Overview

❓ **If Arthur Samuel's checkers-playing program had been allowed to play only 10 games against itself, how would this have affected its performance compared to when it was allowed to play over 10,000 games?**

Answer: Would have made it worse.

Machine learning algorithms

- **Supervised learning:** supervised learning is the type of machine learning that is used most in many real-world applications and has seen the most rapid advancements and innovation.
- **Unsupervised learning**
- **Recommender systems**
- **Reinforcement learning.**

💡 Practical advice for applying learning algorithms

Supervised learning

$$X(\text{input}) \longrightarrow Y(\text{output label})$$

⇒ Learns from being give **"right answers"**

EXAMPLES:

Input(X)	Output(Y)	Application
Email	spam? $(0/1)$	spam filtering
audio	text transcripts	speech recognition
English	Spanish	machine translation
ad, user info	click? $(0/1)$	online advertising
image, radar info	position of other cars	self-driving car
image of phone	defect? $(0/1)$	visual inspection

1. Train your model with examples of inputs x and the right answers, that is the l**abels** y.
2. Give them a brand new input x, something it has never seen before, and try to produce the appropriate corresponding output y.

Regression

💡 **Specific example: Predict housing prices based on the size of the house.**

- Collecting and plotting the data like this:

Ques: what's the price for their 750 square foot house?

⇒ How can the learning algorithm help you?

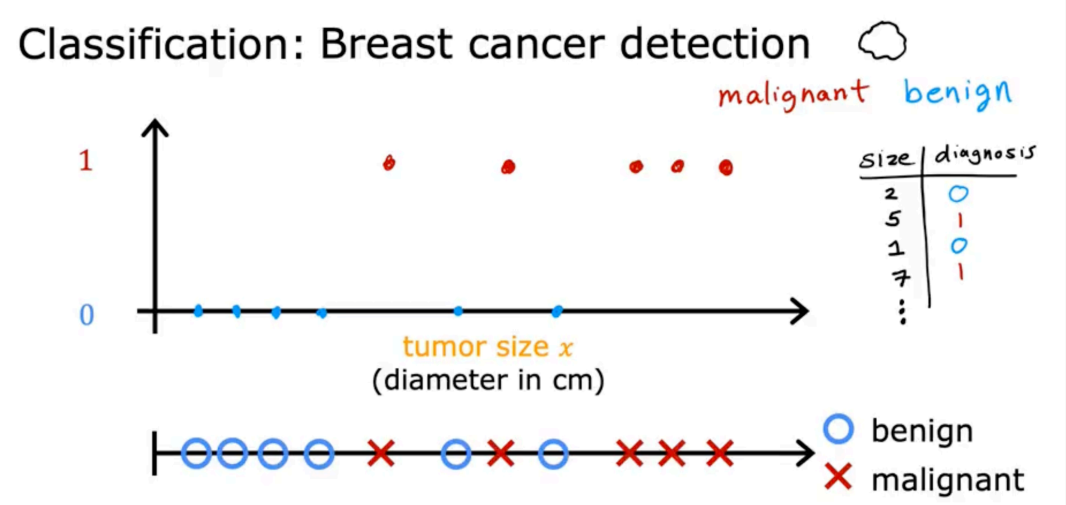
It can fit the straight line or curve line to predict the price with the input.

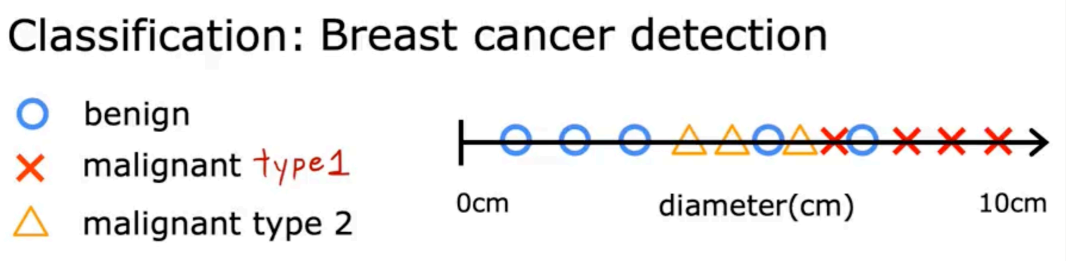
⇒ One of the things you see later in this class is how you can decide whether to fit a straight line, a curve, or another function that is even more complex to the data.

