★Shallow Copy vs Deep Copy

↓	↘
→ pointer points to the Same copy of objects Of class	Create copy of each object inside of Class
→ Store refrences of object to original Memory Address	→ Store copies of object's value.
→ reflect changes made to the new/copied object at Originally object	→ Doesn't reflect changes
→ faster	→ comparitively slower

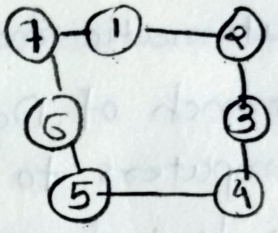
---

ndarray vs array

Data Science Life Cycle :-

- Refer Series of steps - follow by data Scientists - working of Data Driven program
- Include : data - cleaning, preparation, modelling, model, etc.
- lengthy procedure & quitly take few months.
- essential to have generic structure - CRISP - DM FrameWork.

1. Business Understanding
2. Data Understanding
3. Data Preparation
4. Exploritory Data Analysis
5. Data Modeling
6. Data Evulation
7. Model Deployment



1. Business Understanding
- Enterprise goal
- aim of analysis
- aim of Evaluation

→ pridiction

3. preparation of Data.
- choosing applicable Data, Integrate the data

→ formate data into preferred Structure

→ Time Consuming, arguably the most esential Step

→ Model will be accurate as your data

5. Data Modeling :-
- Cornory heart of data Analysis
- Organized Data - input | pref. output
- Selecting Model
- hyperparameter

2. Data Understanding.
- Series of all reachable data
- Explore information using graphical plot
- Exploring data

4. Exploratory Data Analysis.
- getting Some Concept about the answer and elelment affecting it.
- Discover each and Every char indi. & by means Combine them with diff. feature.

6. Data ~~M~~ Evalution.
- geared up to deploy

7. Model Deployment

→ Application of Datascience.

-Search Engine
→ gaming
→ Health Sector

- finance
- Business
- AirLine

- Medical
- Image Recognize
- E-commerce

→ NLP

- Natural Language Processing
- Automatic manipulation of Natural languages
- Branch of Data Science that focus on training computers to process and interpret conversation in text format in a way human do by listen.

→ filling gap between Data Science & human language

→ Difficult & challenging during Development as computer require humans to interect with them Using programming Languages like Java, python, etc.

↓ Structure & unambiguous

⟹ Use Case

1. spam detection
2. Machine translation
3. Virtual Agents & chatbot
4. Social Media Sentiment Analysis
5. Text summerization

☆ Computer Vision

→ field of artificial Intelligence
- train Computers to intrup & understand
  - visual world

→ Machine - accuratly identify objects - then react

Work

→ need lots of data

→ performed over + over analyses of data.
analysis,
- Until it recognizes output

Ex :- tire

→ 2 Essential technologies :- Machine Learning (Deep Lea.)
:- Convolution neural Network (CNN)

Application

1. IBM - 2018, My moments - Masters golf g tournament
2. google translator
3. Self-Driving vehicals
4. IBM - partnership - Verizon to bring AI to Edge.

Big Data

-> collection of huge data, yet growing rapidly with time.

-> Complex & huge in size - none Data Managment System can store it.

-> System - process and store huge data become common component in Big Data Managment in organization - combined tool that support Big Data uses.

-> Charachatrize by three v's :-

① large volume - mermy Enviroment
② wide Varitety - stored in Big Data System
③ velocity - much data is generated, collected & processed.

-> Application

-> Banking
-> Education
-> Healthcare
-> Agri
-> agriculture
-> Media
-> Goverment
-> finance
-> E-commerce
-> Retail

-> Issue In ~~Big Da~~ Data Science.

- empowring companies
- data constantly - has to be handle & cant ignored
- Data Scientist Collect Data Set, remove Data, Analyze.

① Identifying problem
② Fiding appropite data
③ workforce
④ Clensing.

