ILT 3 - Mathematics for Machine Learning

bangkit

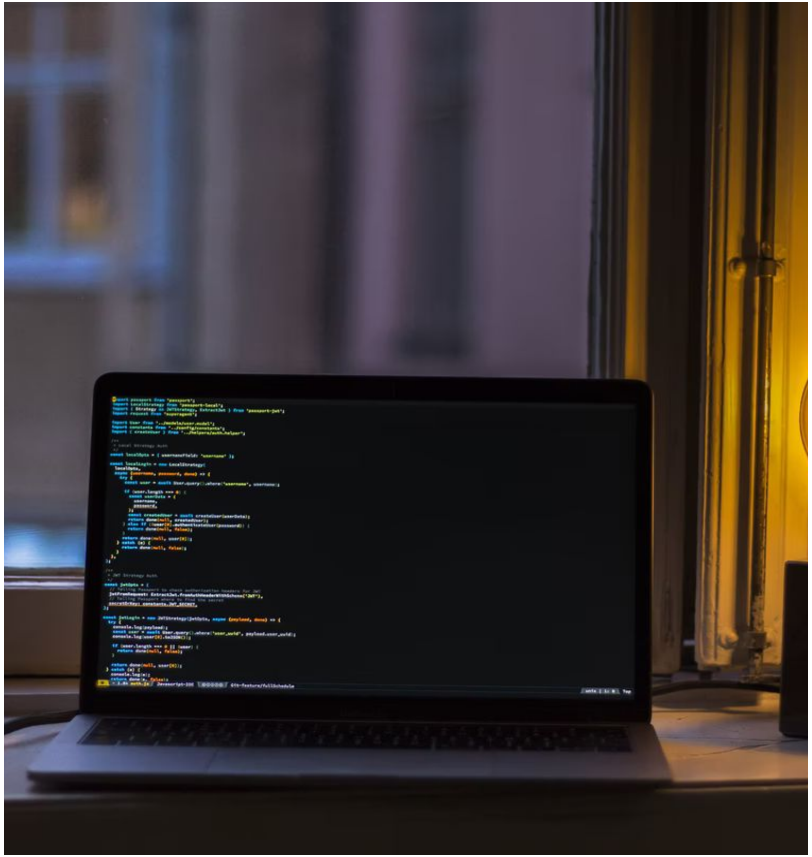
Ground Rules

Observe the following rules to ensure a supportive, inclusive, and engaging classes

- Give full attention in class
- Mute your microphone when you're not talking
- Keep your camera on
- Turn on the CC Feature on Meet
- Use raise hand or chat to ask questions
- Make this room a safe place to learn and share

Outline Session

- Introduction to Linear Algebra & Calculus
- Basic Statistics of Data
- PCA



bangkit

Introduction to Linear Algebra

What is Linear Algebra?

Linear algebra is a mathematical system for manipulating vectors in the spaces described by vectors.

 # Bedrooms

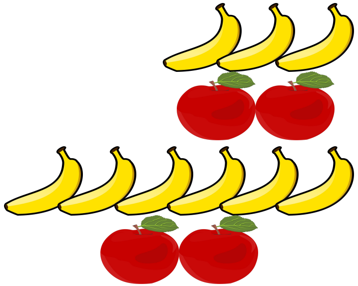
 # Bathrooms

Size

Sale Price

$$\begin{bmatrix} 3 \\ 1 \\ 130 \\ 340000 \end{bmatrix}$$

How to Manipulate a Vector?



