

PCA

* Steps :

- 1- Standardize the range of continuous initial variables
- 2- Compute the covariance matrix to identify correlations
- 3- Compute eigenvectors and eigenvalues of the covariance matrix to identify the principle components
- 4- Create a feature vector to decide which principle components to keep
- 5- Recast the data along the principle components axes.

* Usage :

Dimensionality reduction



First step requirements:

* Skewness : measure of the asymmetry of a distribution.

- right skewness : no longer on right side
- left skewness : no longer on left side
- zero skewness : no longer on zero side.

* Descriptive statistic : summarize and organize characteristics of data set.

in quantitative research → ① collecting data

② statistical analysis (Avg / relation / ...)

③ inferential statistics (data confirm or refutes our hypothesis)

④ generalizable (larger population)

(count each group or value) → ① distribution : frequency of each value

(estimate center/avg using mean / median / mode) → ② central tendency : average of values

(standard deviation and variance affects spread) → ③ variability : how spread out the values are.

* standard deviation is average of variability in our dataset

(on average how far each score lies from mean)

(the larger standard deviation \rightarrow the more variable)

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

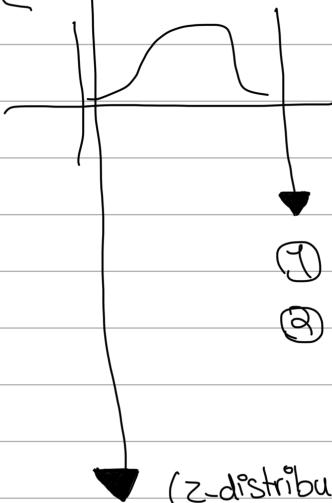
each data point mean
n-1 number of data in our dataset

* variance: average of squared derivations from mean.

(reflects the degree of spread in the data set)

(x-distribution)

* normal distribution: data is symmetrically distributed with no skew (most values clustering around central region)



↳ we can use inferential statistics to compare different groups and make estimation.

① mean / median / mode are the same

② the distribution can be described by two values (mean & standard deviation)

location parameter where the peak is centered
scale parameter squeezes/stretches the curve (width)

(z-distribution)

* standard normal distribution (z-distribution)

↳ mean = 0 / standard deviation = 1 / variance = 1

(squeezed or stretched) and moved horizontally (right or left)

↳ normal distribution can be converted to standard normal distribution by turning individual values into z-scores

* z-scores: how many standard deviations away from mean each value lies

↳ below or above the mean

$$z = \frac{x - \mu}{\sigma}$$

individual mean
standard deviation

positive z-score \rightarrow x-value is greater than the mean

negative z-score \rightarrow x-value is smaller than the mean

zero z-score \rightarrow x-value equals to mean

* converting normal distribution to standard normal distribution:

- ① calculate the probability of certain values occurring in our dataset
- ② compare datasets with different means and standard deviations
- ③ normalizing scores for statistical decision making



First step :

because of above reasons we want to standardize our input, we should convert individual values to Z-values using

$$Z = \frac{\text{Value} - \text{mean}}{\text{Standard deviation}} \rightarrow \text{this makes to have standard normal distribution.}$$

(assumption our input features are normal distribution but not standarized)



Second step prerequisites:

* Variance measure the variation of a single random variable.

* Covariance is a measure of how much two random variables vary together. (covariance value ↑ → more different)

$$\sigma_{(xy)} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

* covariance matrix : ① square matrix

② $C_{(ij)} = \sigma_{(xi,yj)}$

③ $C \in \mathbb{R}^{d \times d}$ (d number of random variables)

④ Symmetric $\sigma_{(xi,yj)} = \sigma_{(yj, xi)}$

⑤ diagonal entries of the covariance matrix are variances and other entries are the covariance

$$C = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T$$



⑥ positive value → positive relationship

shows covariance between

⑦ negative value → negative relationship

pair of datasets

⑧ positive semi-definite ($u = \text{column vector}, M = \text{covariance matrix} \Rightarrow u^T M u \geq 0$)

⑨ eigenvalues of covariance matrix are real and non-negative.

Population Variance: $\text{var}(x) = \frac{\sum_i^n (x_i - \mu)^2}{n}$
Population Covariance: $\text{cov}(x, y) = \frac{\sum_i^n (x_i - \mu_x)(y_i - \mu_y)}{n}$
Sample Variance: $\text{var}(x) = \frac{\sum_i^n (x_i - \bar{x})^2}{n-1}$
Sample Covariance: $\text{cov}(x, y) = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{n-1}$
μ = mean of population data.
\bar{x} = mean of sample data.
n = number of observations in the dataset.
x_i = observations in dataset x.
Using these formulas, the general form of a variance covariance matrix is given as follows:

* covariance of a variable with itself its variance

$$\text{cov}(x, x) = \text{var}(x)$$

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

- * Covariance = the direction of the linear relationship between variables
- Correlation = both the strength and direction of the linear relationship between two variables.
- ↑ not standardized
↓ standardized
can take value between $(-\infty, +\infty)$
↑ scales the value down to a limited range $(-1, +1) \rightarrow$ range of correlation
- Correlation is a function of covariance.
 - calculate correlation coefficient of two variables by dividing the covariance of these variables by the product of the standard deviation of the same values.
 - Correlation is dimensionless
 - Change in scale of variable affects covariance, but not correlation

Second step:

- * If two variables are highly correlated in such a way they contain redundant information.

