CS410: Principles and Techniques of Data Science

Module 1: Introduction to Data Science

What is Data Science?

Make sense of the data by creating insights that every human can benefit from

Example – Sigma Technologies

Ben Runkle, CEO, Sigma Technologies, is trying to resolve a huge problem.

The company is consistently losing long-time customers.

He does not know why they are leaving, but he must do something fast. May be create new products and features.

His chief data scientist, Dr. Jessie Hughan is not convinced that new products and features alone will save the company. She turns to the transcripts of recent customer service tickets and finds something surprising:

- "…. Not sure how to export this; are you?"
- "Where is the button that makes a new list?"
- "Wait, do you even know where the slider is?"
- "If I can't figure this out today, it's a real problem..."

It is clear that customers were having problems with the existing UI/UX, and weren't upset due to a lack of features. Runkle and Hughan organized a mass UI/UX overhaul and their sales have increased.

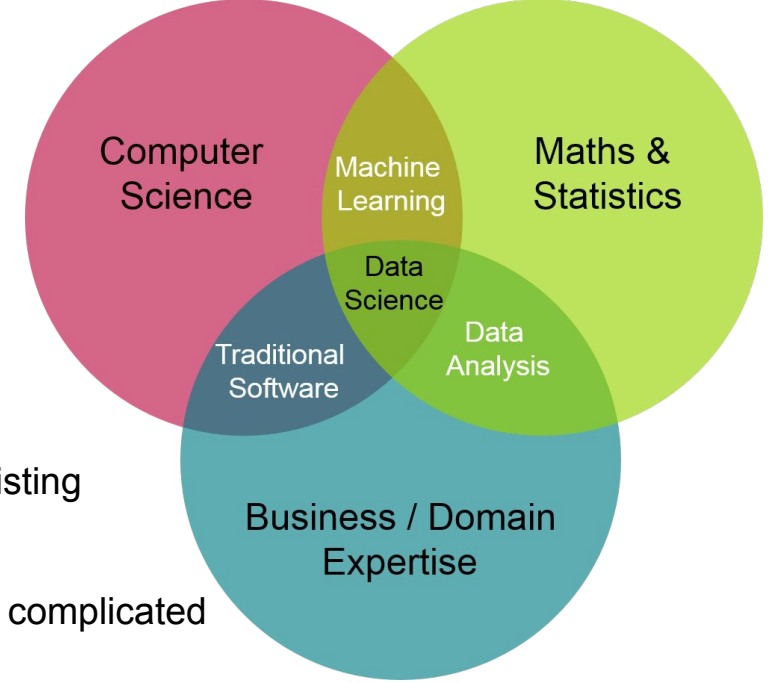
Data Science

Data Science is about drawing useful conclusions from large and diverse data sets through exploration, prediction, and inference.

- **Exploration** involves identifying patterns in information.
- **Prediction** involves using information we know to make informed guesses.
- **Inference** involves quantifying our degree of certainty: will the patterns that we found in our data also appear in new observations? How accurate are our predictions?

Our primary tools for exploration are visualizations and descriptive statistics,
for prediction are machine learning and optimization, and
for inference are statistical tests and models.

Data Science Venn Diagram



Data Science is the intersection of the three key areas:

Math/statistics: Theorize and evaluate algorithms and tweak the existing procedures to fit specific situations

Computer programming: Allows you to conceptualize and program complicated algorithms using computer languages

Domain knowledge: Understanding the problem domain (medicine, finance, social science, and so on)

Having **domain expertise** allows you to apply concepts and results in a meaningful and effective way. If a financial analyst started analyzing data about heart attacks, they might need the help of a cardiologist to make sense of a lot of the numbers.

Why Data Science?

- **Sky-high Salaries**
- **Growing Demand**
- **Plenty of Work Options:** can work in different industries, ranging from sales/marketing to healthcare.
- **Challenging Work:** combination of various disciplines, such as statistics, computer programming, and mathematics that demands you to continuously learn new skills.
- **Evolving Field:** Many exciting new fields are emerging such as Artificial Intelligence (AI), Machine Learning (ML), Big Data.
- **Data Science is the Career of Tomorrow:** Data holds the potential to develop better products.

Case Study: Plotting the Classics

https://colab.research.google.com/drive/16B7aVn3iIcQvrWVGU3ZfiSx3PKfLj30t?usp=sharing

THANK YOU!