1 Inroduction

1.1 欢迎参加《机器学习》课程 | Welcome

- 机器学习在日常生活中有很多应用（e.g.搜索引擎排序，照片人脸识别etc）
- 但是只知道算法和数学公式是不够的，所以本课程还是设计了许多project，让你去了解ML的应用

（1）Machine Learning

- Grew out of work in Al（ML是AI发展出来的一个领域）
- New capability for computers

（2）Examples

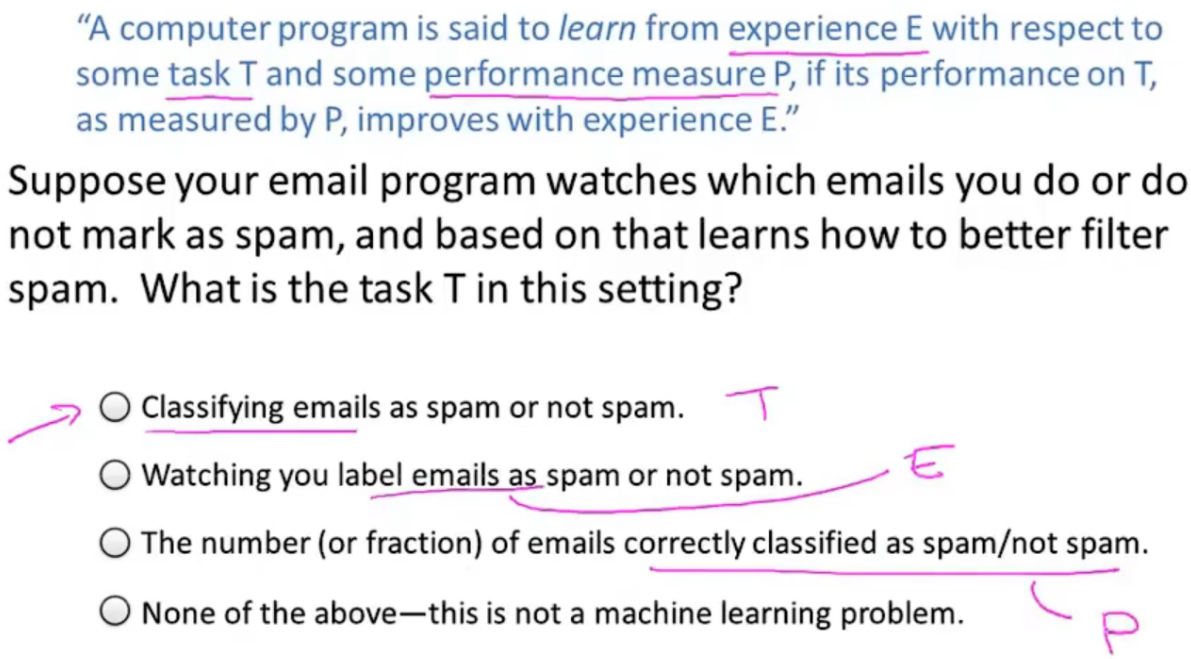
- Database mining
  Large datasets from growth of automation/web.
  E.g., Web click data, medical records, biology, engineering
- Applications can't program by hand.
  E.g., Autonomous helicopter, handwriting recognition, most of Natural Language Processing(NLP), Computer Vision.
- Self-customizing programs
  E.g., Amazon, Netflix product recommendations

1.2 什么是机器学习 | What is machine Learning

（1）Machine Learning definition

- Arthur Samuel (1959).Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed.
- Tom Mitchell (1998) Well-posed LearningProblem: **A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.**

