Deloitte.

Tech Treks

ROI TRAINING

MAXIMIZE YOUR TRAINING INVESTMENT

Machine Learning with Python

Welcome!

- ROI leads the industry in designing and delivering customized technology and management training solutions
- Meet your instructor
  - Name
  - Background
  - Contact info
- Let’s get started!

Session Objectives

In this session, you will:

- Learn the fundamentals of Machine Learning (ML)
- Use scikit-learn to do Cluster, Regression, and Classification Analysis

Session Concepts

Machine Learning Basics

Cluster

Regression

Classification

What Is Machine Learning?

- The idea behind Machine Learning (ML) and ultimately Artificial Intelligence (AI) is quite simple
  - There are patterns in data
  - Determining those patterns can give us better insights into what data means
  - If a reliable pattern can be found, we can use it to make predictions of currently unknown values
- All we need is the right tools
  - Statistics is the mathematical basis for finding these patterns
  - ML Packages encode the math into computer algorithms we can use to find these patterns

How to Do Machine Learning

- We start with data, lots of it
  - We clean it up, and reshape it to keep the parts that are relevant; these are known as **features**
  - We identify the value we are trying to predict; this is called the **label**
- Based on the data we have and the business case we want to make with it, we identify the types of algorithms that can get us there
- We run the data through a process known as **training** which produces a **model**
- The model can then be used to make **predictions**

Python and ML

- ML can be done with any programming language
- Traditionally, **R** was the most popular, but Python is by far the most popular now
- To do ML with Python, you just need to install some additional packages
  - NumPy and pandas are the two main packages for storing data
  - Matplotlib and seaborn are two popular packages for plotting
  - scikit-learn is the first package most people use to get started with
  - TensorFlow, PyTorch, and Keras are more modern and feature rich

Session Concepts

Machine Learning Basics

Cluster

Regression

Classification

Cluster Analysis

- Analysis technique to help make sense of the data before feeding it into other models
- Unsupervised
  - More about discovering patterns in data
  - Not about predicting values for unknown values
- Looks for natural groupings among the data
  - Voter groups (is it just left vs. right, or left, right, or center)
  - Species identification (are two groups of organisms different enough to be considered a different species or not)
  - Identify different types of customers we may have
- Often helpful as a preparatory step before classification to determine how many categories we may want to predict

Session Concepts

Machine Learning Basics

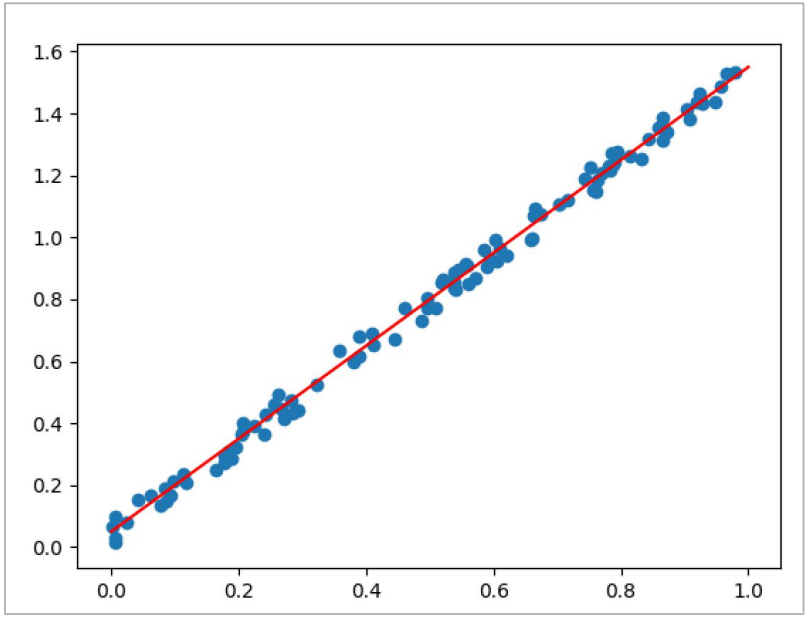
Cluster

Regression

Classification

Regression Analysis

- Given a collection of X, Y points, you could easily see there is a pattern
- If you remember enough algebra, you could describe the pattern of dots as roughly following the red line, which could be described with the formula y = 1.5x + .01



Regression Analysis (continued)

- Whenever we want to try to predict a numeric value, we can use regression
  - Predict a temperature
  - Price of a commodity
  - Time to failure of a part
- Regression finds the pattern among the features set and how it influences the target value we are trying to predict

Session Concepts

Machine Learning Basics

Cluster

Regression

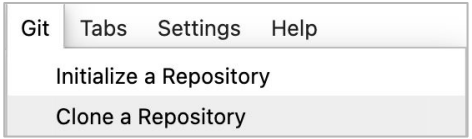
Classification

Classification Analysis

- There are many business use cases for wanting to predict whether a record will fall into one category or another
  - Is a credit card swipe fraudulent or not?
  - Does a patient have a disease?
  - Will an applicant be a profitable customer?
- Classification can make such predictions by using historical data to train the model to look for patterns
- To see how good a job the model does, you test it with another set of data that was not used to train the model
  - By comparing the known values to the predicted ones, you can judge how well the model performs

Jupyter Notebook

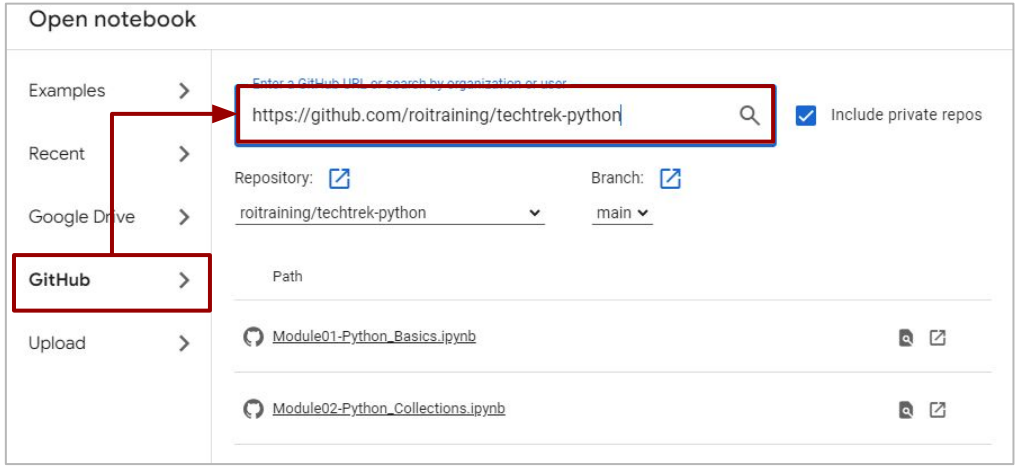
- Jupyter Notebook is perhaps the easiest way to get started with Python
- Let's start a Jupyter session by navigating to: https://notebooks.roitraining.com/
- Select the link nearest to your location
- From the top menu, choose **Git | Clone a Repository** and enter: **`https://github.com/roitraining/techtrek-python`**
  - Sometimes, you may need to add `.git` to the end
- Click the **`Module04-Machine_Learning.ipynb`**



Note: The Notebook server will only be available during class and one hour afterward

After Class

- To access the Notebook server after class, go to: https://colab.research.google.com/
- Select **GitHub** and enter **https://github.com/roitraining/techtrek-python**
- You must have a Google-compatible email account to log in



Session Summary

In this session, you have:

- Learned the fundamentals of Machine Learning
- Used scikit-learn to do Cluster, Regression, and Classification Analysis

Discussion: Recap