PROJECT REPORT ON

MACHINE LEARNING USING PYTHON

Submitted to:
Dr Sanjay Saxena

Submitted by:-
Debayan Das (B218058)
Prateek Sharma (B218031)
Sneha Shukla (B218045)

1.Introduction

Machine learning is a research field in computer science, artificial intelligence, and statistics. The focus of machine learning is to train algorithms to learn patterns and make predictions from data. Machine learning is especially valuable because it lets us use computers to automate decision-making processes. Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves.

The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly. ML is widely being used from banks to detect fraudulent activity in credit card transactions, and healthcare companies to monitor, assess, and diagnose patients.

2.What is Machine Learning?

Two definitions of Machine Learning are offered. Arthur Samuel described it as: "the field of study that gives computers the ability to learn without being explicitly programmed." This is an older, informal definition.

Tom Mitchell provides a more modern definition: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."

Example: playing checkers.

E = the experience of playing many games of checkers

T = the task of playing checkers.

P = the probability that the program will win the next game.

In general, any machine learning problem can be assigned to one of two broad classifications: Supervised learning and Unsupervised learning.

3.Machine Learning using Python

Python's popularity in machine learning may be due to the increased development of deep learning frameworks available for this language recently, including TensorFlow, PyTorch, and Keras. As a language that has readable syntax and the ability to be used as a scripting language, Python proves to be powerful and straightforward both for preprocessing data and working with data directly. The scikit-learn machine learning library is built on top of several existing Python packages that Python developers may already be familiar with, namely NumPy, SciPy, and Matplotlib.

4.Some machine learning methods

Machine learning algorithms are often categorized as.

1. Supervised
    1.1 Regression
    1.2 Classification
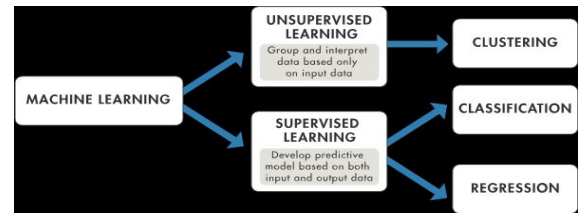
2. Unsupervised
    2.1 Clustering

Fig1: Machine Learning Methods

Supervised Learning

Supervised machine learning algorithms can apply what has been learned in the past to new data using labeled examples to predict future events. Starting from the analysis of a known training dataset, the learning algorithm produces an inferred function to make predictions about the output values. The system is able to provide targets for any new input after sufficient training. The learning algorithm can also compare its output with the correct, intended output and find errors in order to modify the model accordingly.

Regression:-Regression is basically a statistical approach to find the relationship between variables. In machine learning, this is used to predict the outcome of an event based on the relationship between variables obtained from the data-set. Linear regression is one type of regression used in Machine Learning.

Classification:- In machine learning and statistics, classification is a supervised learning approach in which the computer program learns from the data input given to it and then uses this learning to classify new observations.
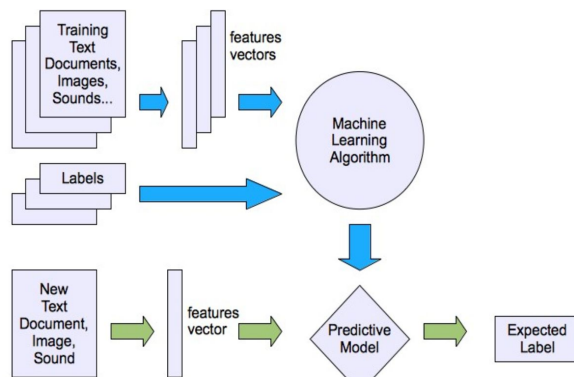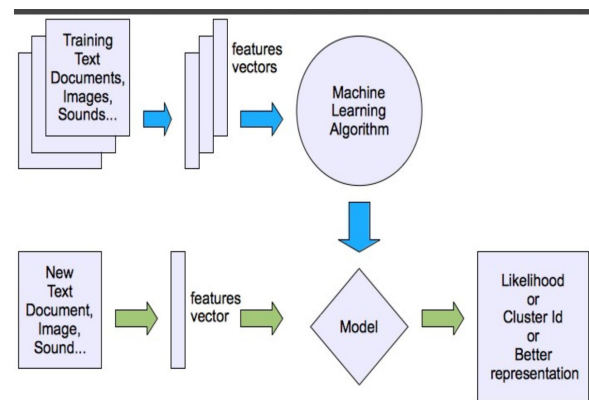
Fig2: Flowchart of supervised learning

Fig3: Flow chart of unsupervised learning

5.Basic Models Used in Machine Learning

Linear Regression: Linear Regression is usually the first machine learning algorithm that every data scientist comes across. It is a simple model but everyone needs to master it as it lays the foundation for other machine learning algorithms.

What is Linear Regression?

