Yeah Education Introduction to Data Science

Syllabus

Spring 2017

Welcome to Yeah Education’s Introduction to Data Science course! Data Science is one of the hottest buzzwords to come out of the tech industry and certainly a versatile skill to have. Data Scientists go into data analyst, business analyst, product analyst, quantitative finance, financial engineering, software engineering, machine learning, AI, actuary, and biostatistician job positions and much more. At its core, data science is a mix of computer science and statistics – leveraging the power of programming with the theory of how data should behave. In this course, I will be giving you a high-level introduction to how to analyze data from a data scientist perspective. The course is designed to be thorough but practical – everything you will learn is used in industry in one way or another.

Learning objectives:

You will gain proficiency in several topics, including but not limited to the following list:

1. Python programming

2. Data interpretation, data munging, and visual analysis

3. Numerical and Text

4. Linear Regression

Course Logistics:

There will be four 3-hour course sessions on Saturdays 9:30am to 12:30pm. The week-by-week schedule can be found below:

4/08 - Python setup. Coding environment setup. Jupyter notebooks. Variables, objects, loops, NumPy.

4/15 – Pandas. Hands-on data exploration. Plots. Summary statistics. Data cleaning. Introduction to regular expressions.

4/22 - Linear Regression. Assumptions for intuition. Interpretation.

4/29 – Review and Final Exam.

Each session will have the following format:

4/08

9:30-11:00am: Course introduction. Python setup. Basic programming.

*10 minute break*

11:10-12:00am: Basic programming. The NumPy library.

12:00-12:20pm: Programming quiz.

12:20-12:30pm: Homework 1 and Reading 1 assigned.

4/15

9:30-9:50am: Homework 1 (due) and Reading 1 (due) quiz.

10:00-11:00am: Pandas data reading, munging, statistics, and visualizations.

*10 minute break*

11:10-12:00am: Analyzing text data. Introduction to regular expressions.

12:00-12:20pm: Data analysis quiz.

12:20-12:30pm: Homework 2 and Reading 2 assigned.

4/22

9:30-9:50am: Homework 2 (due) and Reading 2 (due) quiz.

10:00-11:00am: Linear regression theory

*10 minute break*

11:10-12:00am: Linear regression handwritten and coding practice

12:00-12:20pm: Linear regression quiz.

12:20-12:30pm: Homework 3 and Reading 3 assigned.

4/29

9:30-9:50am: Homework 3 (due) and Reading 3 (due) quiz.

10:00-11:00am: Review

*10 minute break*

11:10-12:20pm: Final Exam.

12:20-12:30pm: Final words

Readings Schedule:

Note: Reading page numbers may differ from online textbook versions.

4/08

Python for Data Analysis Chapter 4 pages 75-90, 94-99, 101-102

4/15

Python for Data Analysis Chapter 5 pages 107-118, 121-124, 133-143

4/22

An Introduction to Statistical Learning Chapter 3 pages 59-68, 71-74, 82-87

Homework Schedule:

4/08

Programming and NumPy Exercises due 4/15

4/15

Data manipulation in Pandas Exercises due 4/22

4/22

Linear regression with Sklearn Exercises due 4/29