Machine Learning

Introduction

Machine Learning Definition

- ML is the science of getting computers to learn and act like humans do, and improve their learning over time in autonomous fashion, by feeding them data and information in the form of observations and real-world interactions.

- ML at its most basic is the practice of using algorithms to parse data, learn from it, and then make a determination or prediction about something in the world.

Machine Learning Definition

- ML [Arthur Samuel(1959)]: Field of study that gives computers the ability to learn without being explicitly programmed.

- Well-posed Learning Problem: 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.

"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."

Suppose your email program watches which emails you do or do not mark as spam, and based on that learns how to better filter spam. What is the task T in this setting?

- ◯ Classifying emails as spam or not spam.
- ◯ Watching you label emails as spam or not spam.
- ◯ The number (or fraction) of emails correctly classified as spam/not spam.
- ◯ None of the above—this is not a machine learning problem.

Where We Use Machine Learning?

- **Recommendations**
  - Intelligent machine learning algorithms analyze your activity and compare it to the millions of other users to determine what you might like to buy or binge watch next.
  - More than 80 percent of TV shows on **Netflix** are found through its recommendation engine. Due to its impact on customer retention, Netflix uses Machine learning to recommend new series.
- **Spam Detection**
- **Image Processing**
- **Natural Language Processing**
- **Text Analysis**
- **Voice recognition**
- **Cancer prediction**

Real-World Examples of ML and AI

1. Siri & Cortana

Voice recognition systems such as Siri and Cortana use machine learning and deep neural networks to imitate human interaction. As they progress, these apps will learn to ‘understand’ the nuances and semantics of our language.

2. Facebook

The social network’s algorithms recognise familiar faces from your contact list, using some impressive technology.

3. Spotify

Much like Netflix, Spotify uses machine learning to figure out your likes and dislikes and provides you with a list of related tracks.

Machine Learning Algorithms

- **With respect to given correct answers**
- Supervised Learning
- Unsupervised Learning

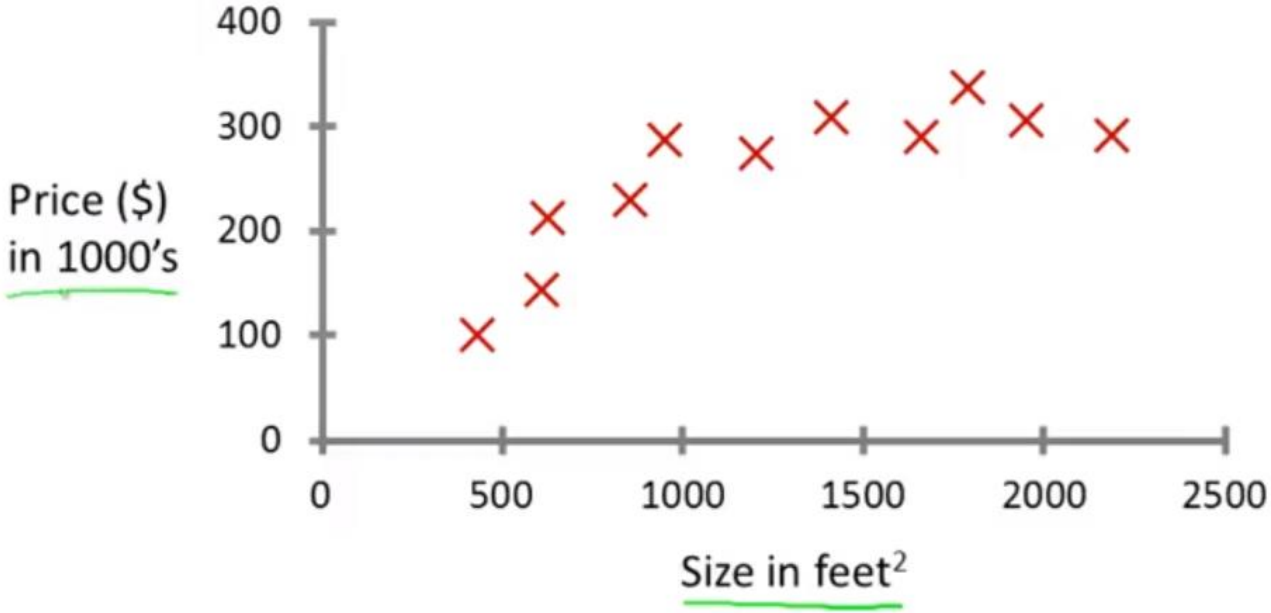
- **With respect to parametrization**
- Parametric learning
- Non-parametric learning