Rp 8.000

Rp 13.000

Matric

Vector

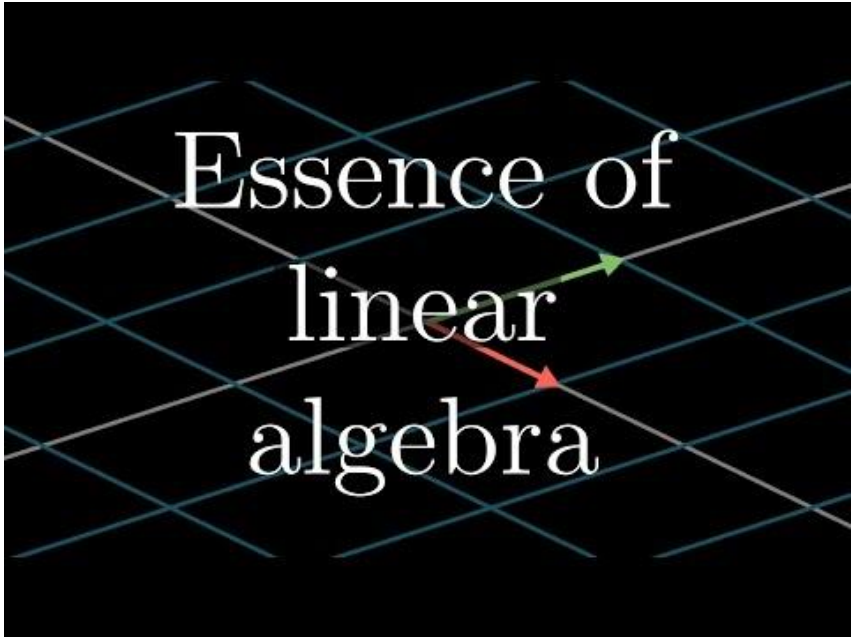
New Vector

$$\begin{bmatrix} 2 & 3 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 8000 \\ 13000 \end{bmatrix}$$

Matrix is an object that will transform a vector into a new vector.

Why Linear Algebra?

- Linear algebra is about vectors and matrices
- Linear algebra is essentially the mathematics of data



Important Concepts

Mathematics for Machine Learning: Linear Algebra

Formula Sheet

Vector operations

$$\mathbf{r}+\mathbf{s}=\mathbf{s}+\mathbf{r}$$
$$2\mathbf{r}=\mathbf{r}+\mathbf{r}$$
$$\|\mathbf{r}\|^2=\sum_i r_i^2$$

- *dot or inner product:*

$$\mathbf{r}.\mathbf{s}=\sum_i r_i s_i$$

$$\begin{aligned}\text{commutative} \quad & \mathbf{r}.\mathbf{s}=\mathbf{s}.\mathbf{r}\\ \text{distributive} \quad & \mathbf{r}.(\mathbf{s}+\mathbf{t})=\mathbf{r}.\mathbf{s}+\mathbf{r}.\mathbf{t}\\ \text{associative} \quad & \mathbf{r}.(a\mathbf{s})=a(\mathbf{r}.\mathbf{s})\end{aligned}$$

$$\mathbf{r}.\mathbf{r}=\|\mathbf{r}\|^2$$

$$\mathbf{r}.\mathbf{s}=\|\mathbf{r}\|\|\mathbf{s}\|\cos\theta$$

- *scalar and vector projection:*

$$\text{scalar projection:} \quad \frac{\mathbf{r}.\mathbf{s}}{\|\mathbf{r}\|}$$
$$\text{vector projection:} \quad \frac{\mathbf{r}.\mathbf{s}}{\mathbf{r}.\mathbf{r}}\mathbf{r}$$

Change of basis

Change from an original basis to a new, primed basis. The columns of the transformation matrix B are the new basis vectors in the original coordinate system. So

$$B\mathbf{r}'=\mathbf{r}$$

where r' is the vector in the B-basis, and r is the vector in the original basis. Or;

$$\mathbf{r}'=B^{-1}\mathbf{r}$$

If a matrix A is *orthonormal* (all the columns are of unit size and orthogonal to eachother) then:

$$A^T=A^{-1}$$

Gram-Schmidt process for constructing an orthonormal basis

Start with n linearly independent basis vectors $\mathbf{v}=\{\mathbf{v}_1,\mathbf{v}_2,...,\mathbf{v}_n\}$. Then

$$\mathbf{e}_1=\frac{\mathbf{v}_1}{||\mathbf{v}_1||}$$

$$\mathbf{u}_2=\mathbf{v}_2-(\mathbf{v}_2.\mathbf{e}_1)\mathbf{e}_1 \quad \text{so} \quad \mathbf{e}_2=\frac{\mathbf{u}_2}{||\mathbf{u}_2||}$$

... and so on for $\mathbf{u}_3$ being the remnant part of $\mathbf{v}_3$ not composed of the preceding $\mathbf{e}$-vectors, etc. ...

bangkit

* Formula Sheet

Eigenstuff

An eigenstuff (eigenvector & eigenvalue) is a set of tools for finding the characteristic properties of a matrix.

matrix

eigenvector

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

eigenvalue

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = 0$$

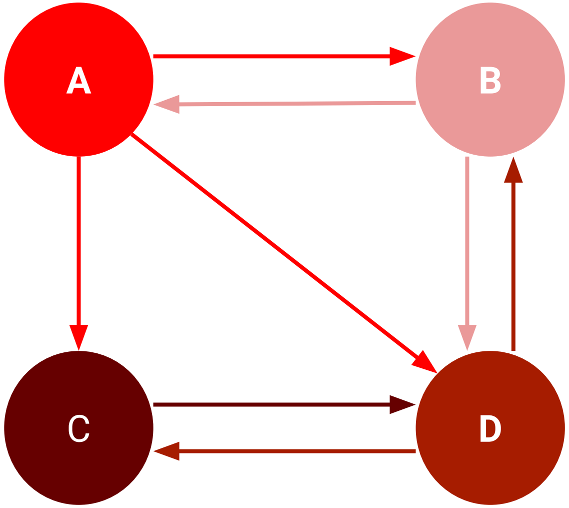
identity matrix

What are the examples of Linear Algebra in Machine Learning?

