Norwalk Campus

Division of Extended Studies

Python: Machine Learning and Data Science – Spring 2025

Instructor: Nadia Udler **Contact:** nudler@ncc.commnet.edu

Location: West Campus NCC **Room:** W137

Days: Saturdays **Time:** 9:00am – 12:00pm

Required Books and Video Lectures

Charles Severance, Python for informatics (free, available online)

https://www.py4e.com

Other useful books (not required!):

William Punch, Richard Enbody, The Practice of Computing Using Python, 2017, 3rd Edition, Pearson

Wes McKinney, Python for Data Analysis, 2022, O'Reilly

Deisenroth, Faisal, Ong, Mathematics for machine learning, 2020, Cambridge University Press

Software:

Anaconda/Python installation:

https://www.anaconda.com/products/distribution

Grading:

Homework: 30%

Project: 40%

Participation: 30%

Attendance:

'Incomplete' grade is given if two or more classes are missed

Course Outline:

Day 1: Working environment, basic data types and control flow structures

Installing Python and Anaconda package, Spyder development environment, writing and running a program. Debugging. Basic data types (strings, numerical types, Boolean type) and basic control flow structures (assignments, conditionals, loops), Input and output.

Day 2: Data structures, functions and modules, introduction to Object Oriented Programming

Tuples, lists, dictionaries, sets, operations on sequential data types, operations on mapping data types, exceptions, recursion, assertions, list comprehension

Functions, module, scope of variables, arguments, parameters and namespace, function annotations, docstrings, functions as parameters

Classes and Objects in Python, defining your own data types.

Day 3: Libraries: NumPy, Pandas, SQLite, Matplotlib

Numpy arrays, Pandas Series and DataFrame, data manipulation with Pandas, reading Excel and text files, database connection with SQLite, working with binary and text data, data visualization

Day 4: Essential machine learning mathematics. Libraries: SciPy, Statsmodels, Scikit-learn

Mathematical modeling, linear models, solving system of linear equations, classification and regression, perceptron algorithm

Day 5: Using machine learning model implementations from Scikit learn and Tensorflow, estimating quality of ML model, optimization methods for machine learning, reinforcement learning

Linear regression, logistic regression, Feed forward neural networks, Support vector machines

Day 6: Scikit-learn: working with different data types, numerical and categorical data, text data models

Data scaling, one-hot encoding of categorical data, bag of words model, topic modeling

Day 7: SHAP library: explain machine learning predictions by estimating feature contributions

Estimating feature contributions for Logistic Regression and Neural Network

Day 8: Advanced machine learning methods and topics for individual projects:

Unsupervised learning methods, dimension reduction, data visualization, different neural network configurations, Bayesian methods in ML, decision trees