练习**1**：判断**E/T/P**



(2)Machine learning algorithms

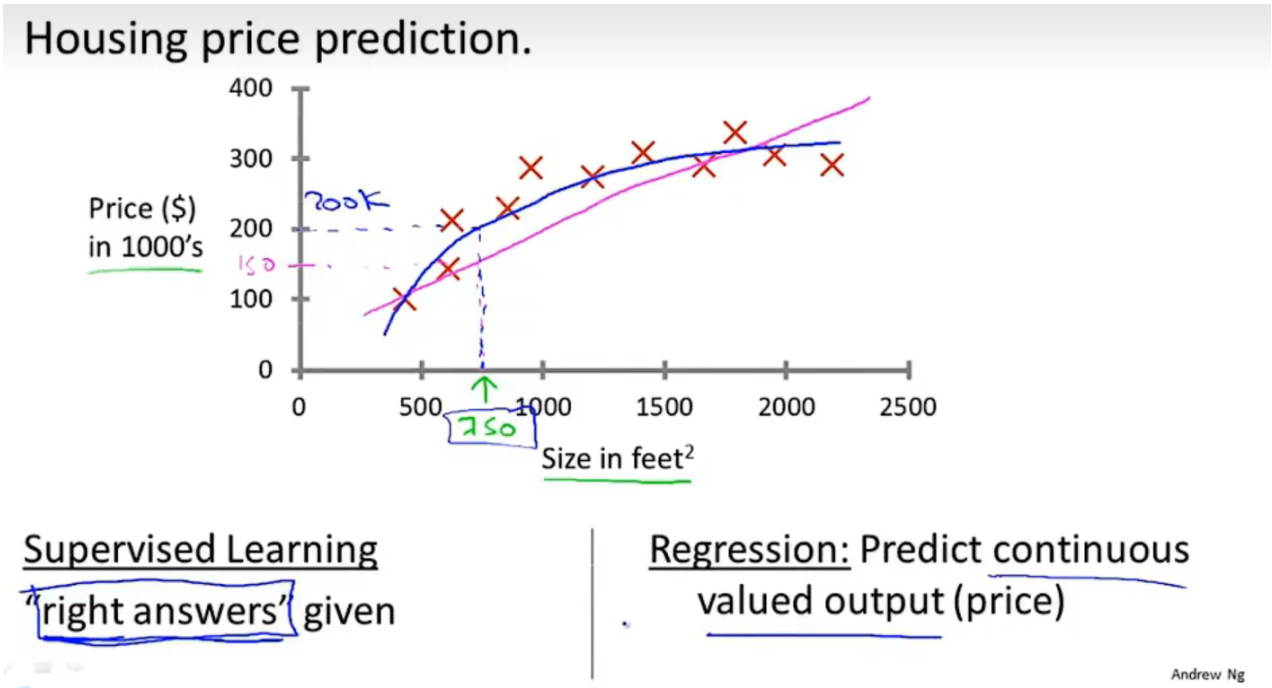
- Supervised learning
- Unsupervised learning

Others: Reinforcement learning, recommender systems.

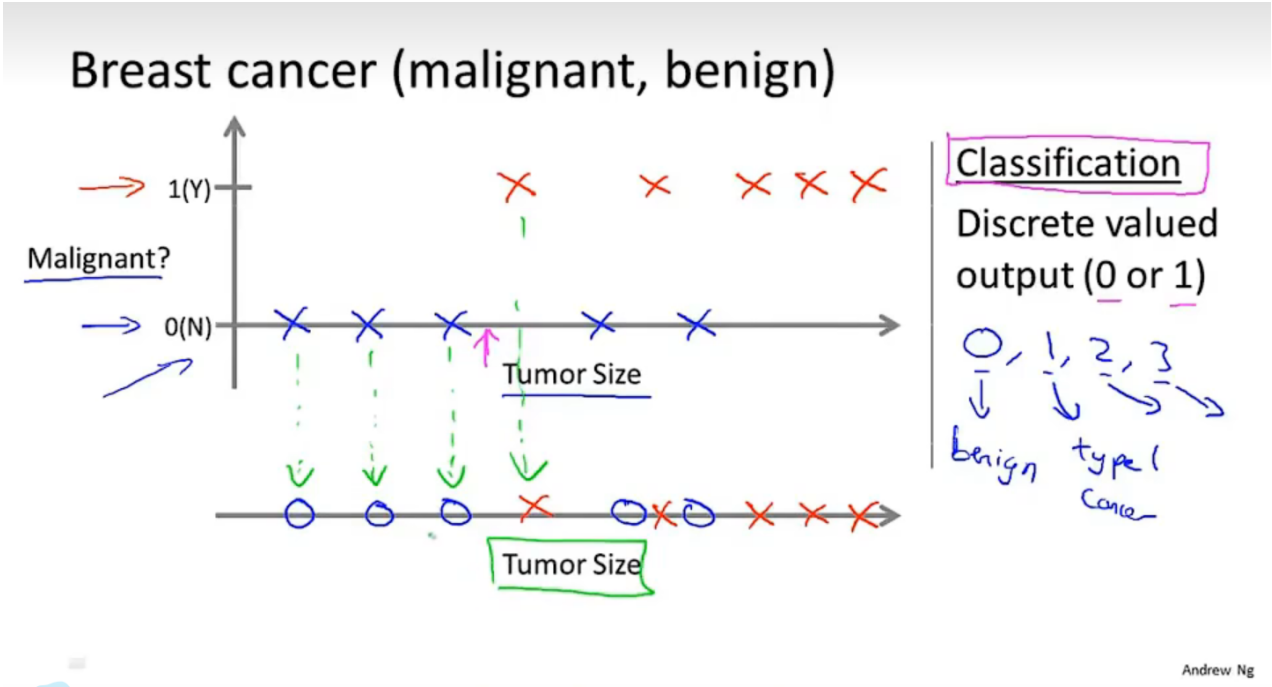
Also talk about: Practical advice for applying learning algorithms.

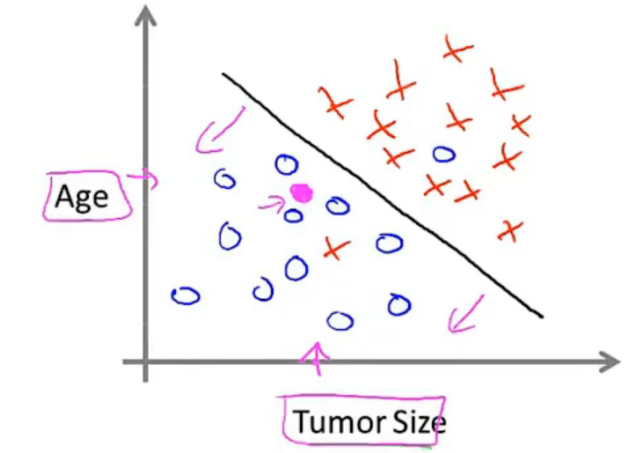
1.3 监督学习 | Supervised learning

（**1**）**Regression** 回归



（2）**Classification** 分类





- Clump Thickness
- Uniformity of Cell Size
- Uniformity of Cell Shape

...

Andrew Ng

练习**2**：判断回归和分类

You're running a company, and you want to develop learning algorithms to address each of two problems.

1000's

Problem 1: You have a large inventory of identical items. You want to predict how many of these items will sell over the next 3 months.

Problem 2: You'd like software to examine individual customer accounts, and for each account decide if it has been hacked/compromised. 0 – not hacked, 1 – hacked

Should you treat these as classification or as regression problems?

- ○ Treat both as classification problems.
- ○ Treat problem 1 as a classification problem, problem 2 as a regression problem.
- → ○ Treat problem 1 as a regression problem, problem 2 as a classification problem.
- ○ Treat both as regression problems.

1.4 无监督学习 | Unsupervised learning

聚类算法（e.g. 新闻算法，市场分类etc）

其他算法，如鸡尾酒酒会算法

练习3：分辨监督学习和无监督学习

Of the following examples, which would you address using an unsupervised learning algorithm? (Check all that apply.)

- ☐ Given email labeled as spam/not spam, learn a spam filter.
- ☑ Given a set of news articles found on the web, group them into set of articles about the same story.
- ☑ Given a database of customer data, automatically discover market segments and group customers into different market segments.
- ☐ Given a dataset of patients diagnosed as either having diabetes or not, learn to classify new patients as having diabetes or not.