Key types of Machine Learning problems

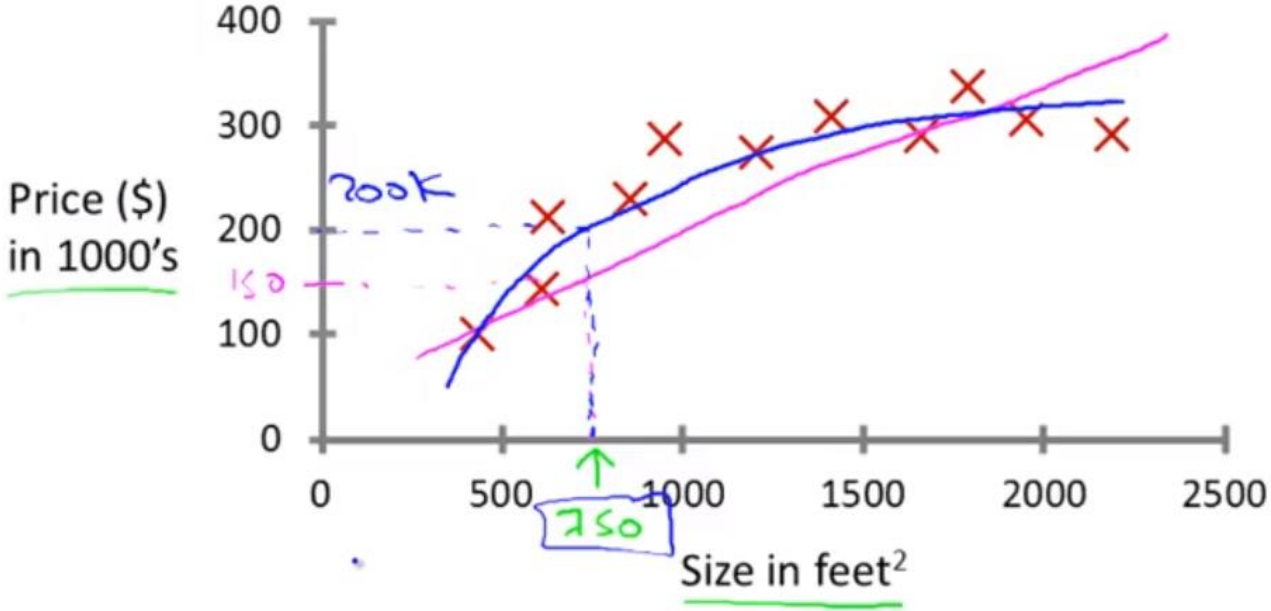
Supervised machine learning: Learn to predict target values from labelled data.

- Classification (target values are discrete classes)
- Regression (target values are continuous values)

Housing price prediction.



Housing price prediction.



Supervised Learning

"right answers" given

Regression: Predict continuous valued output (price)

Supervised Learning (classification example)

Training set

X Sample		Y Target Value (Label)	
	x_1	Apple	y_1
	x_2	Lemon	y_2
	x_3	Apple	y_3
	x_4	Orange	y_4

Classifier

$f : X \rightarrow Y$

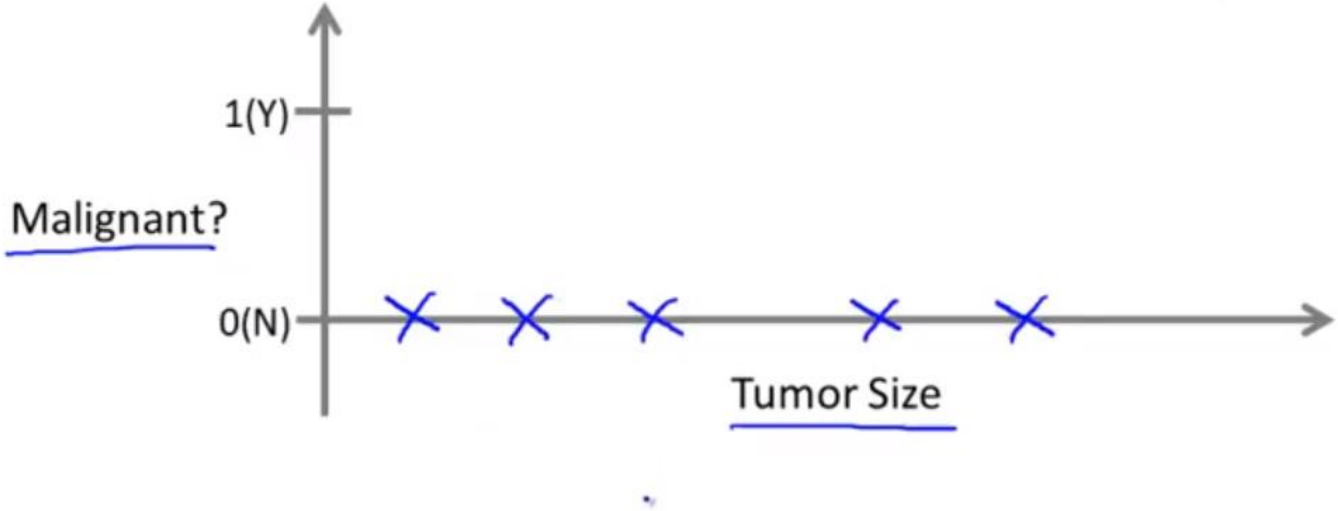
At training time, the classifier uses labelled examples to learn rules for recognizing each fruit type.

Future sample

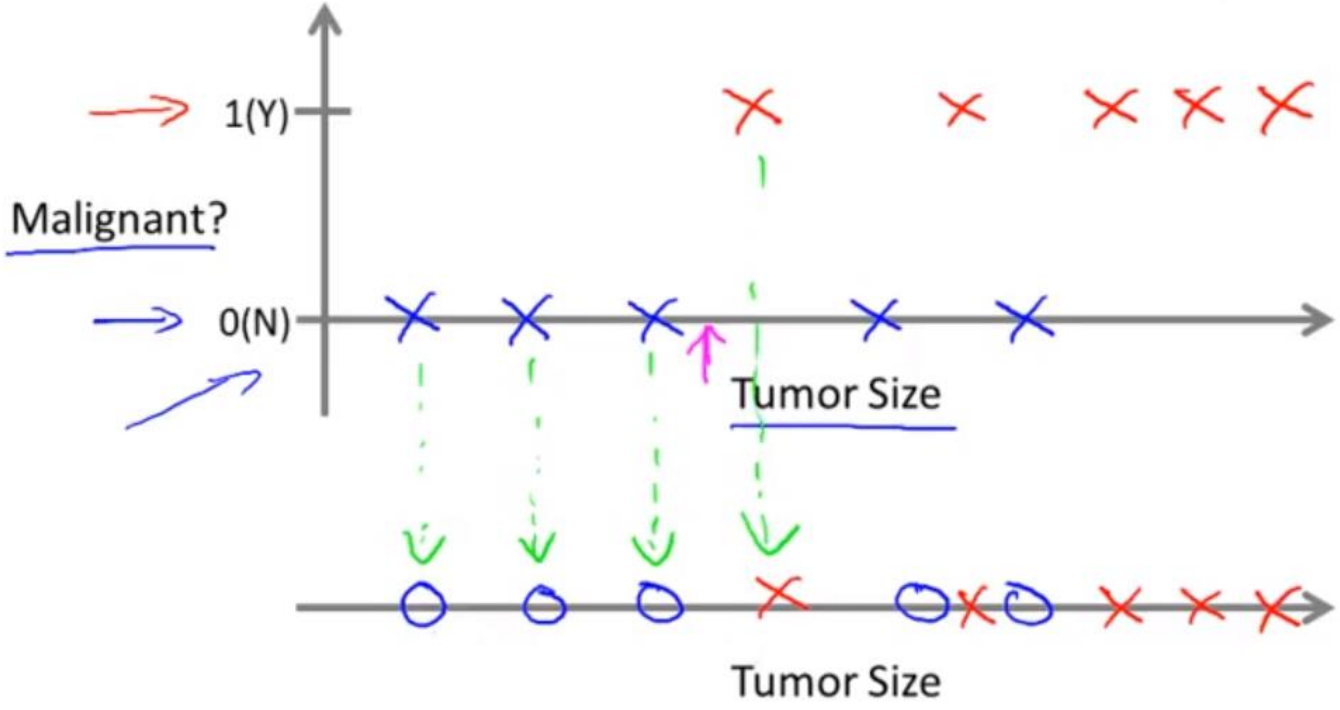
Label: Orange

After training, at prediction time, the trained model is used to predict the fruit type for new instances using the learned rules.