* if covariance is positive: two variables increase or decrease together

~ ~ is negative: one increase and the other decrease

Third step prerequisites:

- * Matrices can be broken down or decomposed in ways that can show information about their functional properties

* One of the most widely decomposition: eigen-decomposition.

we decompose matrix into a set of eigenvectors and eigenvalues

* eigenvectors and values are used to determine a set of important variables

(in form of a vector)

↓
determine a set of important variables along with
scale along different dimensions (key dimensions based on variance)

- * principle components / dimensions = can be seen as eigen vectors with each one of them having its own elements

↓
Data → dependent variable

↓
Data / Data / Data

↓
Initial independent variables

- * The overall data can be seen as transformation matrix.

When data acted on eigen vectors (pc) → eigenvectors multiplied by scale factor (eigenvalue)

new variables that are constructed as linear combinations or mixtures of initial variables. The new variables (pc) are uncorrelated and most of

The information within the initial variables is squeezed or compressed into first component.

* When processing large volume of data representing different variable or dimensions.

↓
to use all of the features is computationally intense

↓
extract smaller set of variables (feature)

information are stored in variables

ensuring most of the information contained in original is retained.

Feature extraction \leadsto PCA and depends on eigenvalue and eigenvectors

- ① reduce dimensionality
② compressing

①

* in PCA we find eigenvalue & vectors of feature covariance matrix

② determine top k eigenvectors based on eigenvalues

③ create projection matrix from these eigenvectors

④ the projection matrix is used to transform the original feature into another feature subspace.

* eigenvalues and eigenvectors are used in transforming a large volume of

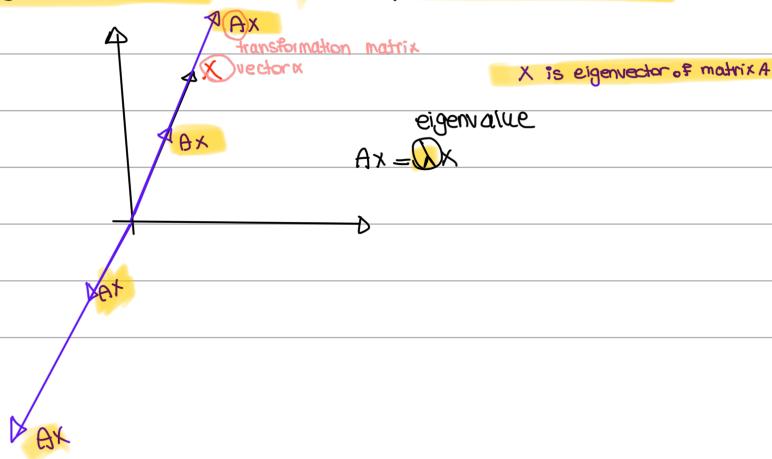
④

multi-dimensional data into another subspace comprising of smaller dimensions

② determine the optimal way to represent the data

③ calculate the distances between data points \rightarrow (used for grouping data based on similarity)

* Eigenvectors = vectors that when multiplied by matrix (linear combination or transformation) result in another vector having the opposite same direction but scaled in forward or reverse direction by scalar eigenvalue



- $\lambda \in \mathbb{R}, \lambda > 0: v$ and Av point in same direction
- $\lambda \in \mathbb{R}, \lambda < 0: v$ and Av point in opposite directions
- $\lambda \in \mathbb{R}, |\lambda| < 1: Av$ smaller than v
- $\lambda \in \mathbb{R}, |\lambda| > 1: Av$ larger than v

① Why do we calculate eigenvalue and vectors for correlation matrix?

because correlation matrix is multidimensional → we want to reduce dimensionality
contain important features, by keeping important features
→ find eigenvalues → create a project matrix → project corr to smaller dimension

When $Ax = \lambda x$ means we can replace A with lambda → means x is really a key component and very similar to A

third step

Compute eigen vectors and values of the cov matrix to identify principle components

① corr matrix is a multi dimensional matrix containing useful information

② we calculate eigenvectors, first because they are principle components of this matrix

second because by creating projection matrix using eigen vector we can transfer it to subspace

③ by having principle components we have information of all initial variables

remember if we have 10 -dimensionality corr matrix we will have 10 PCA, but most of important information are squeezed in first principle components.

fourth step:

* we told that the first principle component has the most information → this means the largest possible variance in the data

* by ranking our eigenvectors in order of eigenvalues, we will get PCA in order

fifth step:

* After ordering principle component using eigenvalues we need to discard those of less significance and form with the remaining ones a matrix called feature vector.

* Feature vector a matrix with its columns are eigenvectors of components we decide to stay.

* it's up to us to decide to keep a component or not.

we can use this metric: $\frac{\text{eigenvalue of the component}}{\text{sum of all eigen values}}$

it will help to understand how much information does our component have and decide to remove or not

Sixth step:



Reorient the data from original axes to the one represented by principle component

Final Dataset = (Featurevector)^T * (standardized original Dataset)^T