SYLLABUS

THE COMPLETE DATA SCIENCE & MACHINE LEARNING BOOTCAMP

APP BREWERY

www.appbrewery.co

INTRODUCTION TO LINEAR REGRESSION AND THE DATA SCIENCE WORKFLOW

- Learn how to specify a Data Science problem.
- Understand how and why you need to clean your data.
- Learn how to incorporate Python modules such as **Pandas** and **Matplotlib** into your Jupyter Notebook.
- Use Matplotlib visualise and better understand your data.
- Learn about the theory behind **Linear Regression** and how it works.
- Estimate and interpret regression coefficients using **scikit-learn**.
- Make a prediction using your model.
- Analyse and evaluate your regression's results using metrics such as the goodness of fit or **R-squared**.

INTRODUCTION TO PYTHON PROGRAMMING

- Understand how to use variables and types to manage data.
- Work with Python collections: **Lists**, **Numpy Arrays**, **Pandas DataFrames**, and **Series**.
- Understand how to use Python **modules** and import Python packages.
- Learn how to use Python **functions** to simplify complex operations.
- Understand function **arguments**, **parameters**, and **return** values.
- Understand how to work with Python **objects**.
- Familiarise yourself with Python naming conventions.
- Learn to how to use the Python **documentation** to solve your own problems.

OPTIMISATION AND GRADIENT DESCENT

- Understand how optimisation works in practice and how parameters in a Machine Learning model are estimated.
- Understand the role of **Cost Functions**.
- Introduction to calculus: **derivatives**, the **power rule**, and **partial derivatives**.
- Understand how to work with Python **Tuples**.
- Work with Python loops to run the **Gradient Descent** optimisation algorithm.
- Understand the effect of the learning rate, multiple minima, and the pitfalls with optimisation algorithms.
- Learn how to manipulate, reshape, concatenate, and transpose data in **N-Dimensional arrays**.
- Learn how to create **3D plots** and **charts**.
- Understand the **Mean Squared Error** cost function.
- Work with a **nested loop** in Python.

SYLLABUS

THE COMPLETE DATA SCIENCE & MACHINE LEARNING BOOTCAMP

APP BREWERY

www.appbrewery.co

MULTIVARIABLE LINEAR REGRESSION

- Explore a dataset by examining summary statistics, finding missing values, and discovering outliers.
- Create powerful visualisations using **Matplotlib** and **Seaborn** to examine how data is distributed and see the relationships present in the data.
- Understand how to work with index and dummy variables in datasets.
- Understand how to interpret and measure correlations in the data.
- Learn how to diagnose and address problems like **Multicollinearity** in Linear Regression.
- Learn how to transform your data to improve your model.
- Evaluate your model's performance and learn how to choose among different regression models using the **Baysian Information Criterion**.
- Learn how to compare the actual and the predicted values of your regression and what you can learn from the **residuals**.
- Understand **control flow** and **logical operators** in Python.
- Understand docstrings and optional arguments in Python functions.

CLASSIFICATION PROBLEMS, PROBABILITY AND DATA PRE-PROCESSING IN PYTHON

- Understand the intuition behind the **Naive Bayes Classification Algorithm**.
- Understand basic, joint- and conditional probability as well as **Bayes Theorem**.
- Understand the role of **Paths** and how to access external resources from your notebook.
- Understand how to use **generator functions** in Python to work with large datasets.
- Clean a dirty dataset and learn how to process text data.
- Visualise your data using Pie and Donut charts as well as Word Clouds.
- Introduction to **natural language processing** to convert, stem, and tokenise text data.
- Learn how to install third party packages and **dependencies**.
- Learn how to check for membership in a Python collection.

TRAIN A NAIVE BAYES MODEL TO CLASSIFY SPAM EMAILS

- Learn how to save and load files external to your Jupyter notebook
- Calculate the spammyness of individual words in emails.
- Build a spam classifier from scratch using **probability theory**, and **Laplace smoothing techniques**.

SYLLABUS

THE COMPLETE DATA SCIENCE & MACHINE LEARNING BOOTCAMP

APP BREWERY

www.appbrewery.co

TEST AND EVALUATE A CLASSIFICATION MODEL

- Understand how to calculate the **joint conditional probability** using the dot product.
- Understand how to calculate the accuracy of your classification algorithm.
- Understand and visualise the **decision boundary** of your classification algorithm.
- Learn techniques for tackling overplotting in data visualisation.
- Understand the role of **false positives** and **false negatives** in classification problems.
- Learn how to calculate the **recall**, **precision**, and **f-scores** of a model.
- Use **sci-kit learn** to implement a Naive Bayes classifier very quickly.

USE PRE-TRAINED DEEP LEARNING MODELS FOR IMAGE CLASSIFICATION

- Understand the theory behind **artificial neural networks**.
- Introduction to **nodes**, **layers**, **activation functions** and **back-propagation** in a neural network.
- Understand the pros and cons of neural networks.
- Introduction to **Tensorflow** and **Keras**.
- Learn how you can incorporate pre-trained models from Model Zoos into your own projects.
- Understand how to work with images and process image data.
- Learn how to make predictions with a pre-trained model.

BUILD YOUR OWN NEURAL NETWORK WITH KERAS

- Learn how to pre-process image data for training a neural network.
- Learn how and why you should create a **validation dataset** for training your model.
- Understand how activation functions work.
- Understand the **cross-entropy loss function** for **multi-class classification**.
- Learn about the role of weights and bias terms in a neural network.
- Understand error handling in Python with **Try-Catch clauses**.
- Learn how to visualise the learning process and spot problems using **Tensorboard**.
- Understand how to break up and batch your training data.
- Use regularisation techniques such as **early stopping** and **dropout** to combat overfitting.
- Use your Keras model to make a prediction on an image.
- Evaluate your model using a **confusion matrix**.
- Use Python's **Itertools** for efficient looping.

SYLLABUS

THE COMPLETE DATA SCIENCE & MACHINE LEARNING BOOTCAMP

USE TENSORFLOW TO CLASSIFY HANDWRITTEN DIGITS

- Learn to use **one-hot encoding** to process data for classification-style problems.
- Understand what a **tensor** is.
- Learn how to set up the Neural Network Architecture in Tensorflow.
- Define **Tensorflow operations** for your loss function, optimiser and evaluation metrics.
- Understand the **Tensorflow graph** and how to use **Context Managers** in Python.
- Understand how to work with **Tensorflow sessions** to train your model and run calculations.
- Learn to use advanced features in Tensorboard to evaluate your model and monitor training over time.

APP BREWERY

www.appbrewery.co

If you click on the "Course Resources" blue button on the top left corner of this lesson, you can download the entire curriculum as a pdf. You can use this list to check off the topics you're confident about and also mark the ones that you might want to review. Other students use the syllabus to plan their learning and schedule their time around the topics. These are just my suggestions, you're the expert on the best way you learn!

All the best,

Your instructor, Philipp