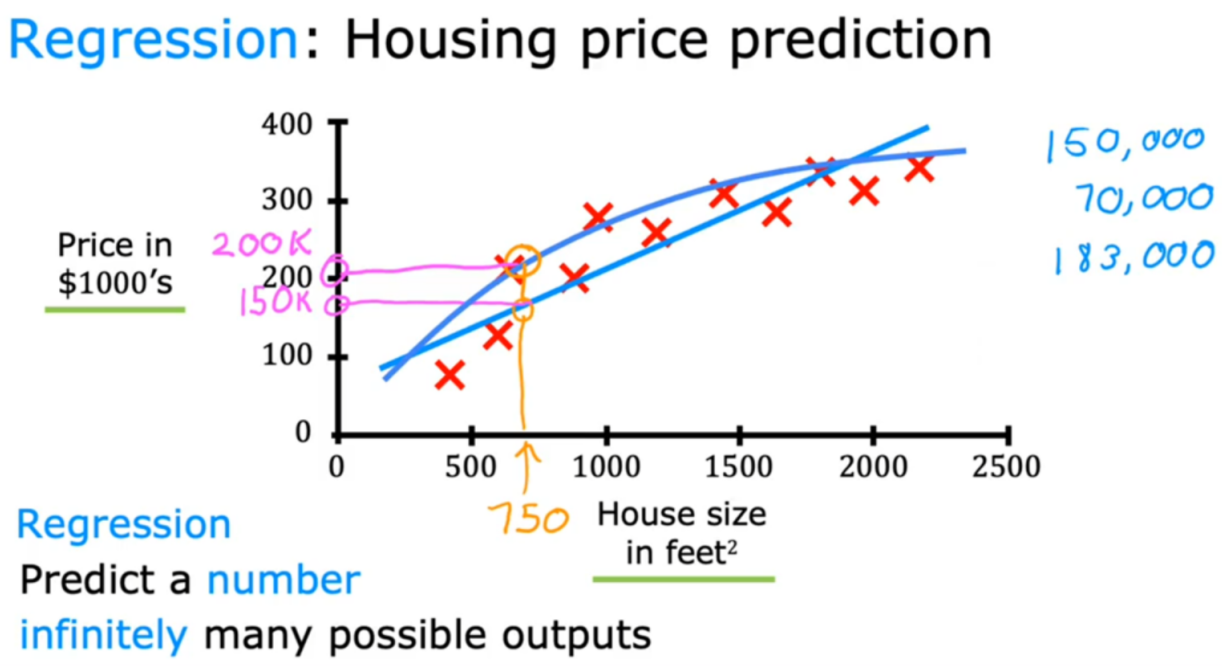
# 2.监督学习和无监督学习

## 2.1 机器学习是什么

- 塞缪尔的定义：“给予电脑不需要明确编程的自学习能力的研究领域”。
- 机器学习的主要分类：
  - 监督学习（使用最多，而且进步速度最快）
  - 无监督学习
  - 强化学习（不在本次课程的范围内）

## 2.2 监督学习part1

- 监督学习中最为常见的是预测类别的学习,也就是回归问题




- 输入x经过函数f的运算得到输出y

$$x \xrightarrow{f} y$$

这个f可以是任意形式的函数，所以无法直接求解通过已知的x_0 y_0,.....等输入输出对，算出一个接近的f'

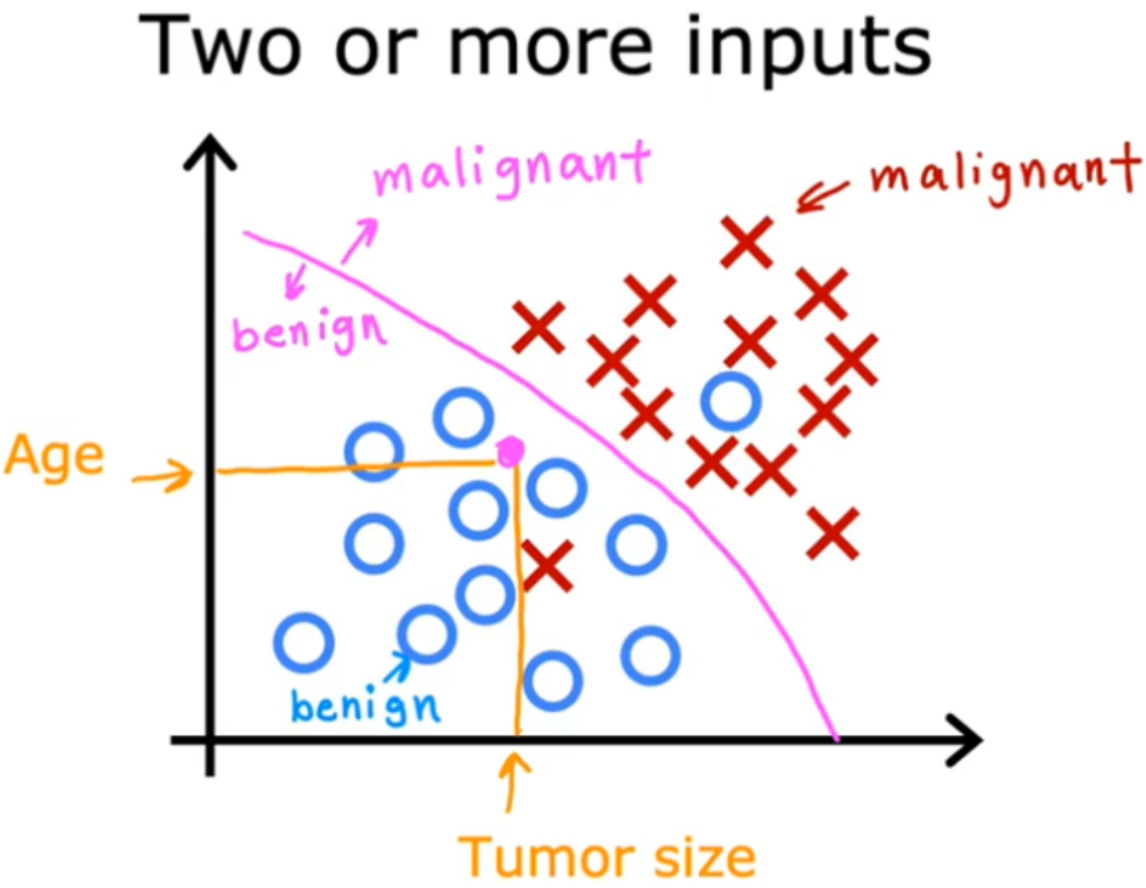
$$ML(x_0, y_0, .....) \longrightarrow f'$$