Scalar

- Singal Numerical Value

Ex [0] [1]

Vector

- Array of Numbers

Ex [0,1,2] [3,8,9]

Matrix

- 2D Array of Shape with m rows & n columns.

Ex $\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$ $\begin{bmatrix} 9 & 0 \\ 10 & 1 \end{bmatrix}$

Descriptiv Statistics

- Describing & Summerizing data ~~numerically~~.
- 2 approches
  1. Quentitve Approche - numerically
  2. Visual Approch - illustrate data with graph, charts, etc.

Types of Mesure

① Central Data (tendency) - Cente of data
- Mean, Median, Mode

② Varibility - Spread of data
- Variance & Standard deviation

③ Correlation or Joint Valiblity :- relation between pair of veriablels.
- corrdation coefficient , covariation

Mean

- The average
- Central value of finite set of number

Ex

$x_1, x_2, \dots x_N$

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

Median

- middle element of dataset.
- Stored in incresing or decreasing
- → Odd. ⇒ middle Value.
- → EVEN ⇒ two value at middle position

Mode

- → Most frequently
- → isn't Such single value, then Set is multimodel Since it has multiple model value.

Variance

→ Messure how far the data points are spread out from the average value, and is equal to the Sum of Squeres of diff between the data value and the average.

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

Standard deviation

⇒ Squre root of Varience and Messure the Extent to which data varies from its Mean

→ prefered over varience.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}}$$

—› popular tools used in data Science.

- Data-preprocessing & Analysis
- -› Data Exploration & visualization
- -› Parallei and distributed Computing incase of BigData

—› Python as programming language.

- -› Support Multiple programming paradigm
- -› Dynamic typing
- -› Reference Counts
- -› Late biding

☆20 Algorithm as described in zen of python by Tim Peters.

—› Data Science Application

1. Search Engine
2. Transport
3. Finance
4. E-commerce
5. Healthcare
6. Image recognization

Covariance

—› Mesure of the joint variability of two random variables & Describes relationship between them

—› Defined as expected value of the product of the two random variables's deviation from their means.

$$\text{covariance}(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

number of Data   Data Value of x   Mean of x

(1) positive

-› heading in Same direction

(2) Negative

-› heading in appoite D

EDA [Exploratory Data Analysis]

→ Critical Step - investigate and Summarize - main char. of Dataset

→ Help to Understand data, identify pattern & uncover insights

1. Loading the Data

- Start by Loading Dataset.

```
import pandas as pd
df = pd.read_csv('your_dataset.csv')
```

2. Inspecting the Data :-

- check first few row and basic info

```
print(df.head())
print(df.info())
```

3. Handling Missing value.

- Identify & Handle Missing value.

```
print(df.isnull().sum())
df.fillna(df.mean(), inplace=True)
```

4. Descriptive Statistics :

- calculate summary Statics to understand data's central tendency & varibility

```
print(df.describe())
```

5. Data Visualization :

- Create visualization to better understand the data.
- lib :- Matplotlib , Seaborn

```
import matplotlib.pyplot as plt
import Seaborn as sns

    plt.hist(df['Age'])
    plt.xlable('Age')
    plt.ylable('Frequency')
    plt.show()
```

6. Feature Relationships.

- Explore relation beetween Variable.

```
correlation-matrix = df.corr()
sns.heatmap(correlation-matrix, annot=True)
Plt.show()
```

7. Data transformation

- perform Data transform is needed

```
df = pd.get_dummies(df, columns=['category'])
from sklearn.preprocessing import MinMax Scaler
scaler = MinMaxScaler()
df['Age'] = scaler.fit_transform(df[['Age']])
```

8. Out~~er~~liner Detection:

- Identify + Deal with outliner in data

9. feature Engineering.

- create new feature or drive insight from existing features

```
df['Total-score'] = df['score1'] + df['score2']
```

10. final Summary:-

Provide a final summary of your finding, insight, and any recommandation based on your analysis.