Classification

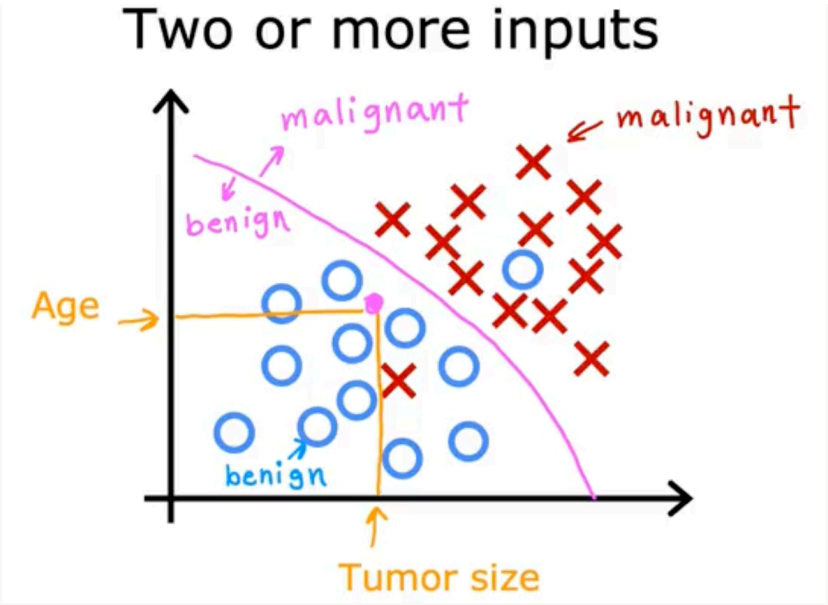
💡 Specific example: **Breast cancer detection**

1. Breast Cancer Detection
    - **Classification Problem**: The algorithm predicts if a tumor is benign (ác tính) (0) or malignant (lành tính) (1) based on medical records.
    - **Data Representation**: Tumors are plotted on a graph with tumor size on the horizontal axis and categories (0 or 1) on the vertical axis.





2. **Multiple Input Variables**:
    - **Example**: Predicting tumor classification using both tumor size and patient age.
    - **Boundary Line**: The learning algorithm finds a boundary that separates benign tumors from malignant ones, aiding in diagnosis.



Types of Supervised Learning

- Regression
  - Predict a number from infinitely many possible numbers.
  - Such as the house prices in our example, which could be 150,000 or 70,000 or 183,000 or any other number in between.
- Classification
  - Classification algorithms predict categories.
  - Predicts a small finite limited set of possible output categories.
  - Such as 0, 1 and 2 but not all possible numbers in between like 0.5 or 1.7. That's makes **classification** different from **regression** when you're interpreting the numbers.

💡 **Supervised learning is when we give our learning algorithm the right answer *y* for each example to learn from. Which is an example of supervised learning?**

Ans: Spam filtering.

- For instance, emails labeled as "spam" or "not spam" are examples used for training a supervised learning algorithm. The trained algorithm will then be able to predict with some degree of accuracy whether an unseen email is spam or not.

Unsupervised learning

- Unsupervised learning involves analyzing data without labeled outputs, aiming to identify patterns or structures within the data.
- The algorithm autonomously discovers interesting features in the dataset, rather than being guided by predefined labels.