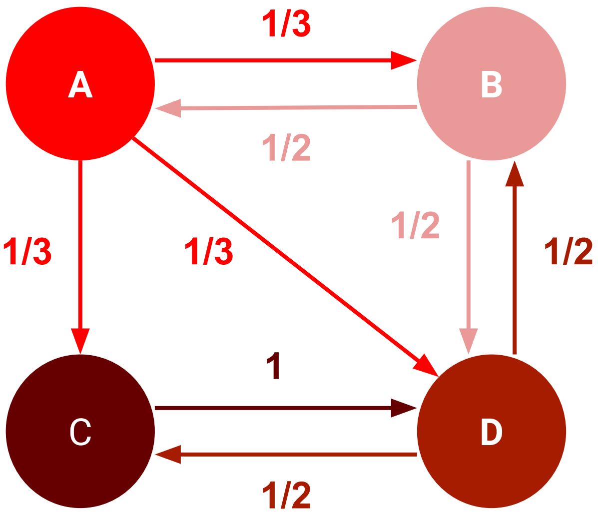
PageRank Algorithm

- The PageRank algorithm was published by Larry Page (Google founder) and colleagues in 1998.

- Used by Google to help them decide which order to display their websites when they returned from the search.



How PageRank Algorithm Works?



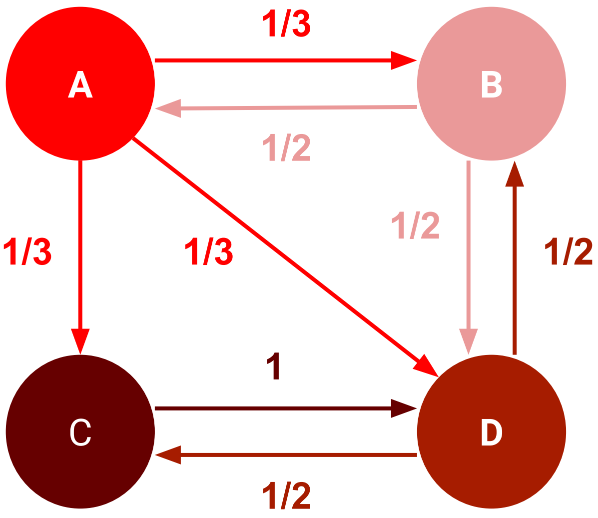
La = [0 ⅓ ⅓ ⅓]

Lb = [½ 0 0 ½]

Lc = [0 0 0 1]

Ld = [0 ½ ½ 0]

How PageRank Algorithm Works?



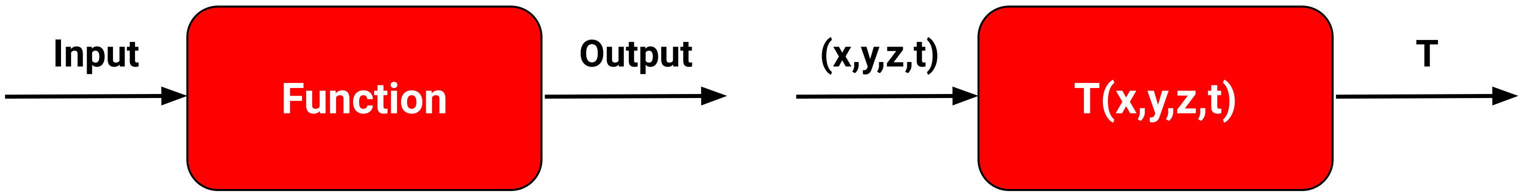
$$L = \begin{bmatrix} 0 & 1/2 & 0 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 1/2 & 1 & 0 \end{bmatrix}$$