The objective of a linear regression model is to find a relationship between one or more features(independent variables) and a continuous target variable(dependent variable). When there is only one feature, it is called Uni-variate Linear Regression and if there are multiple features, it is called Multiple Linear Regression.

Where can Linear Regression be used?

It is a very powerful technique and can be used to understand the factors that influence profitability. It can be used to forecast sales in the coming months by analyzing the sales data

Unsupervised Learning

Unsupervised machine learning algorithms are used when the information used to train is neither classified nor labeled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabeled data. The system doesn't figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabeled data.

Clustering:-Clustering is a Machine Learning technique that involves the grouping of data points. Clustering is a method of unsupervised learning and is a common technique for statistical data analysis used in many fields.

for previous months. It can also be used to gain various insights about customer behaviour.

Code on Simple Linear Regression using one variable

```python
import numpy as np

import matplotlib.pyplot as plt


def estimate_coef(x, y):

    # number of observations/points

    n = np.size(x)


    # mean of x and y vector

    m_x, m_y = np.mean(x), np.mean(y)


    # calculating cross-deviation and deviation about x

    SS_xy = np.sum(y*x) - n*m_y*m_x

    SS_xx = np.sum(x*x) - n*m_x*m_x


    # calculating regression coefficients

    b_1 = SS_xy / SS_xx

    b_0 = m_y - b_1*m_x


    return(b_0, b_1)


def plot_regression_line(x, y, b):

    # plotting the actual points as scatter plot

    plt.scatter(x, y, color = "m",
```

```python
        marker = "o", s = 30)


    # predicted response vector

    y_pred = b[0] + b[1]*x


    # plotting the regression line

    plt.plot(x, y_pred, color = "g")


    # putting labels

    plt.xlabel('x')

    plt.ylabel('y')


    # function to show plot

    plt.show()

def main():

    # observations

    x = np.array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9])

    y = np.array([1, 3, 2, 5, 7, 8, 8, 9, 10, 12])


    # estimating coefficients

    b = estimate_coef(x, y)

    print("Estimated coefficients:\nb_0 = {}  \

    \nb_1 = {}".format(b[0], b[1]))
```

```
        # plotting regression line

        plot_regression_line(x, y, b)


if __name__ == "__main__":

        main()
```

Output

b_0 = 1.2363636363636363
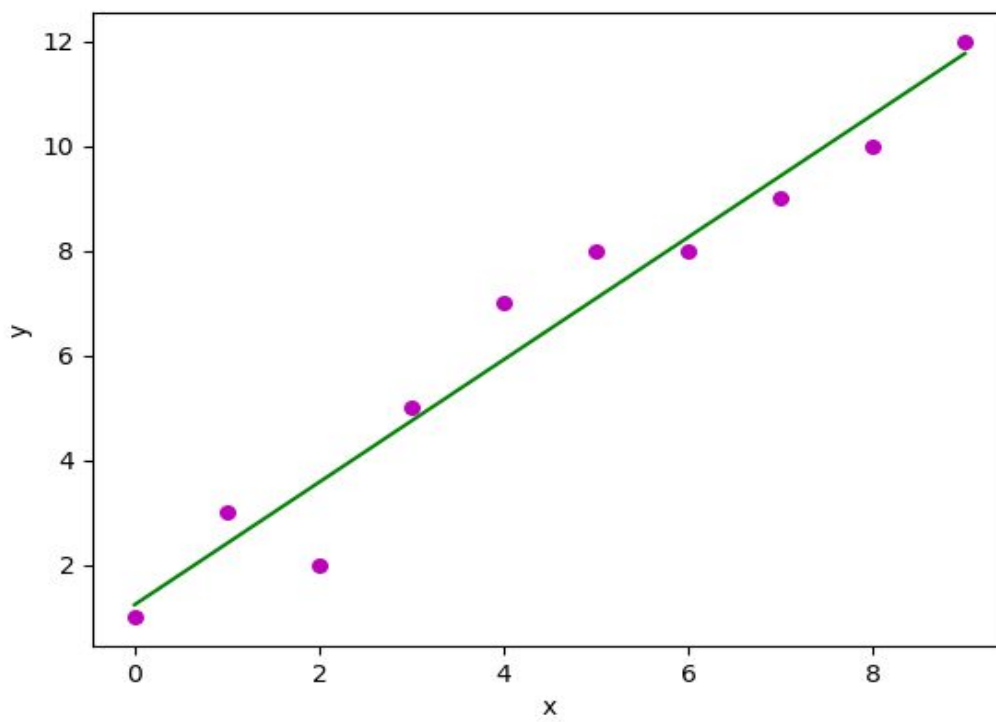
b_1 = 1.1696969696969697

Fig4: Linear regression plot

6.Logistic Regression

What is Logistic Regression?

