

# SUPERVISED LEARNING

## Supervised Learning

- **Definition:**
  - Algorithms learn a mapping from **input** ( $x$ ) to **output** ( $y$ ).
- **Training data:**
  - Consists of labeled examples  $((x, y))$ .
- **After training:**
  - The model predicts ( $y$ ) for **new, unseen inputs** ( $x$ ).

## Supervised Learning Workflow

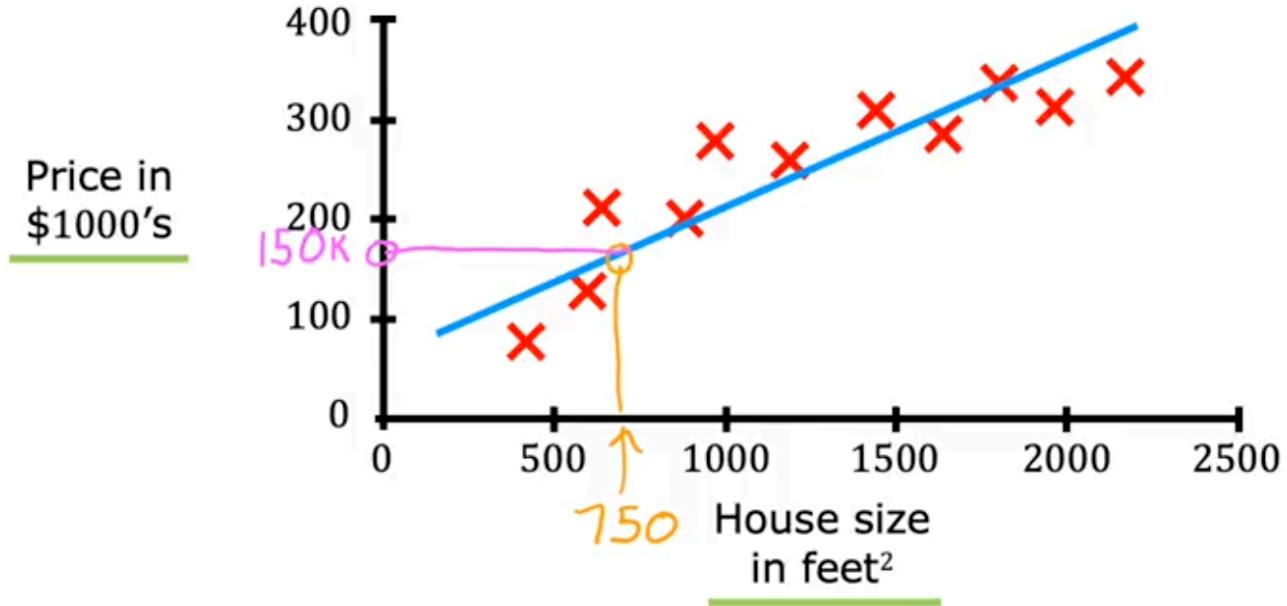
1. Collect labeled data pairs  $((x, y))$
  2. Train a model on this data
  3. Use the trained model to predict ( $y$ ) for new inputs ( $x$ )
- 

## Regression

- **Type:** Supervised learning
- **Goal:**
  - Predict a **numeric value** from infinitely many possible values
- **Example:**
  - Predicting house prices (e.g., \$150,000, \$183,000)

## Example: Housing Price Prediction

- **Task:**
  - ( $x$ ): house size (square feet)
  - ( $y$ ): house price (thousands of dollars)
- **Learning algorithm:**
  - Fits a function to labeled data (e.g., a line or curve)
  - Predicts price for a new house (e.g., 750 sq ft)