Breast cancer (malignant, benign)

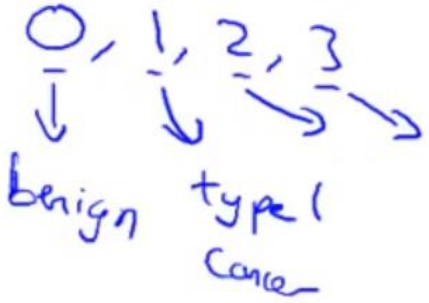


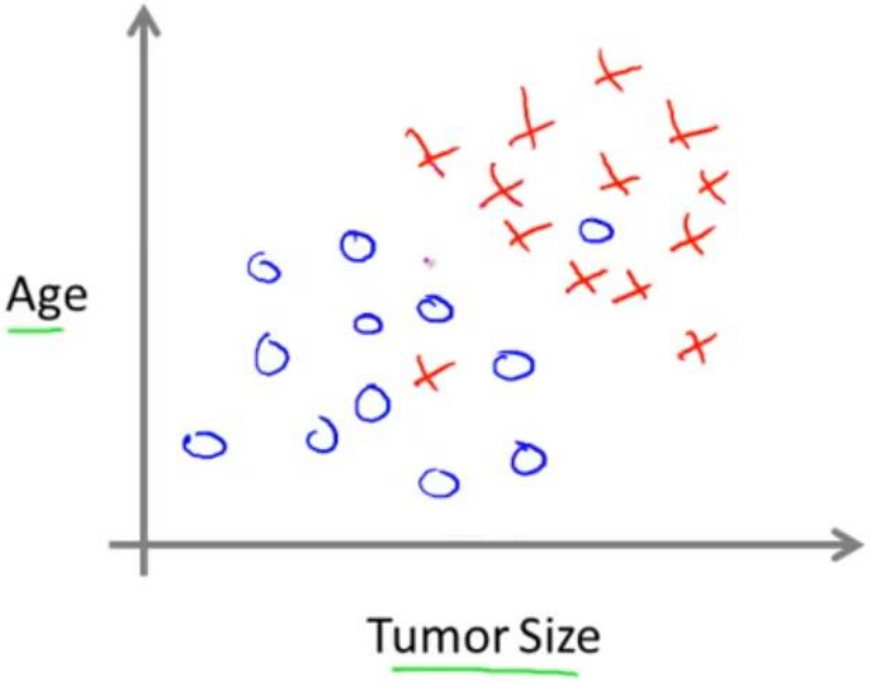
Breast cancer (malignant, benign)

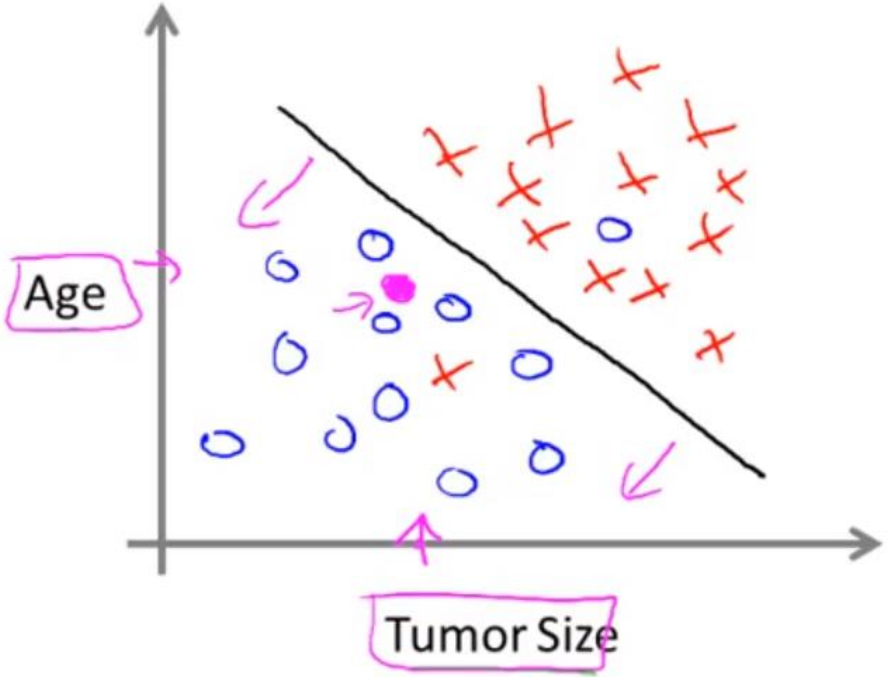


Classification

Discrete valued output (0 or 1)







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

...

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

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.

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.

UNSUPERVISED LEARNING

Key types of Machine Learning problems

Supervised machine learning: Learn to predict target values from labelled data.

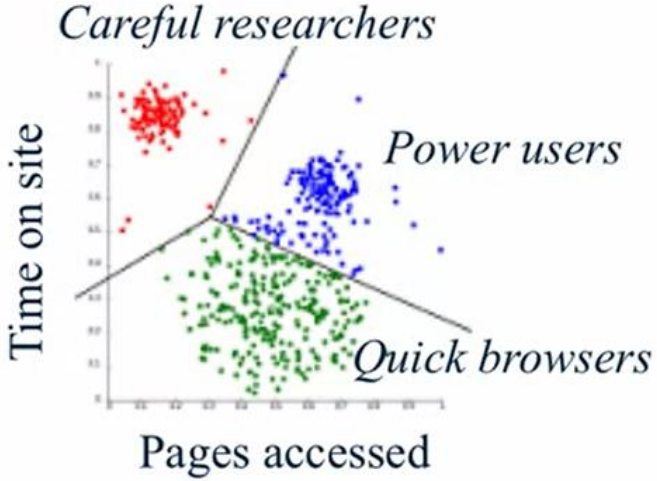
- Classification (target values are discrete classes)
- Regression (target values are continuous values)

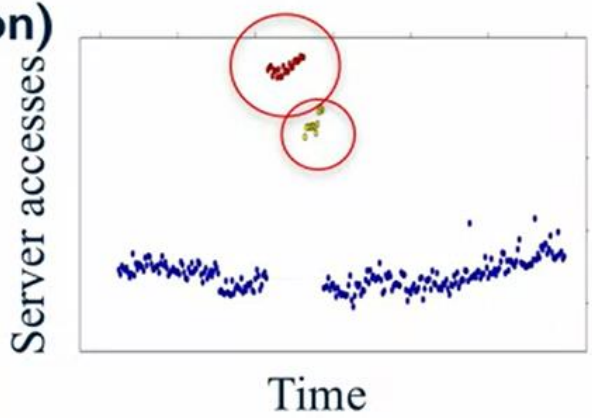
Unsupervised machine learning: Find structure in *unlabeled data*

- **Find groups of similar instances in the data (clustering)**
- **Finding unusual patterns (outlier detection)**

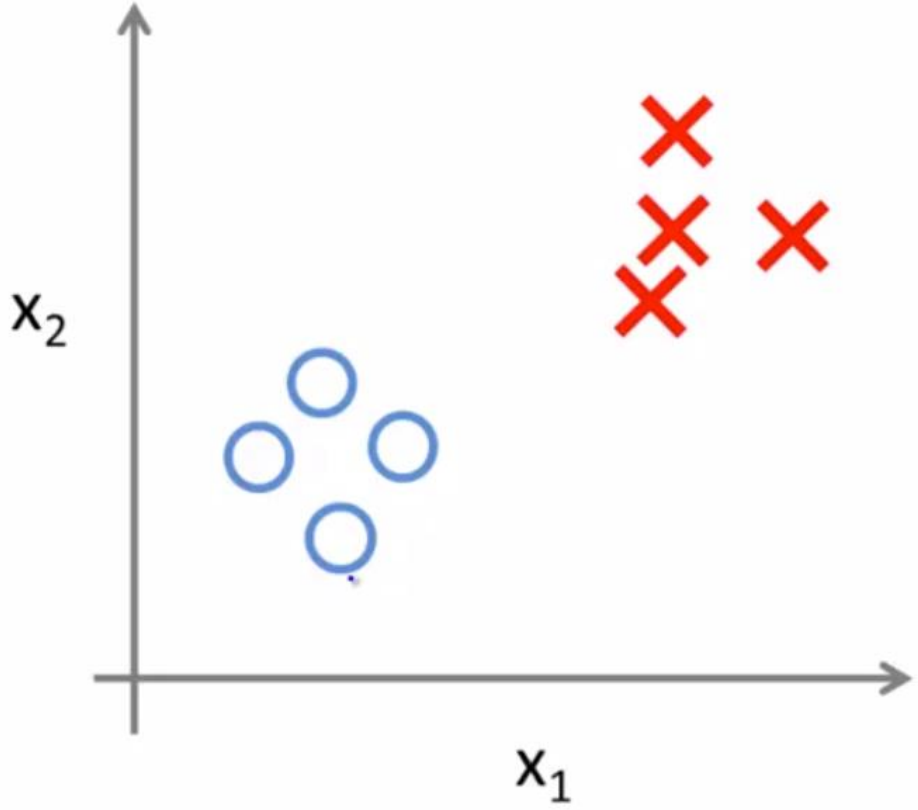
Unsupervised learning: finding useful structure or knowledge in data when no labels are available

- Finding clusters of similar users (clustering)
- Detecting abnormal server access patterns (unsupervised outlier detection)

