

# Lecture 1: What is Machine Learning

---

When a computer program uses its **experience E** to do **task T** and then depending on its **performance P**, improve its ability to do task T again.

## Supervised Learning

---

When the computer program is given a dataset **with labeled data** and it is then programmed to find the right answer from its experience with the features of the problem.

An example, an email program is given a labelled dataset of spam and not spam emails and it is then programmed with predetermined features to detect whether an email is spam or not.

## Unsupervised Learning

---

When the computer program is given a dataset **without any labeled data** and it is programmed to find structures/patterns/"potential-features" among the data in that dataset.

An example, an email program has been given a unlabeled dataset of emails. It is then programmed to find similarities between those emails and categorize/separate them into groups or clusters. (Algorithm that group data into clusters is called Clustering Algorithm).

## Classification vs. Regression

---

### Classification

---

When there are only two outputs (like yes or no, spam or not, healthy or not) or the output can be categorized (according to the authors or researchers) into two groups, the problem is said to be addressed with Classification.

### Regression

---

When there are a spectrum of outputs (like cost of houses per sqft, ) The problem is said to be addressed through regression.