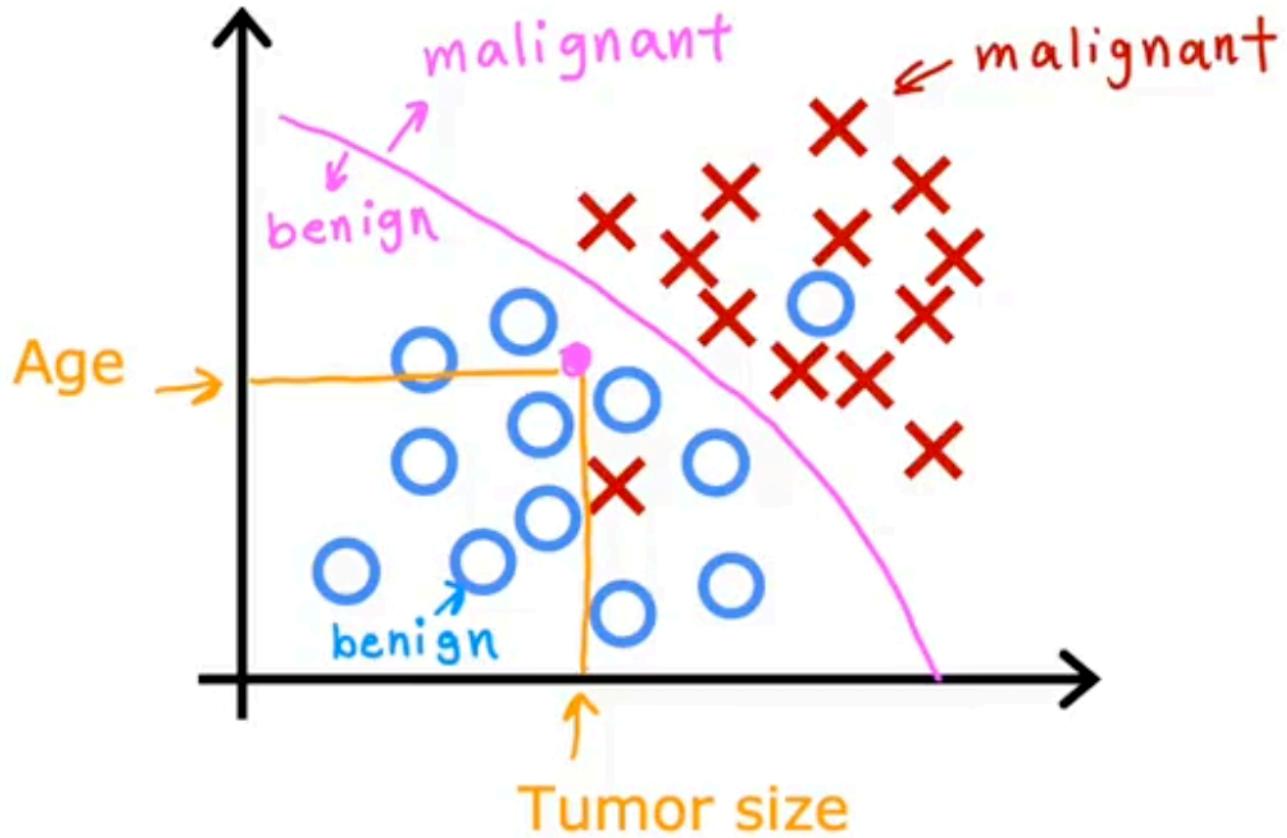


## Classification

- **Type:** Supervised learning
- **Goal:**
  - Predict a **category** from a small, finite set of outputs
- **Terminology:**
  - Outputs are called **classes** or **categories**
- **Key difference from regression:**
  - Does **not** predict arbitrary numeric values

## Example: Breast Cancer Detection

- **Task:**
  - ( $x$ ): tumor features (e.g., size, patient age)
  - ( $y$ ): tumor type
    - **0**: benign
    - **1**: malignant



## Classification vs. Regression

- **Regression:**
  - Predicts numbers from infinitely many values
  - Example: house price = \$183,000
- **Classification:**
  - Predicts one of a limited number of categories
  - Example: 0 or 1 (benign / malignant)
  - Cannot output intermediate values (e.g., 0.5)

## Binary vs. Multiclass Classification

- **Binary classification:**
  - Two possible output classes
  - Examples:
    - Spam / not spam
    - Benign / malignant

- **Multiclass classification:**
    - More than two output classes
    - Example:
      - Cancer type 1, type 2, type 3
- 

## Inputs with Multiple Features

- Supervised learning can use **many input features**
  - Example features:
    - ( $x_1$ ): tumor size
    - ( $x_2$ ): patient age
    - Additional features:
      - Cell size uniformity
      - Cell shape uniformity
      - Tumor thickness
  - Data points exist in a **high-dimensional feature space**
- 

## Decision Boundary

- Classification algorithms learn a **decision boundary**
- The boundary:
  - Separates different classes in feature space
  - Is used to classify new examples