最后通过f'来预测新输入x'的输入y'

$$x' \xrightarrow{f'} y''$$

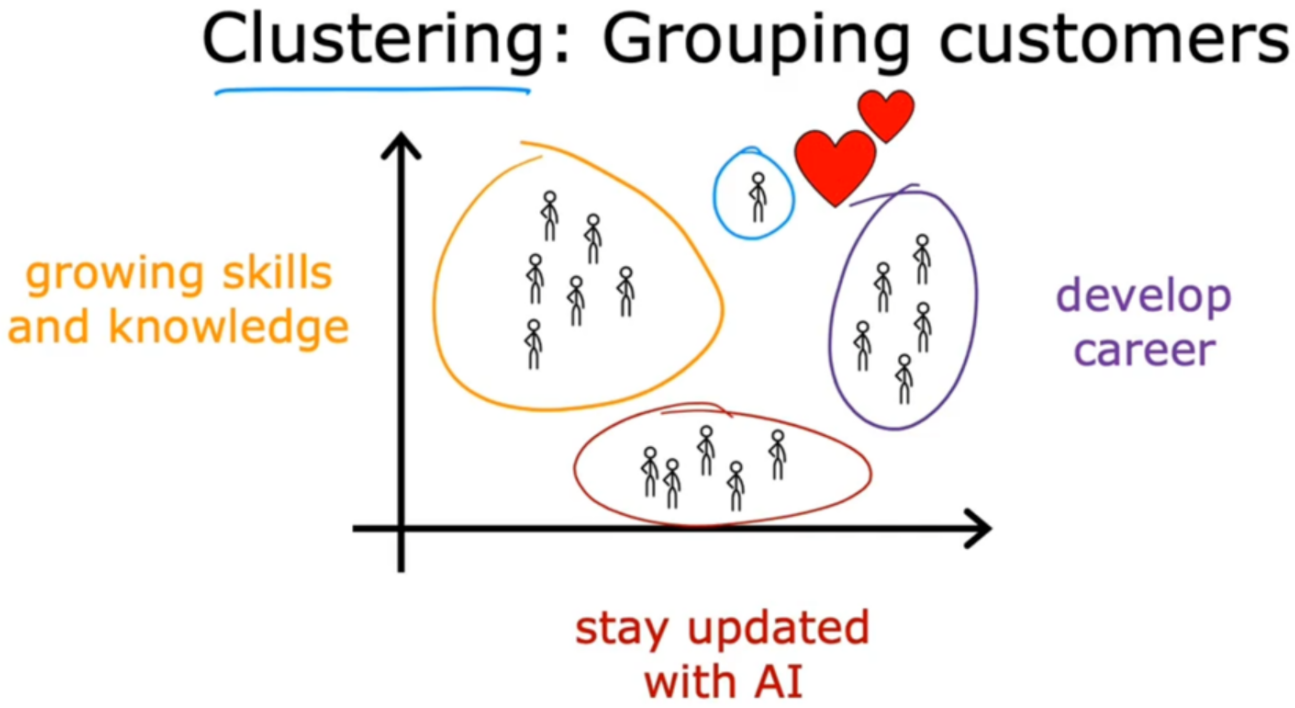
## 2.3 监督学习part2

- 如果输出是受限的，比如只有有限的选择，那么回归问题会变为分类问题。因此，输出可以不再是数，而是由分类方式决定的。




## 2.4 无监督学习part1

- 无监督学习不再有预先准备好的正确数据，而是交由机器去发掘数据中呈现出的特殊性。
- 代表算法为聚类算法（在大量数据中，划分出（给定或不给定数量的）多块，这些块内部更加相关，而且互相之间差异较大）




## 2.5 无监督学习part2

- 异常检测也是一种无监督学习
- 数据压缩：将一个大的数据集减小到一个较小的数据集

## 2.6 jupyter入门

本节课主要引入了jupyter notebook的应用。

lab1-0：尝试最为基本的ipynb文件