

SUMMARY

Why machine learning?

Machine learning overcomes the problem of having to write the rules containing if else statements for a case that may contain a lot of conditions such as identifying spam and non spam emails. Spam emails have features that make it spam and non spam emails also have some features that make it non spam, these features are so many that to distinguish the features by if else statement we might end up writing a ton of codes. Using machine learning we can enable the machine to learn and extract features that will help it to make the decision of whether the email is spam or not.

It also saves time and helps humans analyze and make decision faster and more accurately.

Problems machine learning can solve.

most successful kinds are those that automate decision making process by generalizing from known examples.

If your application can be formulated as a supervised learning problem, and you are able to create a dataset that includes the desired outcome, machine learning will likely be able to solve your problem.

Types of machine learning: -

SUPERVISED LEARNING

Examples

- 1)Identifying the zip code from handwritten digits on an envelope
- 2)Determining whether a tumour is benign on a medical image
- 3)Detecting fraudulent activity in credit card transactions

UNSUPERVISED LEARNING

Examples

- 1)Identifying topics in a set of blog posts
- 2)Segmenting customers into groups with similar preferences
- 3)Detecting abnormal access patterns to a website

Why python? lingua franca

- It is the lingua franca for programmers around the world.
- Python has libraries for data loading, visualization, statistics, natural language processing, image processing, and more. It is important for these processes to have tools that allow quick iteration and easy interaction.
- Ability to interact direct with code, using a terminal or other tools like Jupyter notebook.
- As a general purpose programming language, python also allows for the creation of complex graphical user interfaces(GUI) and web services

Scikit-learn library

- It is an open source project.
- It contains a number of state-of-the-art machine learning algorithms, as well as comprehensive documentation about each algorithm.
- most prominent python library for machine learning.
- It is widely used in industry and academia, and a wealth of tutorials and code snippets available online.

NumPy

- it contains functionality for multidimensional arrays, high-level mathematical functions such as linear algebra operations and the Fourier transform, and pseudorandom number generators.
- It is the fundamental data structure in scikit-learn
- all numpy array elements must be of the same type.

SciPy is a collection of scientific computing in Python.

- It provides among other functionality, advanced linear algebra routines, mathematical function optimization, signal processing, special mathematical functions, and statistical distributions.
- Scikit-learn draws from SciPy's collection of functions for implementing its algorithms.

matplotlib

- Is the primary scientific plotting library in Python.
- It provides functions for making publication-quality visualizations such as line charts, histograms, scatter plots and so on in 2D

pandas

- Is a Python library for data wrangling and analysis.

- It is built around a data structure called the DataFrame that is modeled after the R DataFrame.
- simply pandas DataFrame is a table, similar to an Excel spreadsheet.
- pandas provides a great range of methods to modify and operate on this table.
- allows sql like and joins of tables.
- In contrast to NumPy, each column can have different or separate types.

Python 2.7 vs Python 3

- Python 2 code usually does not run on python 3
- python 3 is the newest python version
- Output may differ slightly in python 2.7 comparing to python 3
- there is some syntax changes in in python 3 compared to python 2

//Logistic regression

//K-nearest neighbor

//supportive-vector

//mean squared error

//accuracy score

//logarithmic loss