$$r_{i+1} = Lr_i$$

Introduction to Calculus

bangkit

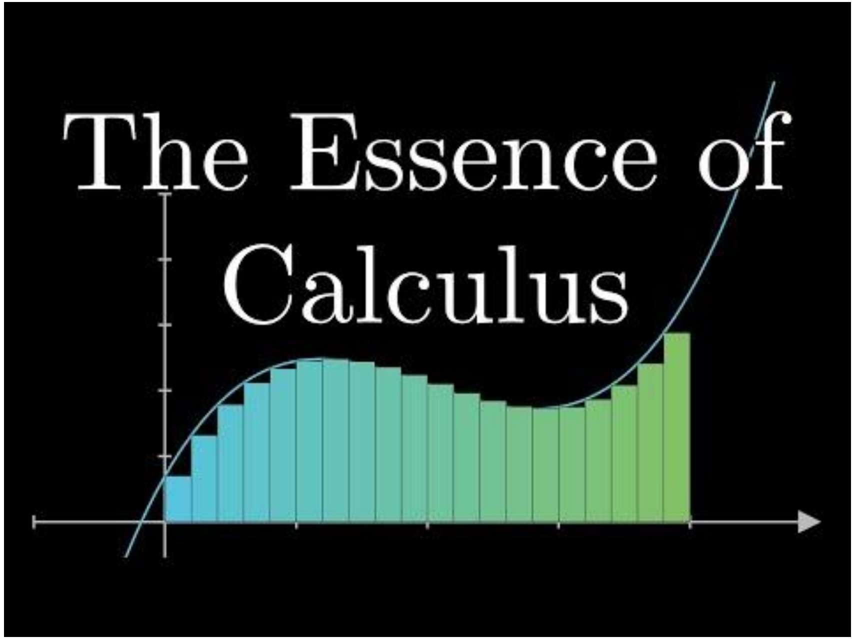
Concept of Function



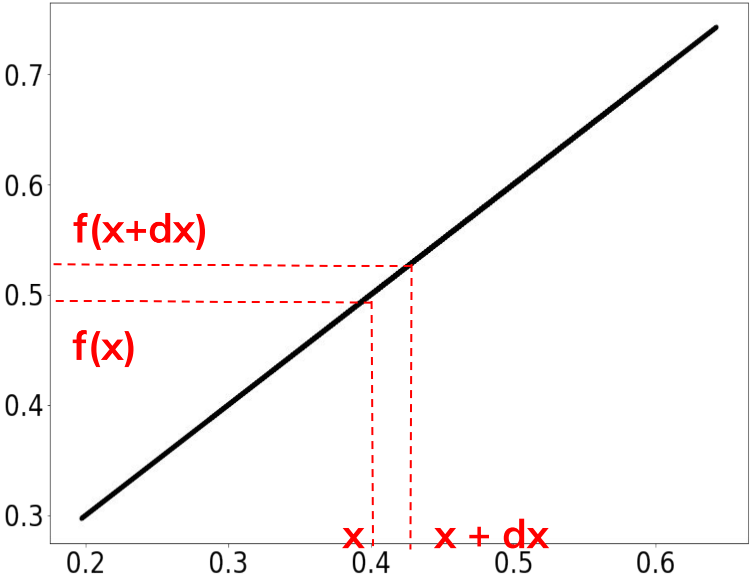
A function is a relationship between some inputs and an output.

Concept of Calculus

- Calculus is the study of how these functions change with respect to their input variables.
- Many important machine learning approaches have calculus at their core



Definition of Derivative



$$gradient = \frac{f(x+dx) - f(x)}{dx}$$

$$f'(x) = \lim_{dx \to 0} \frac{f(x+dx) - f(x)}{dx}$$

Important Concepts

Mathematics for Machine Learning

Multivariate Calculus

Formula sheet

Dr Samuel J. Cooper
Prof. David Dye
Dr A. Freddie Page

Definition of a derivative

$$f'(x) = \frac{\mathrm{d}f(x)}{\mathrm{d}x} = \lim_{\Delta x \to 0}\left(\frac{f(x+\Delta x) - f(x)}{\Delta x}\right)$$

Time saving rules

- *Sum Rule*:

$$\frac{\mathrm{d}}{\mathrm{d}x}(f(x) + g(x)) = \frac{\mathrm{d}}{\mathrm{d}x}(f(x)) + \frac{\mathrm{d}}{\mathrm{d}x}(g(x))$$

- *Power Rule*:

Given $f(x) = ax^b$,
then $f'(x) = abx^{(b-1)}$

Derivatives of named functions

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{x}\right) = -\frac{1}{x^2}$$
$$\frac{\mathrm{d}}{\mathrm{d}x}(\sin(x)) = \cos(x)$$
$$\frac{\mathrm{d}}{\mathrm{d}x}(\cos(x)) = -\sin(x)$$
$$\frac{\mathrm{d}}{\mathrm{d}x}(\exp(x)) = \exp(x)$$

Derivative structures

Given $f = f(x, y, z)$

* [Formula Sheet]

Basic Statistics of Data

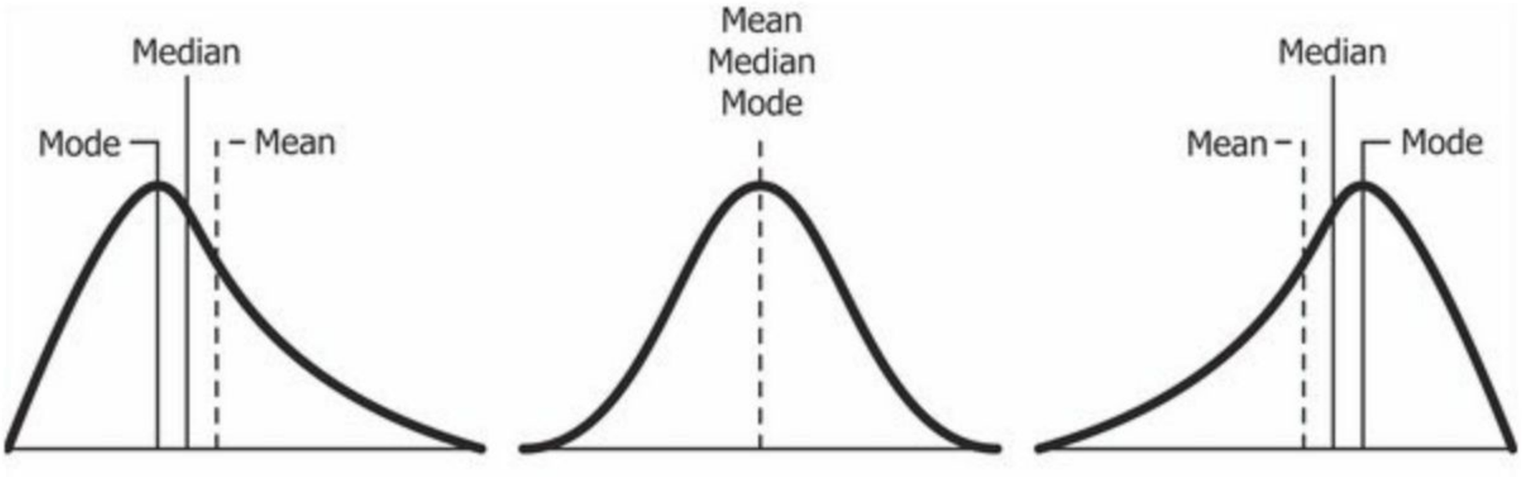
bangkit

Analyzing Quantitative Data

There are 4 main aspects of analyzing quantitative data

- **Shape** describes the distribution of data.
- **Spread** describes the variation of data.
  Two measures of spread are variance and standard deviation.
- **Outliers** is an observation that lies outside the overall pattern of a distribution
- **Center**
  The measure of center can give us an idea of the central position of the data set.
  - Mode
  - Median
  - Mean

Data **Distribution**

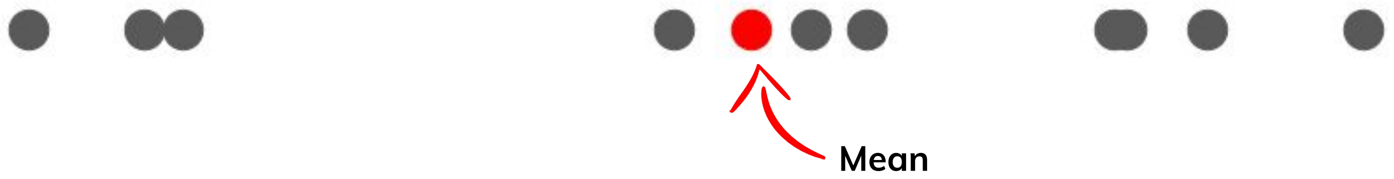


Right Skewed
Distribution

Normal
Distribution

Left Skewed
Distribution

Mean



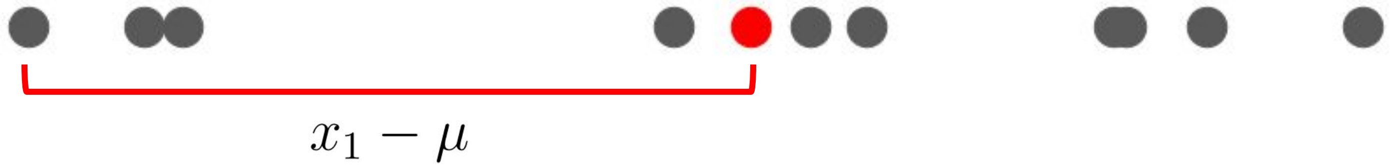
The mean of a data set describes the average data point.

$$D = \{x_1, x_2, x_3, \ldots, x_N\}$$

$$E[D] = \frac{1}{N} \sum_{n=1}^{N} x_n$$

bangkit

Variance



Variance is used to characterize
the variability or spread of data points in a dataset.

$$D = \{x_1, x_2, x_3, ..., x_N\}$$

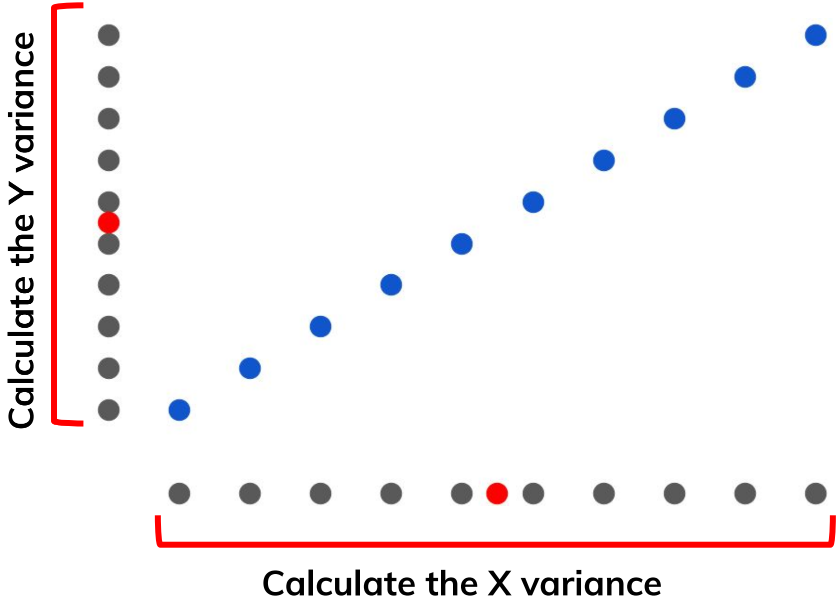
$$Var[D] = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu)^2 \quad \mu = E[D]$$

Why Variance is Important?

- Variance can describe how well the mean or median represents the data
- Variance can tell how much do we trust conclusions based on the mean and median.

bangkit

Variance 2D

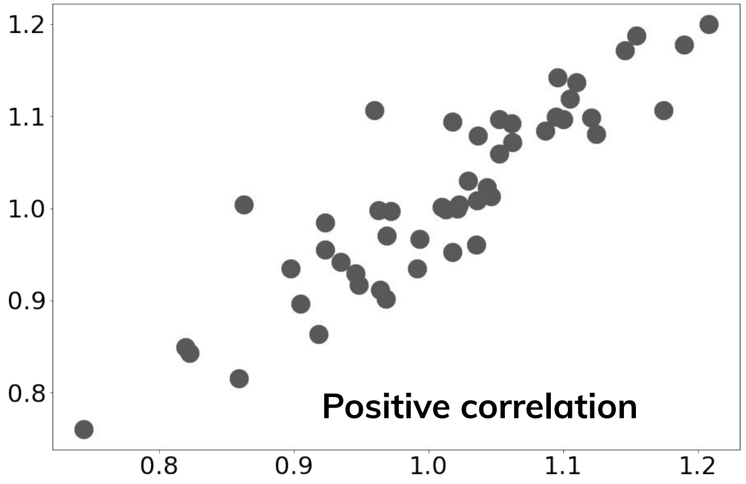


bangkit

Covariance

- The correlation between two variable can be captured by extending the notion of the variance to what is called the covariance of the data.

- There 3 types of covariance:
  - Positive correlation.
  - Negative correlation.
  - Zero correlation.



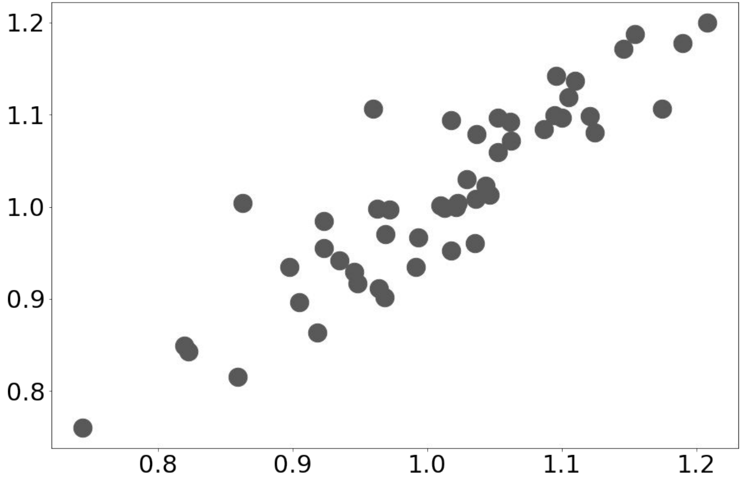
$$cov[x, y] = E[(x - \mu_x)(y - \mu_y)]$$

bangkit

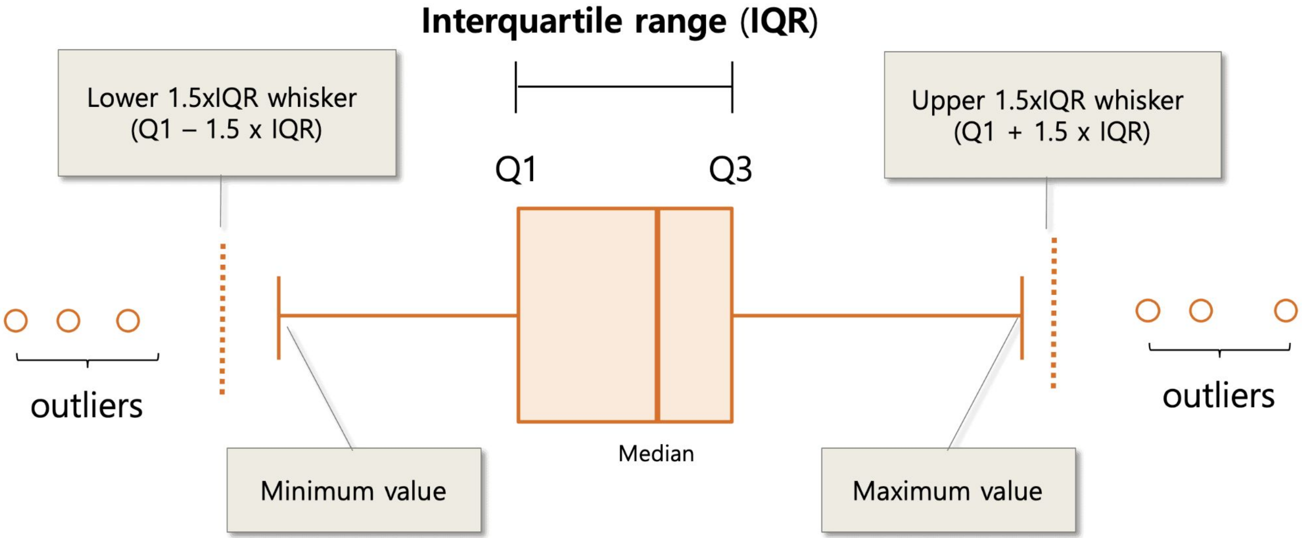
Covariance Matrix

- The variance of X
- The variance of Y
- The covariance of X,Y
- The covariance of Y,X

$$\begin{bmatrix} var[x] & cov[x,y] \\ cov[y,x] & var[y] \end{bmatrix}$$



Outliers



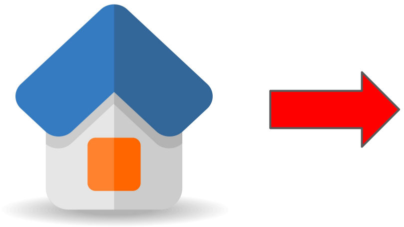
bangkit

PCA

Principal Component Analysis

Data in Real Life

- Data in real life is often high dimensional.
- Working with high dimensional data comes with some difficulties.
- Dimensionality reduction allows us to work with a more compact representation of the data.



 # Bedrooms

 # Bathrooms

Size

Sale Price

Crime Rate

Condition

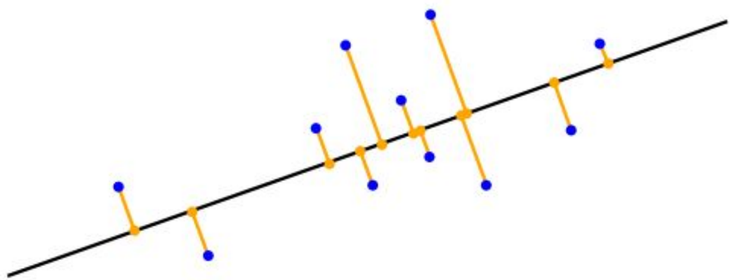
Type of House

Year Built

etc..

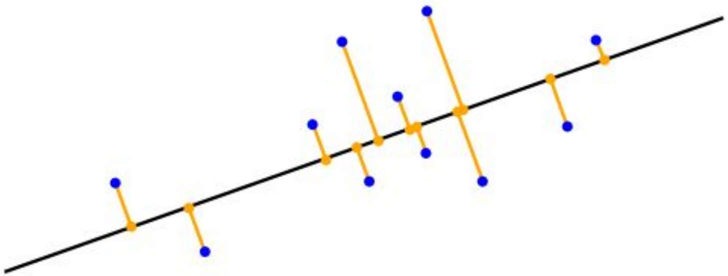
What is PCA?

- Principal Component Analysis (PCA) is a classical algorithm for linear dimensionality reduction
- PCA reduced the dimensionality of the data by projecting them into a lower-dimensional subspace
- In PCA we'll use some concepts such as eigenstuff, variance, and covariance

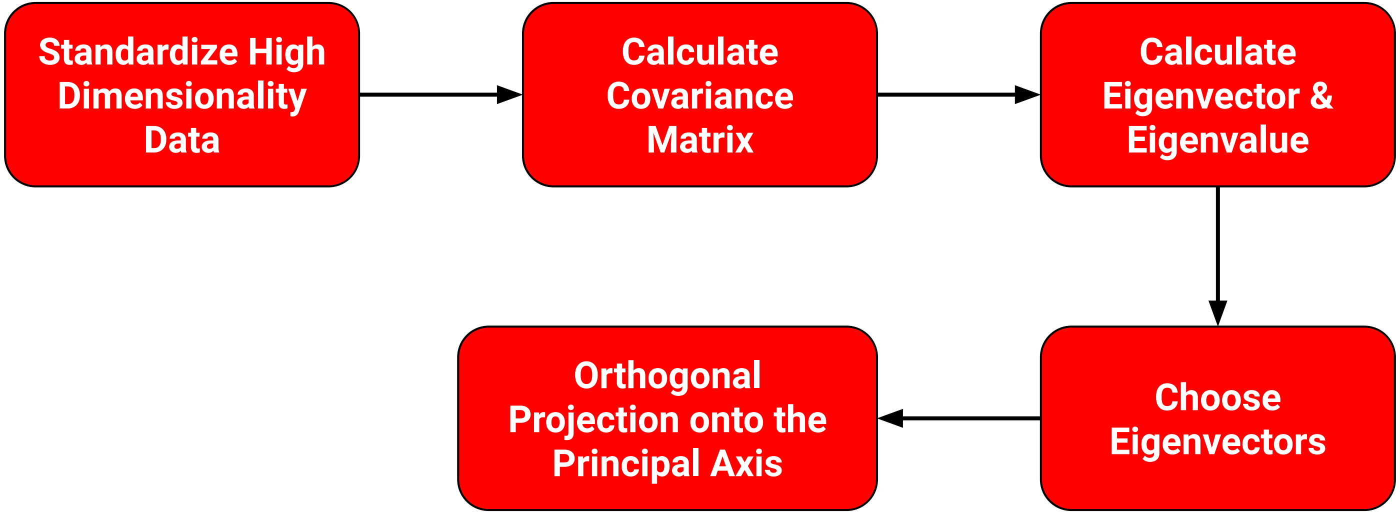


When Should We use PCA?

- If you want to reduce the dimensionality of the data but can't identify which variables to completely remove from consideration
- If you are comfortable making your data & model less interpretable
- If you want to ensure your variables are independent of one another

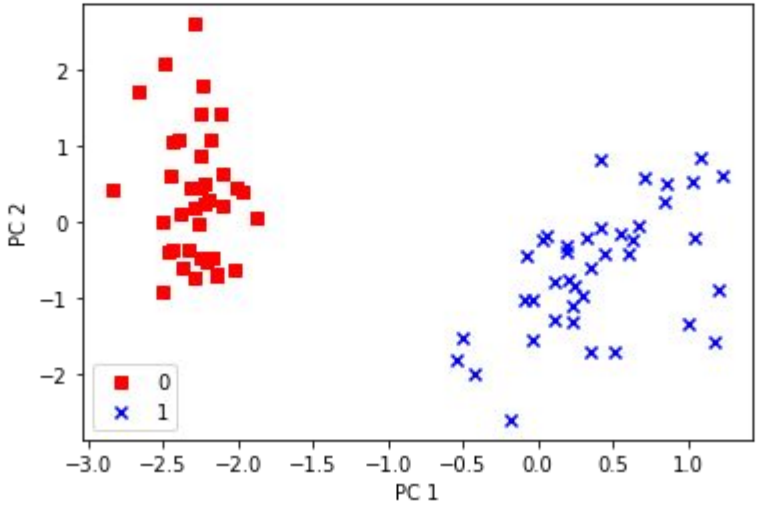


Key Step of PCA



How PCA Impact the Training Process?

- Speed up the training process by reducing the dimensionality of the data
- Improve the accuracy of the classification model



Demo Link

Demo PCA from scratch & use scikit-learn:
https://colab.research.google.com/drive/1tqCvpEcDDB9_kZ0PihRLDg8ZWmyJCDcg?usp=sharing

Demo Improve the Accuracy of the Classification Model using PCA :
https://colab.research.google.com/drive/1qKSszYc3TMTEP7xsHbq0X-i_AluJqNtw?usp=sharing

Sharing Session

Discussions

Quiz

Thank You