Notes of Machine Learning

In a dataset, labels and features are two fundamental components that are used in supervised machine learning. They play distinct roles in training and building predictive models. Here's the key difference between them:

Features

- Features are the input variables or attributes that describe the characteristics of each data point in the dataset.
- Features are used to represent the data and provide information about the input space. They are the independent variables that are used to make predictions or classifications.
- Features can be numeric (e.g., age, temperature), categorical (e.g., color, category), or even more complex representations like text or images.
- Features are the aspects of the data that the model uses to learn patterns and relationships.

Labels

- Labels are the output variable or target variable in a supervised learning dataset. Labels represent the ground truth or the desired outcome that the model is supposed to predict or classify.
- For a classification task, labels are typically categorical (e.g., class labels such as "spam" or "not spam").
- For a regression task, labels are typically continuous (e.g., predicting a numerical value like house prices).
- Labels are used to train the machine learning model by providing it with examples of input data (features) along with their corresponding correct output (labels).
- The model's primary goal is to learn a mapping from features to labels, so it can make predictions on new, unseen data.