Introduction to Machine Learning Notes

High Level Mathematical Definition

- Machine Learning **encapsulates a function** to calculate an output based on known input values (can be more than one as represented by f(x, y, z) for instance.
    - Attempting to define that function to calculate this output is **training**.
- After this function has been defined for the data set, you can use it to predict new values, which is known as **inferencing**.

High-Level Understanding of Training

1. Training data consists of using past observations or known values, called **features.**
    a. The output is denoted by a term called **label**.
    b. Example taken from **learn.microsoft.com**:
        i. "In the ice cream sales scenario, our goal is to train a model that can predict the number of ice cream sales based on the weather. The weather measurements for the day (**temperature, rainfall, windspeed,** and so on) would be the **features (x)**, and the number of ice creams sold on each day would be the **label (y)**."
2. An algorithm is then applied to the data set to determine some kind of relationship between the features and the label to generalize a relationship (exactly a mathematical concept).
    a. The basic principle is to try and fit a function to the data.
3. Finally, the result of the algorithm is a **machine learning model** that summarizes the calculation derived by the algorithm as **a function**.
    a. In mathematical terms, it's simple represented by **y = f(x)**.
4. Now that the training is finished, the model can be used for **inferencing** to predict a corresponding label when you input features. Note that the output is denoted as **ŷ** and not y because they are predicted labels by the model, not the actual label.

Types of Machine Learning (Basics and High-Level Overview)

Supervised Machine Learning

- A general term for machine learning algorithms that use both features and labels from past data to predict future data. It determines a relationship

Linear Regression

- One such example is a **linear regression model (label predicted by the model is numerical)**.
    - For instance, the number of ice creams sold on a day based on temperature, rainfall, etc.

Classification

- The labels in a classification machine learning model are categorical variables. There are two main classification model types.
- **Binary Classification**
    - Simply determines whether the observed value belongs or doesn't belong to a certain group.
    - Predicts true/false scenarios.
- **Multiclass Classification**
    - Predicts a label to fit into one of **many different groups**, not just two. This successfully extends the binary classification to many groups.
        - For instance, classifying a species of something (several different species groups).
    - Predicts **mutually exclusive labels** (one entity cannot be part of two groups simultaneously).

Unsupervised Machine Learning

- Involves training machine learning models using data that has only feature values without known labels (no sample outputs of the data to work with).

Clustering

- This is the most common form of unsupervised machine learning which basically finds similar features between data sets.
- This is very similar to multiclass classification from supervised machine learning, but we don't have any known labels to work with.
    - For example, grouping flowers by number of petals, etc.

Links

https://learn.microsoft.com/en-us/training/modules/fundamentals-machine-learning/3-types-of-machine-learning