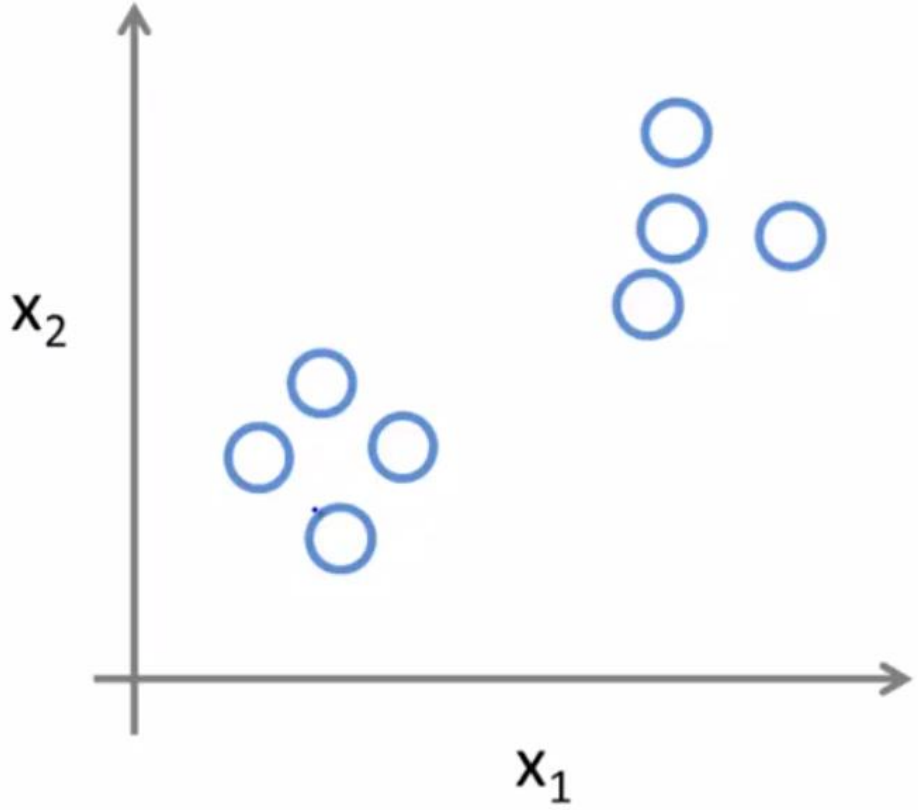




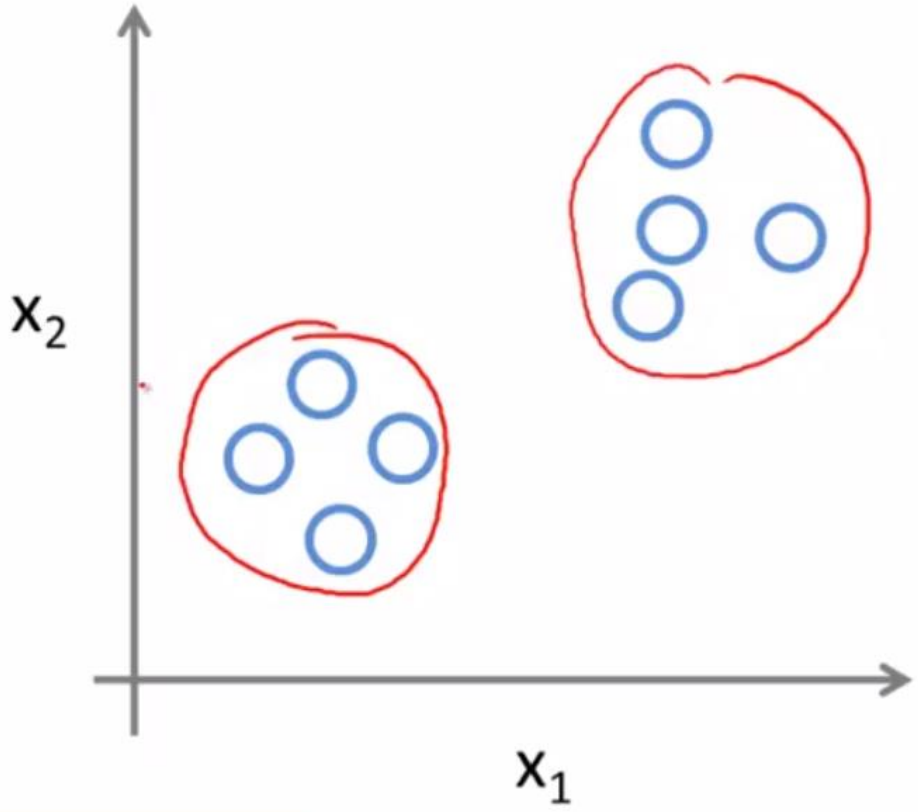
Supervised Learning



Unsupervised Learning



Unsupervised Learning



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.

A Basic Machine Learning Workflow

Choose:
- A feature representation
- Type of classifier to use

e.g. image pixels, with k-nearest neighbor classifier

Choose:
- What criterion distinguishes good vs. bad classifiers?

e.g. % correct predictions on test set

Choose:
- How to search for the settings/parameters that give the best classifier for this evaluation criterion

e.g. try a range of values for "k" parameter in k-nearest neighbor classifier

Feature Representations

Feature representation

Email

```
To:  Chris Brooks
From:  Daniel Romero
Subject:  Next course offering
Hi Daniel,
Could you please send the outline for the
next course offering?  Thanks! -- Chris
```

Feature	Count
to	1
chris	2
brooks	1
from	1
daniel	2
romero	1
the	2
...	

A list of words with their frequency counts

Picture

A matrix of color values (pixels)

Sea Creatures

Feature	Value
DorsalFin	Yes
MainColor	Orange
Stripes	Yes
StripeColor1	White
StripeColor2	Black
Length	4.3 cm

A set of attribute values

Supervised Learning (classification example)

Training set

X Sample		Y Target Value (Label)	
	x_1	Apple	y_1
	x_2	Lemon	y_2
	x_3	Apple	y_3
	x_4	Orange	y_4

Classifier

$f : X \rightarrow Y$

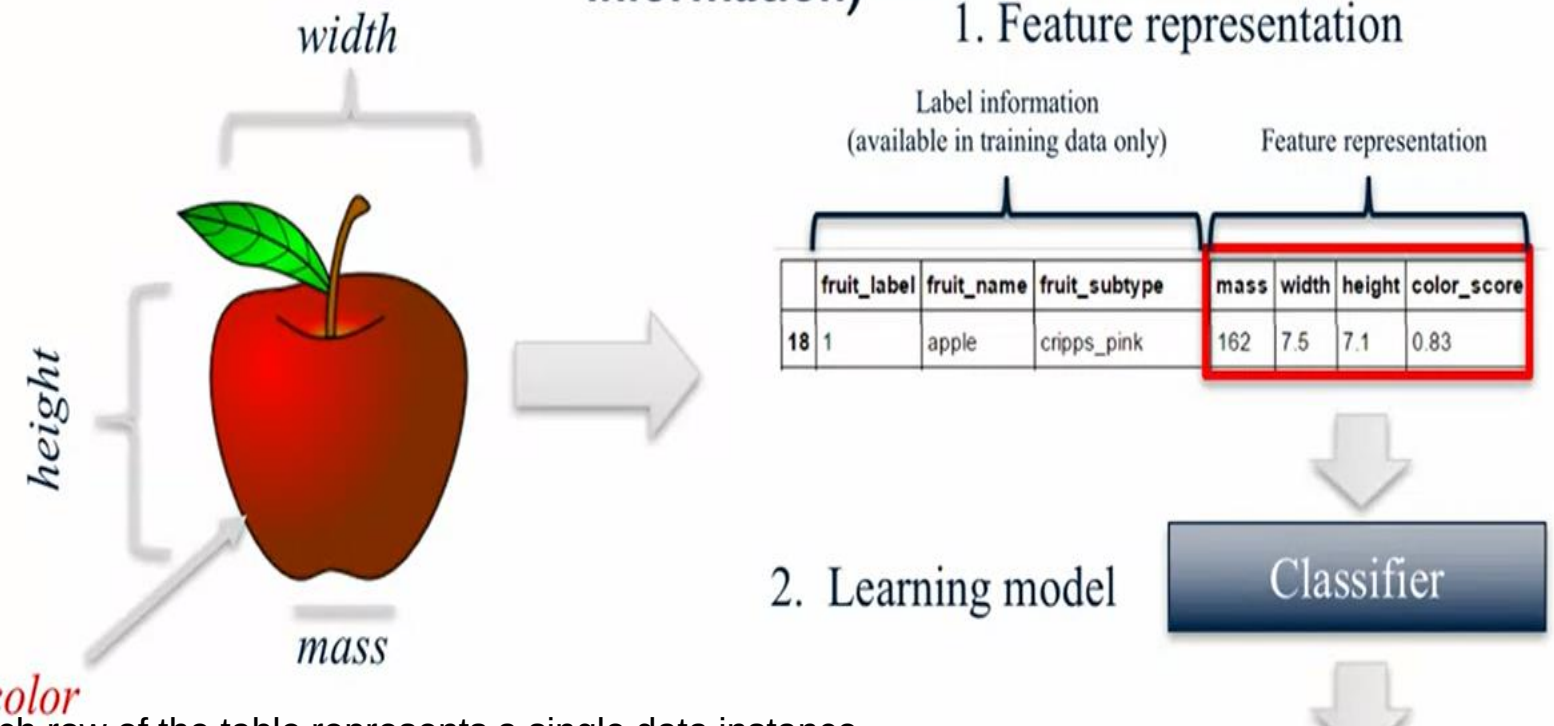
At training time, the classifier uses labelled examples to learn rules for recognizing each fruit type.

Future sample

Label: Orange

After training, at prediction time, the trained model is used to predict the fruit type for new instances using the learned rules.

Representing a piece of fruit as an array of features (plus label information)



- Each row of the table represents a single data instance.
- The columns of the table represent the features of the object
- There could be many different choices for how to represent an object using features.

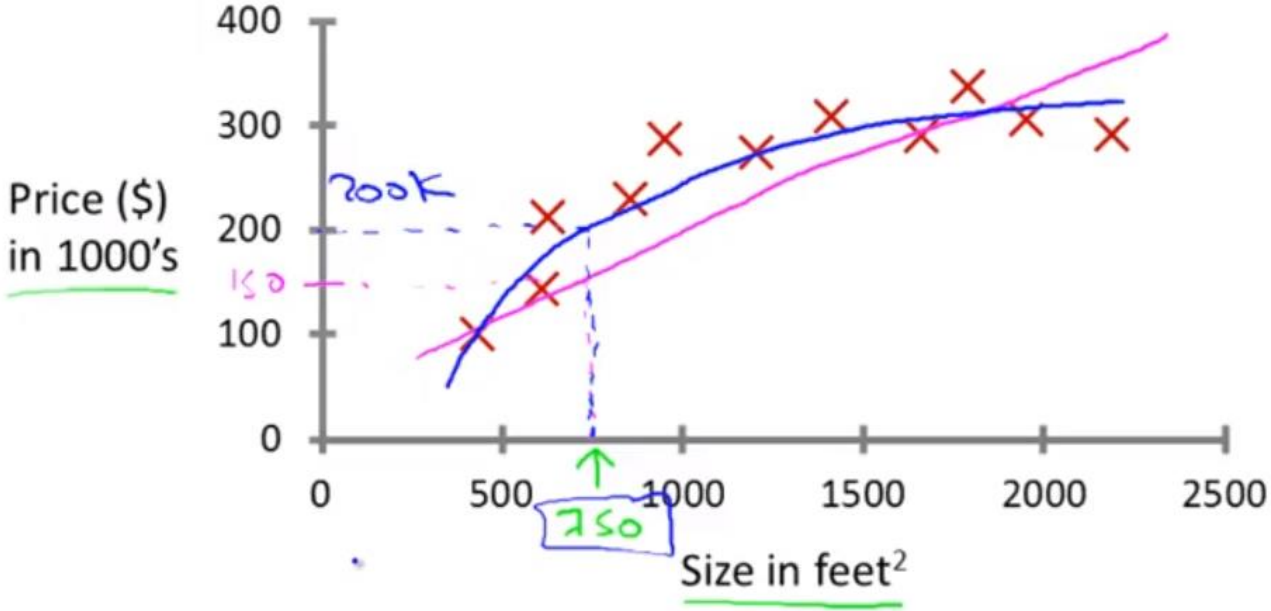
Feature Representation

- So this problem of trying to figure out how to represent an object for machine learning algorithm it's a challenge and it's known as **feature engineering** or **feature extraction**.

- The other key part of representing a machine learning problem is choosing the type of classifier that's appropriate for the problem.
    - They all have different trade-offs in terms of their accuracy, their interpretability, and their speed, and so forth.

- Often the process of addressing a machine learning task is a cycle.
- It involves an iterative process,
    - we make an initial guess about features and the classifier
    - We then train the system using our training data.
    - Produce an evaluation, see how well the classifier works
    - Failure analysis: Check what works and what does not work, which examples are misclassified, etc.
    - Refine the set of feature.

Housing price prediction.



Supervised Learning

"right answers" given

Regression: Predict continuous valued output (price)

Credits

As I wrote in the syllabus, I am following the lectures of Prof. Andrew NG from Stanford University in Coursera and capture the slides from his online lecture. Please visit the following page for that lecture:

https://www.coursera.org/learn/machine-learning