Clustering Algorithms

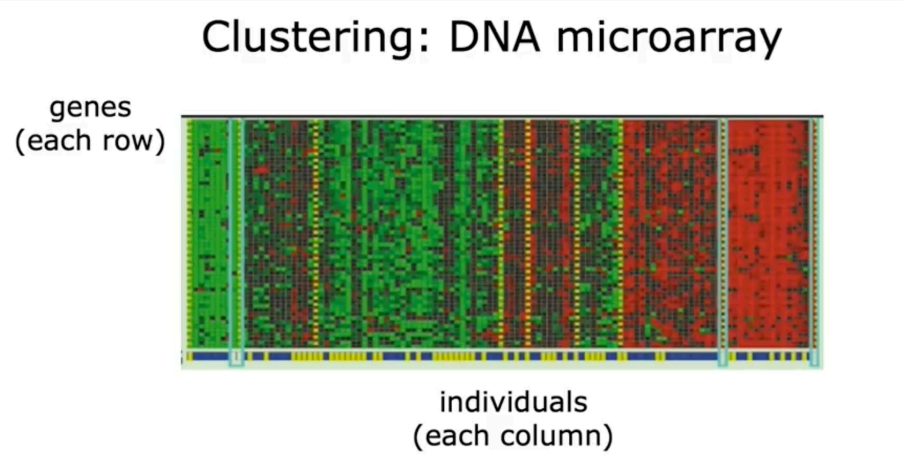
- Clustering is a common type of unsupervised learning that groups unlabeled data into clusters based on similarities.

💡 Example: **Google News Clustering**

- Google News analyzes thousands of articles daily and groups related stories together based on shared keywords.
- For instance, articles about pandas and twins are clustered together because they contain similar terms, allowing users to find related content easily.

💡 The Google News algorithm operates autonomously, identifying clusters of related articles without human intervention. Given the daily influx of diverse news topics, it's impractical for employees to manually categorize articles. Instead, the algorithm independently determines the clusters based on shared content, exemplifying unsupervised learning.

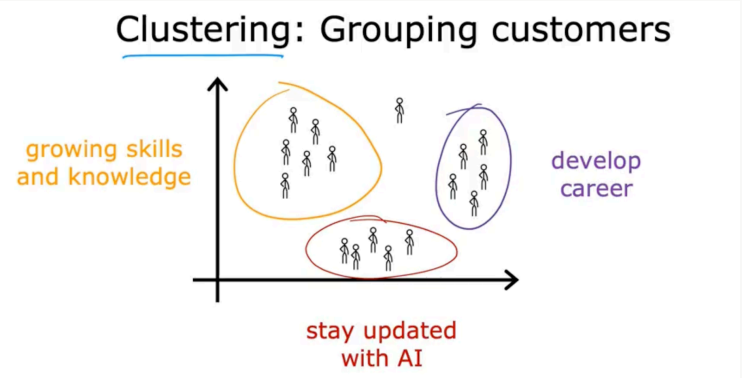
💡 Example: **DNA Microarray Data**:



- This example involves analyzing genetic data where each column represents an individual's DNA activity, and each row corresponds to a specific gene.
- Clustering algorithms group individuals based on gene expression patterns, identifying different types of people without prior labels.

💡 Example: Grouping customers

1. **Market Segmentation**:
   - Companies use clustering algorithms to analyze customer data and identify distinct market segments.
   - This helps in understanding different customer behaviors and preferences.
2. **Deep Learning Community Analysis**:
   - The deep learning community identified groups of learners based on their motivations:
     - Some seek knowledge to improve skills.
     - Others aim to advance their careers.
     - Some want to stay updated on AI's impact in their fields.



Unsupervised Learning

Data only comes with inputs x, but not output labels y. Algorithm has to find *structure* in the data.

- **Clustering** is a common type of unsupervised learning that groups unlabeled data into clusters based on similarities.
- **Anomaly detection** is used to identify unusual events, which is crucial for applications like fraud detection in finance.
- **Dimensionality reduction (PCA)** allows for the compression of large datasets into smaller ones while retaining essential information.

💡 Ques: