**IIT-HYDERABAD**

**AI WORKSHOP DAY-2 (09-04-2023)**

**(Lib.gl.is – json dolmey machine learning book)**

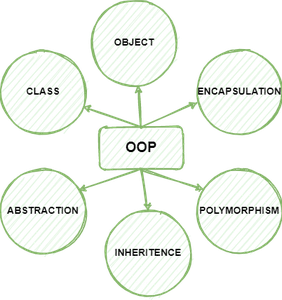
**OBJECT ORIENTED PROGRAMMING:**

In Python, object-oriented Programming (OOPs) is a programming paradigm that uses objects and classes in programming. It aims to implement real-world entities like inheritance, polymorphisms, encapsulation, etc. in the programming. The main concept of OOPs is to bind the data and the functions that work on that together as a single unit so that no other part of the code can access this data.

**Main Concepts of Object-Oriented Programming (OOPs)**

* Class
* Objects
* Polymorphism
* Encapsulation
* Inheritance
* Data Abstraction

Self.first -> to assign a value/ to initialise a value.



## Class

A class is a collection of objects. A class contains the blueprints or the prototype from which the objects are being created. It is a logical entity that contains some attributes and methods.

**Some points on Python class:**

* Classes are created by keyword class.
* Attributes are the variables that belong to a class.
* Attributes are always public and can be accessed using the dot (.) operator. Eg.: Myclass.Myattribute

-> the first variable is reserved for a class.

-> in instance method the first variable will be reserved for instance.

**Class Definition Syntax:**

class ClassName:

# Statement-1

.

.

.

# Statement-N

VARIABLE = reserved for class/instance.

Class variable can be changed using this.

-> class variable will not have any impact until it is altered.

**Virtual environment:**

A virtual environment is a tool that helps to keep dependencies required by different projects separate by creating isolated python virtual environments for them. This is one of the most important tools that most Python developers use.

**Installing virtualenv:**

$ pip install virtualenv

Test your installation:

$ virtualenv --version

You can create a virtualenv using the following command:

$ virtualenv my\_name

**PYCHARM =** used to create a project.

**NUMPY:**

NumPy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays. It is the fundamental package for scientific computing with Python. It is open-source software. It contains various features including these important ones:

* A powerful N-dimensional array object
* Sophisticated (broadcasting) functions
* Tools for integrating C/C++ and Fortran code
* Useful linear algebra, Fourier transform, and random number capabilities

# numpy.linspace() in Python:

The **numpy.linspace()** function returns number spaces evenly w.r.t interval. Similar to [numpy.arange() function](https://www.geeksforgeeks.org/numpy-arange-python/) but instead of step it uses sample number.

**Syntax :**

numpy.linspace(start,

stop,

num = 50,

endpoint = True,

retstep = False,

dtype = None)

**Parameters :**

-> **start :** [optional] start of interval range. By default start = 0

-> **stop :** end of interval range

-> **restep :** If True, return (samples, step). By default restep = False

-> **num :** [int, optional] No. of samples to generate

-> **dtype :** type of output array

**Return :**

-> **ndarray**

-> **step :** [float, optional], if restep = True

**PANDAS:**

Pandas is a Python library used for working with data sets.

It has functions for analyzing, cleaning, exploring, and manipulating data.

The name "Pandas" has a reference to both "Panel Data", and "Python Data Analysis" and was created by Wes McKinney in 2008.

WHY Pandas?

Pandas allows us to analyze big data and make conclusions based on statistical theories.

Pandas can clean messy data sets, and make them readable and relevant.

Relevant data is very important in data science.

Pandas gives you answers about the data. Like:

* Is there a correlation between two or more columns?
* What is average value?
* Max value?
* Min value?

Pandas are also able to delete rows that are not relevant, or contains wrong values, like empty or NULL values. This is called *cleaning* the data.

## Where is the Pandas Codebase?

The source code for Pandas is located at this github repository <https://github.com/pandas-dev/pandas>

**DEPENDENT vs INDEPENDENT VARIABLE:**

-> input variables are dependent variables.

-> DVs are called input featured terms.

-> labels are independent variables.

-> variables which are functionally defined are independent.

* -> The [**independent variable**](https://www.scribbr.com/methodology/independent-and-dependent-variables/#independent) is the cause. Its value is **independent**of other variables in your study.
* -> The [**dependent variable**](https://www.scribbr.com/methodology/independent-and-dependent-variables/#dependent) is the effect. Its value **depends**on changes in the independent variable.

-> Dependent variables are nothing but the variable which holds the phenomena which we are studying. Independent variables are the ones which through we are trying to explain the value or effect of the output variable (dependent variable) by creating a relationship between an independent and dependent variable.

**Regression:** if output is continuous.

-> Regression is a process of finding the correlations between dependent and independent variables. It helps in predicting the continuous variables such as prediction of **Market Trends**, prediction of House prices, etc.

**Classification:** if output is discrete.

-> Classification is a process of finding a function which helps in dividing the dataset into classes based on different parameters. In Classification, a computer program is trained on the training dataset and based on that training, it categorizes the data into different classes.

**fit()-function:**

The fit() method takes the training data as arguments, which can be one array in the case of unsupervised learning, or two arrays in the case of supervised learning.

**Multiple linear regression:**

Multiple Linear Regression is an extension of Simple Linear regression as it takes more than one predictor variable to predict the response variable. We can define it as:

Multiple Linear Regression is one of the important regression algorithms which models the linear relationship between a single dependent continuous variable and more than one independent variable.

**Some key points about MLR:**

* For MLR, the dependent or target variable(Y) must be the continuous/real, but the predictor or independent variable may be of continuous or categorical form.
* Each feature variable must model the linear relationship with the dependent variable.
* MLR tries to fit a regression line through a multidimensional space of data-points.

**-> Roboflow – online platform for detection applications.**