Tutorial 5: Machine Learning I
An Introduction to Machine Learning

COMP-SCI 490-0025
Research Skills Training (RST) in Distributed AI

Prof. Yusuf Sarwar Uddin (RST Instructor) & Vijaya Yeruva (Graduate Teaching Asst.)
University of Missouri - Kansas City, Spring 2023

On-track completion due date: March 10, 2023

Purpose

This tutorial aims to familiarize students with Machine Learning (ML), Types ML, Liner Regression technique, and its evaluation metrics using Python. This tutorial builds on Python skills from previous Tutorials.

This tutorial is divided into three sections:

- Introduction to Machine Learning (ML)
- Types of Machine Learning (ML)
- Liner Regression technique and its evaluation metrics

Note: This tutorial is designed to be completed on the UMKC SCE computing labs, and students who elect to use their computer would need to install appropriate software such as Python and Anaconda. Please refer to Tutorial 2 for installation instructions.

Outline

Section	Topic	Outcome
1	Introduction to Machine Learning (ML)	Learn what Machine Learning (ML) is and Where to use it
2	Types of Machine Learning (ML)	Learn about three types of ML algorithms
3	Regression technique and its evaluation metrics	Learn and apply Liner Regression using the “scikit-learn” library

Expected Deliverable & Assessment

Please submit a **tutorial completion document** as a **single PDF** file to Canvas Assignment by the tutorial due date. Your goal with this document is to demonstrate your level of competency and understanding for all competency-building **activities** in this tutorial. These will be found within *some* of the TASKS in the tutorial. These types of TASKS should have their titled page (or pages) in which you illustrate (with screenshots, etc.) and summarize (with bullets and brief statements) your work and answers. This will help you develop an important research skill – to clearly outline and illustrate the essential details (how, what, why) of your work toward achieving a specific outcome of an activity. In short, strive to be clear, and concise.

Additionally, other TASKS will ask you to simply practice the steps (*learn-by-doing!* without the need for you to report on your practice. Simply execute the practice steps to be prepared for competency-building **activities**.

You can use Microsoft Word, MacOS Keynote (available on UMKC SCE computing labs), or any other professional presentation software to make your PDFs.

Competency-based Assessment and Feedback

- What: your tutorial completion document will be assessed following the criteria in the Grading Rubric in your syllabus. On canvas, you will receive a tutorial grade and brief but actionable feedback for all items assessed as less than Excellent.
- When: submissions meeting the on-track deadline will be assessed in a timely fashion by the course TA in consultation with the professor. Typically, this should be within a week. Late submissions may take longer to assess.

Collaboration with peers is allowed. Please make sure you understand the collaboration and student conduct policies provided in the syllabus before choosing to collaborate on your assignment. Students **must** include the names of collaborating peers in relevant parts of their **tutorial completion documents**.

Note: Because we primarily focus on coding, there will be no option to revise and resubmit your work from Tutorial 3.

1 Introduction to Machine Learning (ML)

This section will mainly focus on What is Machine Learning? and What is it used for?

1.1 What is Machine Learning?

Machine learning is a field of computer science concerned with programs that learn.

> The field of machine learning is concerned with the question of how to construct computer programs that automatically improve with experience. — Machine Learning, 1997.

That is super broad. The below picture 1 can give a simple idea of how ML differs from traditional programming.

Figure 1: Basic Difference in ML and Traditional Programming

Unlike previous tutorials that followed this PDF instruction document, we will use the pre-filled Jupyter notebook to follow along and complete the activities and tasks in this tutorial.

First, download the **"data"** folder and **"Tutorial5.ipynb"** notebook from Canvas. Open the "Tutorial5.ipynb" notebook with Anaconda (Start Anaconda Jupyter Notebook as explained in Tutorial 2 and open this file).

The "Tutorial5.ipynb" notebook contains detailed instructions for all activities and tasks related to machine learning (ML). Please include a summary of your findings in the **Tutorial Completion Document**. Refer to Figure 2.

1.2 What is it used for?

Fraud detection	Prediction of equipment failures
Web search results	New pricing models
Real-time ads on web pages	Network intrusion detection
Credit scoring	Text Sentiment Analysis
Pattern and image recognition	and more...

