REPORT02-ML

INTRODUCTION

What is Machine Learning?

Machine learning is the study of computer algorithms that can improve automatically through experience and by the use of data.?It is seen as a part of AI. The main idea is to build a model based on collected data.?

What is supervised learning?

Supervised learning is machine learning task of learning a function that maps an input to an output based on example input-output pairs. It means that we know both, input and expected output before running the code.

What is unsupervised learning?

Unsupervised learning is machine learning task which uses machine learning algorithms to analyze and cluster unlabeled datasets. It means that we can realy only on our code. We are not given any sort of guidance.

Which programming languages are most popular for ML?

1. Python
2. Java
3. C++
4. R
5. Javascript

Summary of common libraries

- NUMPY

It is a library which contain multisdimensional array and matrix data structures. It is well-known for it's speed of operating on arrays (e.g. sorting, discrete Fourier transforms, basic linear algebra, basic statistical operations). In Python there are lists that serve the purpose of arrays, but they are slow to process. Numpy is up to 50x faster than traditional Python lists. Numpy arrays are faster beacuse they are stored at one continuous place in memory unlike lists.

- PANDAS

?Pandas is a?software library written for the Python programming language for data manipulation and analysis.? Pandas contains tools for reading and writing data in different formats, merging and joining data sets and many others. The main aim of pandas is to a perfect library for data analysis.

- SCIKIT-LEARN

This library contains efficient tools for predictive data analysis. It provides us with Classification, Regression, Clustering and many others.?

- MATPLOTLIB

It is a library that is most used for dara visualizastion and plotting for Python. Matplotlib has its roots in MATLAB. One important big-picture matplotlib concept is its object hierarchy. A ?hierarchy? means that there is a tree-like structure of matplotlib objects underlying each plot.

- SEABORN

Seaborn is a Python data visualization library. It provides a efficient tools for drawing statistical graphics.

STEPS OF MACHINE LEARNING

1. Data Collection: It means to gather data which will be used in further steps
2. Data Preparation: It means to create numpy arrays, split data into sets, consider normalization/standardization or reduction of dimensions etc.
3. Chosing a Model: The aim is to chose the right model for a specific dataset?
4. Training the Model:?The goal of training is to answer a question or make a prediction correctly as often as possible
5. Model Evaluating: This includes the selection of the measure as well as the actual evaluation, seemingly a smaller step than others, but important to our end result
6. Parameter Tunning: This step is about improving the performance
7. Making Predictions: The goal is to try to predict on validation set as well as on some single samples - randomly chosen from dataset or custom

MACHINE LEARNING MODEL

A machine learning model is next course of action once data preapration is done. Well-prepared data for your model can improve its efficiency. It can help in reducing the blind spots of the model which translates to greater accuracy of predictions.

Examples: Binary Classification Model ("Is this email spam or not spam?"), Multiclass Classification Model ("Is this product a book, movie, or clothing?"), Regression Model ("What price will this house sell for?").

MODELS OF EVALUATION

Model Evaluation is the process through which the quality of a system?s predictions is quantified . To do this, we measure the newly trained model performance on a new and independent dataset. There are four outcomes that can perform classification predictions.

- **True positives**?occur when the system predicts that an observation belongs to a class and it actually does belong to that class.
- **True negatives**?occur when the system predicts that an observation does not belong to a class and it does not belong to that class.
- **False positives**?occur when you predict an observation belongs to a class when in reality it does not.?
- **False negatives**?occur when you predict an observation does not belong to a class when in fact it does.?

LABORATORIES

Our first task was to prepare work environment. W installed numpy, pandas, scikit-learn, matplotlib and seaborn libaries.

Down below, I imported the Jupyter project and added descriptions and short summaries describing methods and ways of solving problems.

RESULTS

I guess that the outcomes of this little project are satysfying. The model is well performing and accurate enough to put into production. The set contained around 800 people with general information about them. As I presume, the given data was sufficient for the analysis. I don't think that a larger training set would improve a lot my model's preformance.? Also chosing linear discriminant as an estimation model was a good choice. The probability of linear regression was around 80% which is a great result.?

PROBLEMS ENCOUNTERED

I had a few problems on our laboratories with working on dataset, but finally I figured out which columns are necessary and which columns are meant to be change (from 'male'/'female' to 1/0).?

Bibliography:

https://realpython.com/python-matplotlib-guide/

https://www.w3schools.com/python/matplotlib_pyplot.asp

https://docs.microsoft.com/en-us/windows/ai/windows-ml/what-is-a-machine-learning-model

https://blogs.nvidia.com/blog/2021/08/16/what-is-a-machine-learning-model/