Lecture 02

- Machine learning: historical data only. Majority of AI that exists is you take historical data, you have a model, extract insights, then generate an output.
- Input → Output
    - The arrow is the AI. Data in the input and output is the prediction
    - Ex: email → model → spam = spam filter
    - Ex: english → model → spanish = speech translation
- Why do we have this explosion?
    - These algorithms have been around since the late 50s-60s. Recently, Google invented big data (Hadoop). They realized that they wanted to archive the entire internet. They wanted to shard the data instead of having one warehouse (data is in many warehouses).
    - We now have around 44 zettabytes of data.
    - You give a neural network enough data it can identify all kinds of patterns.
        - The more data you give the neural network the better it becomes.
- Google tag manager
- Managing data is hard and an underrated skill is being able to clean the data.
- Mechanical turk is using humans to make a process better
- A lot of the time data is being misused and there is a bunch of hype around it.
    - Tons of data does not correlate to valuable data
- Data is messy
    - Missing values
    - Incorrect labels
    - Multiple types of data. Unstructured = images, audio, text. Structured data = excel sheets
- Difference between machine learning and data science
    - Machine learning engineers are more of the tech and programming (software). But, these engineers do not have deep industry experience.
    - Data scientists ideate the solutions. They find insights in the data and see how it applies (powerpoint presentation).
- Deep learning is a type of AI that models the human brain. Neurons identify a different pattern in the data.
- What makes a company good at AI?