Logistic regression is the appropriate regression analysis to conduct when the dependent variable is dichotomous (binary). Like all regression analyses, logistic regression is a predictive analysis. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables

Where can Logistic Regression be used?

Logistic regression is the appropriate regression analysis to conduct when the dependent variable is dichotomous (binary). Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables.

7.Applications

Image Recognition

It is one of the most common machine learning applications.There are many situations where you can classify the object as a digital image.Interpreting the visual world is one of those things that's so easy for humans we're hardly even conscious we're doing it.When we see something, whether it's car, or a tree, or our grandma, we don't (usually) have to consciously study it before we can tell what it is.For a computer, however, identifying a human being at all (as opposed to a dog or a chair or a clock) is difficult problem.Thus,through machine learning we teach the computer about an object by showing thousands of pictures of the object.

Medical Diagnosis

ML provides methods, techniques, and tools that can help in solving diagnostic and prognostic problems in a variety of medical domains. It is being used for the analysis of the importance of clinical parameters and of their combinations for prognosis, e.g. prediction of disease progression, for the extraction of medical knowledge for outcomes research, for therapy planning and support, and for overall patient management.ML is also being used for data analysis, such as detection of regularities in the data by appropriately dealing with imperfect data, interpretation of continuous data used in the Intensive Care Unit, and for intelligent alarming resulting in effective and efficient monitoring.It is argued that the successful implementation of ML methods can help the integration of computer-based systems in the healthcare environment providing opportunities to facilitate and enhance the work of medical experts and ultimately to improve the efficiency and quality of medical care.

Statistical Arbitrage

In finance, statistical arbitrage refers to automated trading strategies that are typical of a short-term and involve a large number of securities. In such strategies, the user tries to implement a trading algorithm for a set of securities on the basis of quantities such as historical correlations and general economic variables.These measurements can be cast as a classification or estimation problem. The basic

assumption is that prices will move towards a historical average.

Prediction

Consider the example of a bank computing the probability of any of loan applicants faulting the loan repayment.To compute the probability of the fault, the system will first need to classify the available data in certain groups.It is described by a set of rules prescribed by the analysts.

8.Companies using machine learning

1. Pinterest

One who is Pinterest user knows very well that Pinterest has attracted a large number of people surfing for the preferences; it has great influence over the internet and social media system. In 2015 Pinterest acquired Kosei, a machine learning automated system of a commercial application that allows specific content searching and recommendations. This touches the business operations and spam moderation of Pinterest, this also allows additions for the users

like email newsletter subscription, monetization of the advertising. This strengthens the ecosystem of the interest which is basically based on the curating existing online content and processing the effective advertisement in the basic priority.

2. Facebook

Introduction of facebook messenger has snuggled off the difference between the virtual and human conversation. This new feature allows any developer to create as well as submit what not, that is inclusively for facebook messenger. This makes retention of customer services more focused and easier for many startups who have limited engineering support. Facebook has introduced other machine learning applications too like computer vision algorithms which read the image and enhances the visual experience by uplifting the poor quality content. Facebook has also introduced the read option for visually impaired people where the image description can be read out. This new addition has given a different experimental testing dimension to the world's largest social platform.

3. Twitter

On a daily routine, we come across numerous controversies and many of them are created over twitter, why Twitter is catching so many eyes? Answer to this the outreach of the tweet to millions of people in just one tagging system with @ either in replies or comment section. One more aspect of machine learning technology-driven change is the addition of the algorithmically curates twitter timelines that evaluates each and every tweet in real time and score them according to internal scoring and produce them in a chronological driven manner. This allows maximum engagement on the displayed tweets done by the machine learning applications technology which studies the individual preferences and past data and publishes the algorithmically managed feeds which completely change the social media platform.

4. Google

Google is expanding its database on the basis of research and development which reflects the machine learning ambition of Google. Have you heard of deep mind network that has been recently launched by Google which is the recent update in neural network research of Google? This allows the virtual twist in the classical analog of the machine learning, this also includes natural language processing, speech recognition, and translation which further translates into search ranking and future prediction.

5. eCommerce companies

Gone are the days where retailers have to continue working on the connecting port between the stores and the consumers, online shopping has changed everything, the introduction of companies like edge case earlier known as comparing metric that has evolved machine learning technology to help

e-commerce website retailers improve the experience of shoppers. Increase in the conversion rate edge case streamlines the consumer experience by analyzing the behavioral pattern of the customer and their actions while browsing online content.

Conclusion

We learnt types of machine learning, common methods and popular approaches used in the field, why python is suitable machine learning programming language, and also covered its application in real life. Because machine learning is a field that is continuously being innovated, it is important to keep in mind that algorithms, methods, and approaches will continue to change and improvise.

Bibliography

- https://www.statisticssolutions.com/
- https://www.geeksforgeeks.org/
- https://www.wikipedia.org/
- https://expertsystem.com/
- https://www.techtarget.com/
- https://www.digitalocean.com/