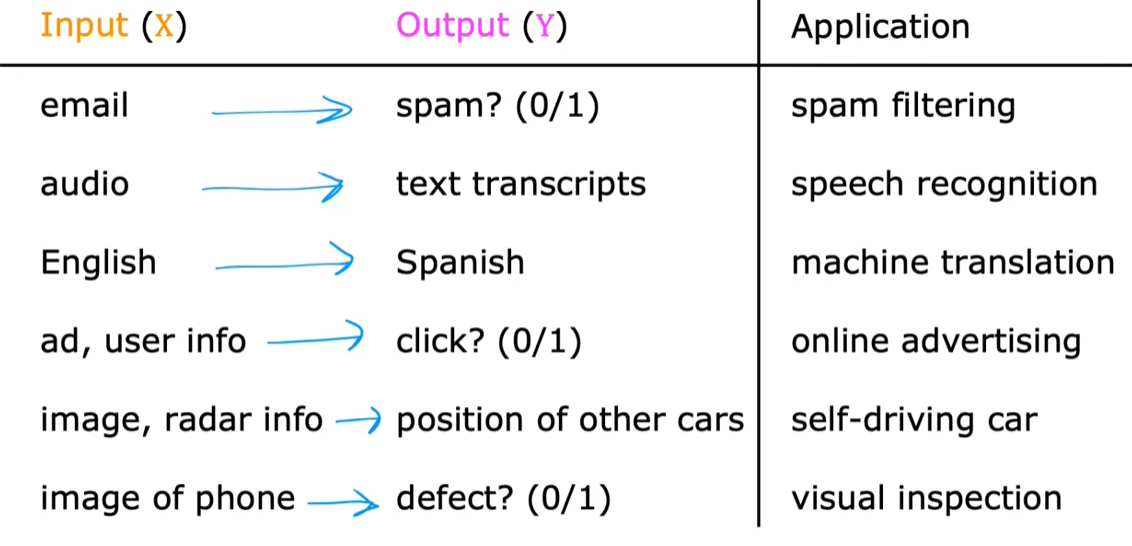
**What is machine learning?**

Field of study that allows the computers to learn without being explicitly programmed.

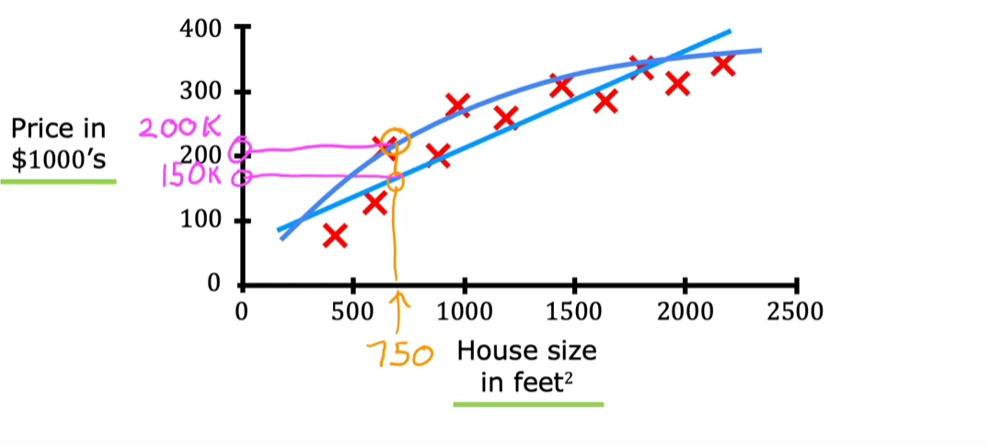
**SUPERVISED LEARNING:** used most in real-world applications (rapid advancements)

You will give input and desired output, and they learn from the desired outputs, i.e. give input X and also the label output Y.



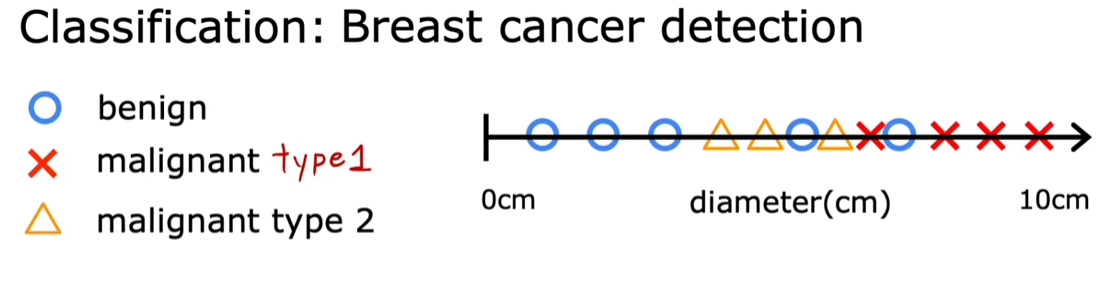
1. **Regression:** Predict an output out of many possible outputs.