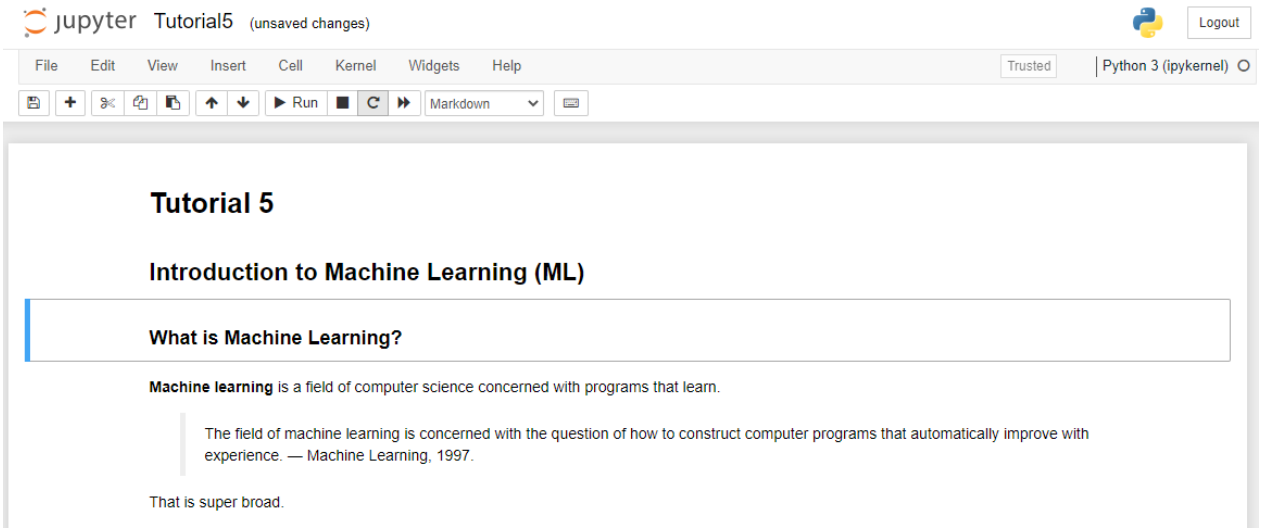

Figure 2: Introduction to Machine Learning (ML) in the Jupyter Notebook

2 Types of Machine Learning (ML)

There are different ways an algorithm can model a problem based on its interaction with the experience, environment, or whatever we want to call the input data.

There are three different learning styles in machine learning algorithms

1. Supervised Learning
2. Unsupervised Learning
3. Semi-Supervised Learning

The "Tutorial5.ipynb" notebook contains detailed instructions for all activities and tasks related to Types of Machine Learning (ML). Please include a summary of your findings in the **Tutorial Completion Document**. Refer to Figure 3.

