \begin{pmatrix} x_1 = x_3 - x_2 \\ x_2 \\ x_3 \end{pmatrix} x_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + x_2 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} for\ any\ x_2, x_3 \in R\{0\}$$

are the eigenvectors of A associated with the eigenvalue λ = −2.