Fundamentals of Mathematics for Data Science

Heidi Perry, PhD

Introduction

Learn the fundamentals of data preparation and analysis needed to transform data into insight. This course will introduce a broad range of mathematics fundamentals used in data science and machine learning, including probability distributions, statistics, regression, linear algebra, and calculus, with a focus on practical application to real-world data sets. Bring your own data set or contribute to a Hack Oregon project; students will apply methods to a project of their choosing.

Prerequisites: familiarity with the Python Programming Language and proficiency in basic algebra.

Course Format

Evening lectures will include IPython notebooks for interactive exploration of the weeks' topic. At Sunday workshops, students have a chance to work together on weekly assignments and receive feedback from the instructor.

Students will choose a data set of interest to analyze, or use data sets from Hack Oregon volunteer projects. Throughout the course the student will perform exploratory analysis on the data set, choose a research question, and present the results to receive peer feedback during the final class.

Weekly Lecture Topics

This course will introduce a very broad spectrum of topics. The goal of the course is to familiarize students with terminology and methods used in data science.

1. Introduction to Data and the Scientific Method
    - The "science" in data science
    - Introduction to data
    - Creating, reading, and describing graphs
    - Identifying trends

2. Math fundamentals
    - Terminology and notation
    - Common functions
    - A brief introduction to calculus
3. Cleaning and transforming data
    - Handling missing data
    - Transformations to identify trends
4. Linear algebra
    - Introduction to matrices and vectors
    - Operations: addition, multiplication, transpose
    - Inverting matrices
5. Probability distributions and stochastic processes
    - Common probability distributions
    - Descriptive statistics
    - Expected value, variance
    - Independent events
    - Correlation
    - Conditional probability
    - Central limit theorem
    - Bayes's theorem
6. Statistical inference
    - Hypothesis testing and p-values
    - Sampling
    - Confidence intervals
7. Linear regression
    - Fitting a line to data
    - Residuals
    - Evaluating error
    - Interpretation
8. Present projects
    - Give and receive peer evaluation
    - Course wrap-up