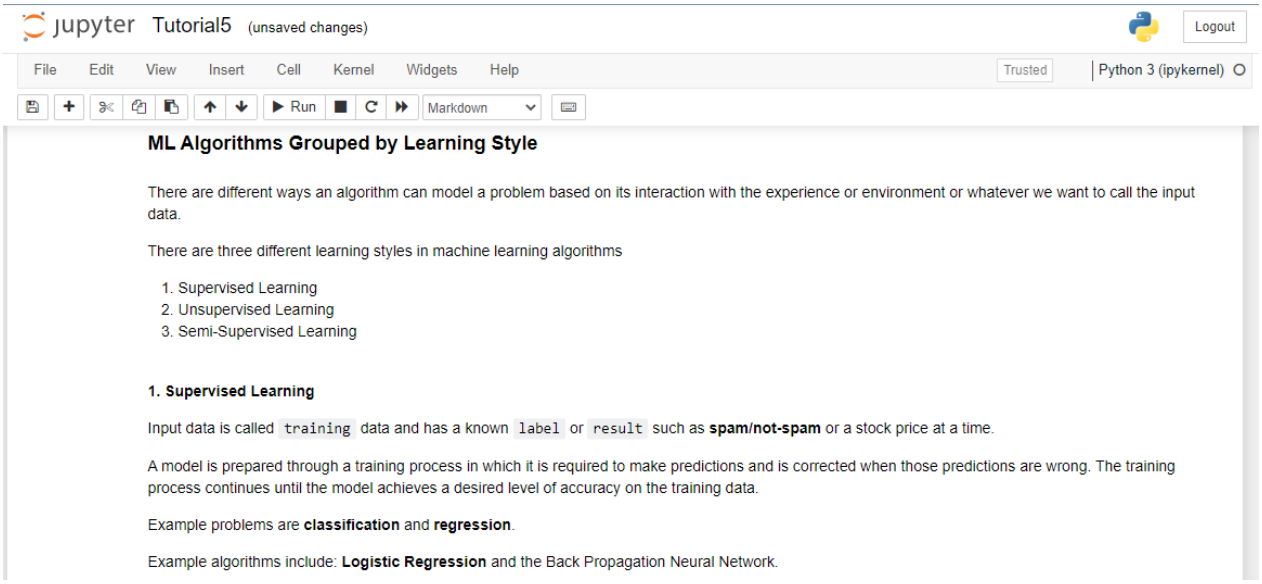

Figure 3: Types of Machine Learning (ML) content in the Jupyter Notebook

3 Regression

Predictive modeling techniques such as **regression** analysis may be used to determine the relationship between a dataset's dependent (goal) and independent variables. It is widely used when the dependent and independent variables are linked in a linear or non-linear fashion, and the target variable has a set of continuous values.

Regression analysis is used for one of two purposes: predicting the value of the dependent variable when information about the independent variables is known or predicting the effect of an independent variable on the dependent variable.

The "Tutorial5.ipynb" notebook contains detailed instructions for all activities and tasks related to Linear Regression. Please include a summary of your findings in the **Tutorial Completion Document**. Refer to Figure 4.

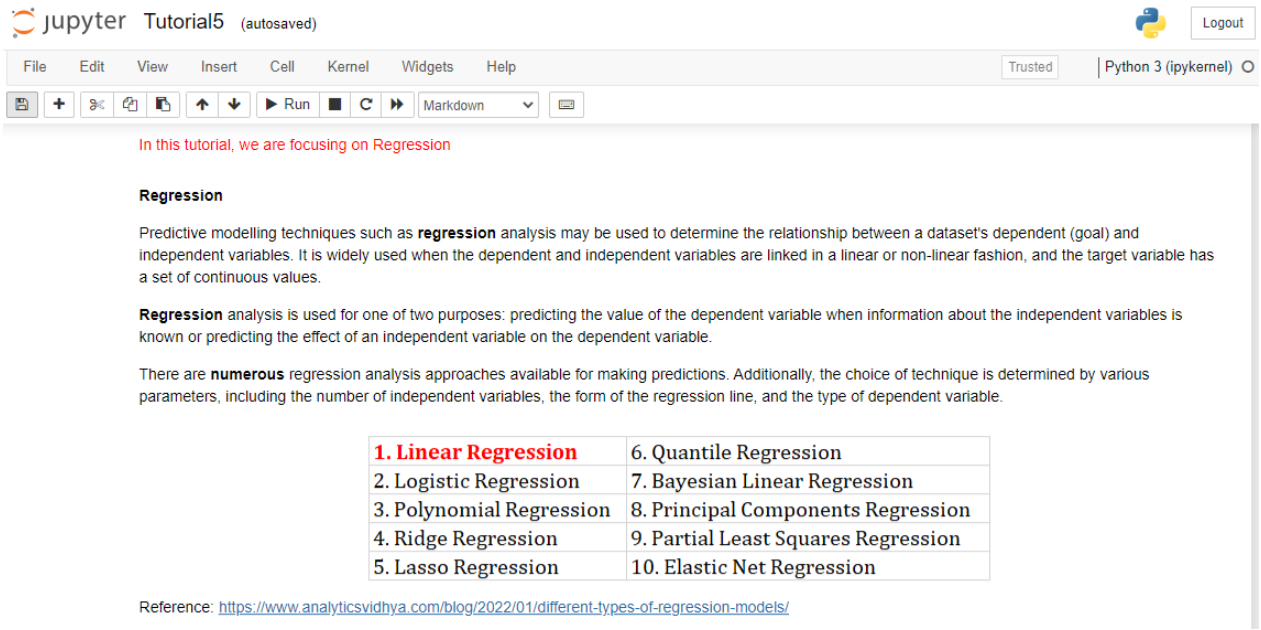

Figure 4: Regression content in the Jupyter Notebook

4 References for Machine Learning

1. https://scikit-learn.org/stable/
2. https://machinelearningmastery.com/
3. https://pieriantraining.com/7-machine-learning-regression-algorithms-p
4. https://www.analyticsvidhya.com/blog/2022/12/analytics-vidhyas-top-10-