Example:



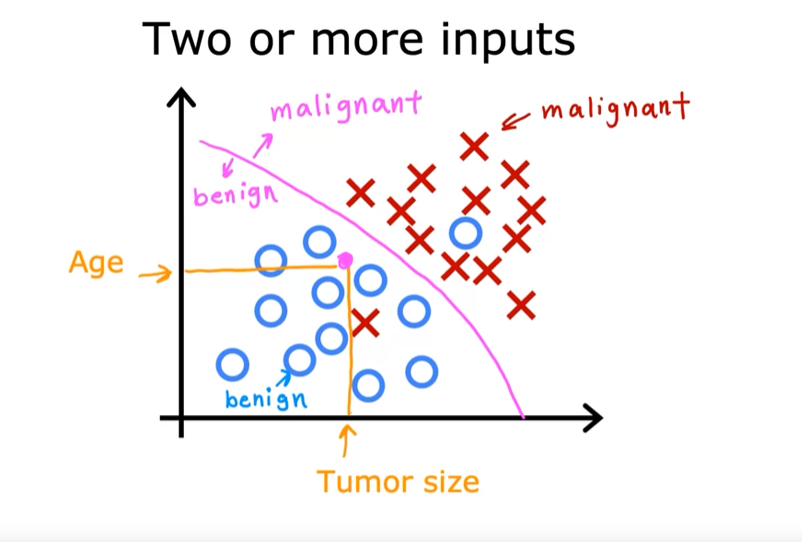
**Here we will try to draw a line or some other curve to get close to the ideal function which will predict the output.**

1. **Classification**: Predict an output out of a limited no.of possible outputs (generally yes/no).

Example:

We can even take two or more inputs.

Example: Age and Tumor size are the inputs



Unsupervised learning involves analyzing data without predefined labels, allowing the algorithm to discover patterns independently. For example, given patient data without tumor classifications, the algorithm seeks to find interesting structures.

So data comes with inputs x, but no labelled outputs y.

**UNSUPERVISED LEARNING:**

It will find something important from the unlabelled data unlike supervised learning which learns from the supervised data.

1. **Clustering:**

Data points are grouped into clusters based on similarities without any labels.

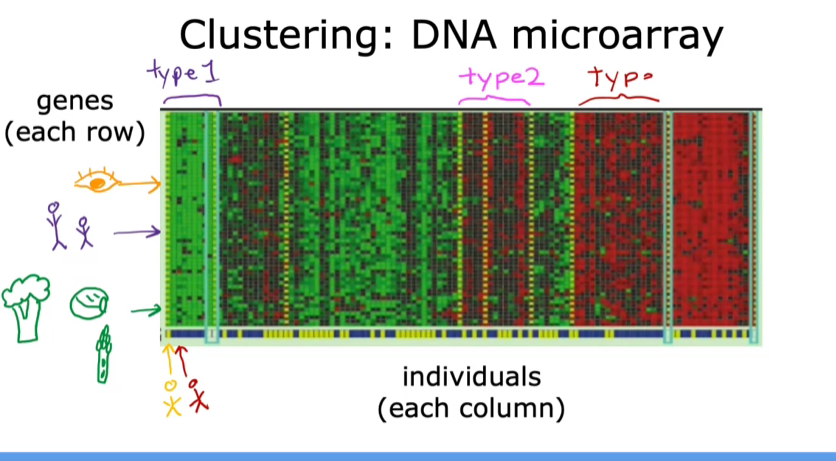
Example:

* Google News uses clustering algorithms to group related news articles, identifying common keywords to create clusters without human intervention.

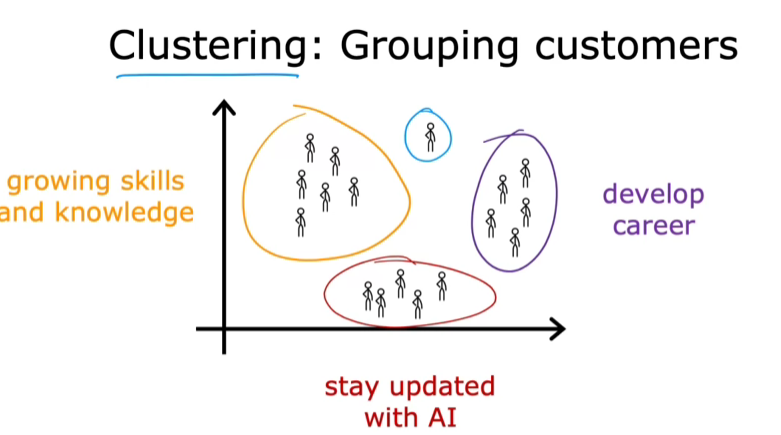
****

Example:

* In genetic research, clustering can categorize individuals based on DNA activity, helping researchers identify different types of people based on genetic traits. Here we don’t tell the machine that there are different type of people here, all we give to it is the data



Example:



1. **Anomaly Detection:**

used to identify unusual events, which is crucial for applications like fraud detection in financial systems.

1. **Dimensionality reduction:**

allows for the compression of large datasets into smaller ones while retaining as much information as possible.