Al-Mustaqbal University
College of Sciences
Intelligent Medical System Department

جامعة المستقبل
AL MUSTAQBAL UNIVERSITY

كلية العلوم
قسم الانظمة الطبية الذكية

Introduction to Python for AI and Machine Learning: Setting up the Environment and libraries

Lab: (1)

Subject: Artificial Intelligence
Class: Third
Lecturer: Dr. Maytham N. Meqdad

Introduction to Python for AI and Machine Learning: Setting up the Environment and libraries

Machine Learning with Python

Why Python?

- Many data-heavy applications are now developed in Python
- Highly readable, less complexity, fast prototyping
- Easy to offload number crunching to underlying C/Fortran/…
- Easy to install and import many rich libraries
  - numpy: efficient data structures
  - scipy: fast numerical recipes
  - matplotlib: high-quality graphs
  - scikit-learn: machine learning algorithms
  - tensorflow: neural networks

- Machine Learning is making the computer learn from studying data and statistics.
- Machine Learning is a step into the direction of artificial intelligence (AI).
- Machine Learning is a program that analyses data and learns to predict the outcome.

Data Set

In the mind of a computer, a data set is any collection of data. It can be anything from an array to a complete database.

Example of an array:

```
[99,86,87,88,111,86,103,87,94,78,77,85,86]
```

Example of a database:

Carname	Color	Age	Speed	AutoPass
BMW	red	5	99	Y
Volvo	black	7	86	Y
VW	gray	8	87	N
VW	white	7	88	Y
Ford	white	2	111	Y
VW	white	17	86	Y
Tesla	red	2	103	Y
BMW	black	9	87	Y
Volvo	gray	4	94	N
Ford	white	11	78	N
Toyota	gray	12	77	N
VW	white	9	85	N
Toyota	blue	6	86	Y

By looking at the array, we can guess that the average value is probably around 80 or 90, and we are also able to determine the highest value and the lowest value, but what else can we do?

And by looking at the database we can see that the most popular color is white, and the oldest car is 17 years, but what if we could predict if a car had an AutoPass, just by looking at the other values?

That is what Machine Learning is for! Analyzing data and predicting the outcome!

In Machine Learning it is common to work with very large data sets.

Data Types

To analyze data, it is important to know what type of data we are dealing with.

We can split the data types into three main categories:

- **Numerical**
- **Categorical**
- **Ordinal**

Numerical data are numbers, and can be split into two numerical categories:

- Discrete Data
  - counted data that are limited to integers. Example: The number of cars passing by.
- Continuous Data
  - measured data that can be any number. Example: The price of an item, or the size of an item

Categorical data are values that cannot be measured up against each other. Example: a color value, or any yes/no values.

Ordinal data are like categorical data, but can be measured up against each other. Example: school grades where A is better than B and so on.

By knowing the data type of your data source, you will be able to know what technique to use when analyzing them.

Machine Learning - Mean Median Mode

Mean, Median, and Mode

In Machine Learning (and in mathematics) there are often three values that interests us:

- **Mean** - The average value
- **Median** - The mid point value
- **Mode** - The most common value

Example: We have registered the speed of 13 cars:

speed = [99,86,87,88,111,86,103,87,94,78,77,85,86]

What is the average, the middle, or the most common speed value?

Mean

The mean value is the average value.

To calculate the mean, find the sum of all values, and divide the sum by the number of values:

(99+86+87+88+111+86+103+87+94+78+77+85+86) / 13 = 89.77

The NumPy module has a method for this. Learn about the NumPy module in our NumPy Tutorial.

Example

Use the NumPy `mean()` method to find the average speed:

```
import numpy

speed = [99,86,87,88,111,86,103,87,94,78,77,85,86]

x = numpy.mean(speed)

print(x)
```

Median

The median value is the value in the middle, after you have sorted all the values:

77, 78, 85, 86, 86, 86, 87, 87, 88, 94, 99, 103, 111

It is important that the numbers are sorted before you can find the median.

The NumPy module has a method for this:

Example

Use the NumPy `median()` method to find the middle value:

```
import numpy

speed = [99,86,87,88,111,86,103,87,94,78,77,85,86]

x = numpy.median(speed)

print(x)
```

If there are two numbers in the middle, divide the sum of those numbers by two.

`77, 78, 85, 86, 86, 86, 87, 87, 94, 98, 99, 103`

`(86 + 87) / 2 = 86.5`

Example

Using the NumPy module:

```
import numpy

speed = [99,86,87,88,86,103,87,94,78,77,85,86]

x = numpy.median(speed)

print(x)
```

Mode

The Mode value is the value that appears the most number of times:

`99, 86, 87, 88, 111, 86, 103, 87, 94, 78, 77, 85, 86 = 86`

The SciPy module has a method for this. Learn about the SciPy module in our SciPy Tutorial.

Example

Use the SciPy `mode()` method to find the number that appears the most:

```
from scipy import stats

speed = [99,86,87,88,111,86,103,87,94,78,77,85,86]

x = stats.mode(speed)

print(x)
```

==========