Intro to Machine Learning

Knowledge Sharing for CPE/SKE students

Sirakorn Lamyai

Student, Kasetsart U.

October 10, 2018

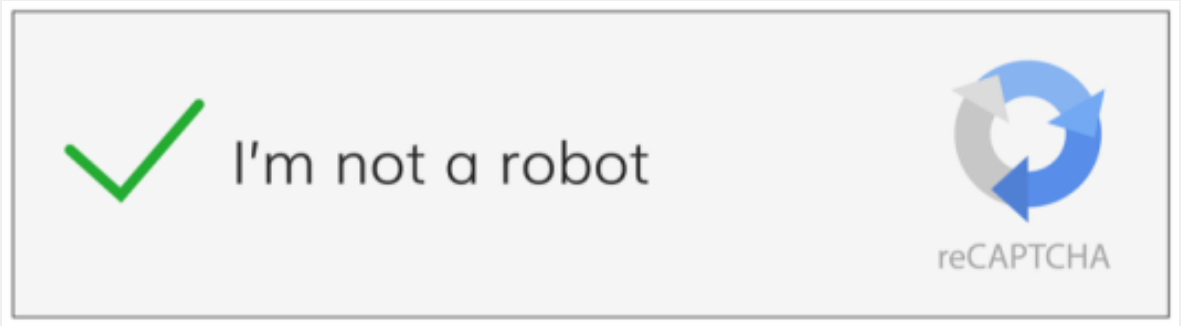
Outline

1. Introduction to Machine Learning
   - What is Machine Learning?
     - Traditional programming approach
     - Machine learning approach
2. Machine Learning Problems
   - Supervised learning
   - Unsupervised learning
   - Reinforcement learning
3. Model
   - A good model
     - Overfitting and underfitting

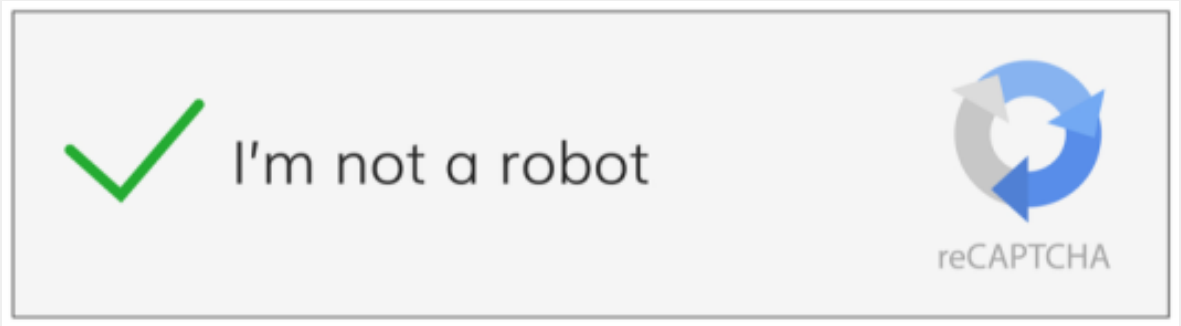
Introduction to Machine Learning

What is Machine Learning?

What is Machine Learning?

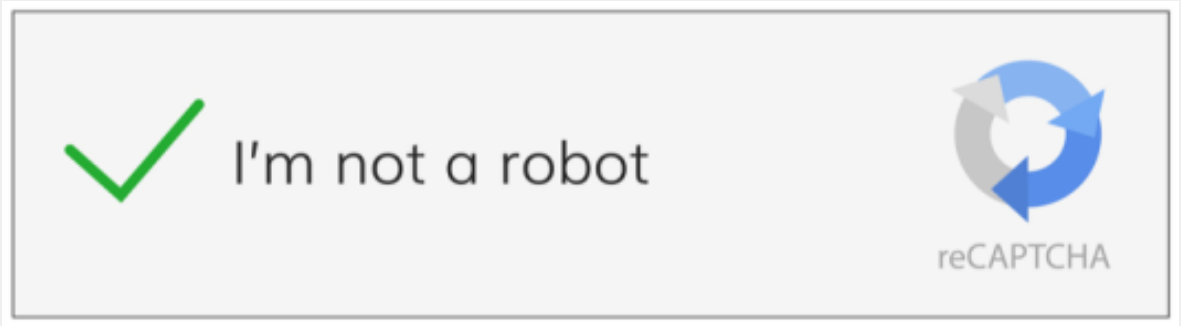


What is Machine Learning?



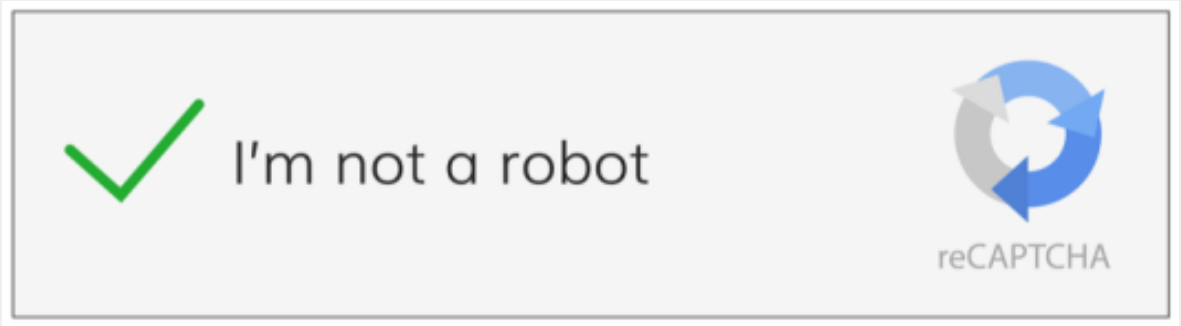
- This is Recaptcha.

What is Machine Learning?



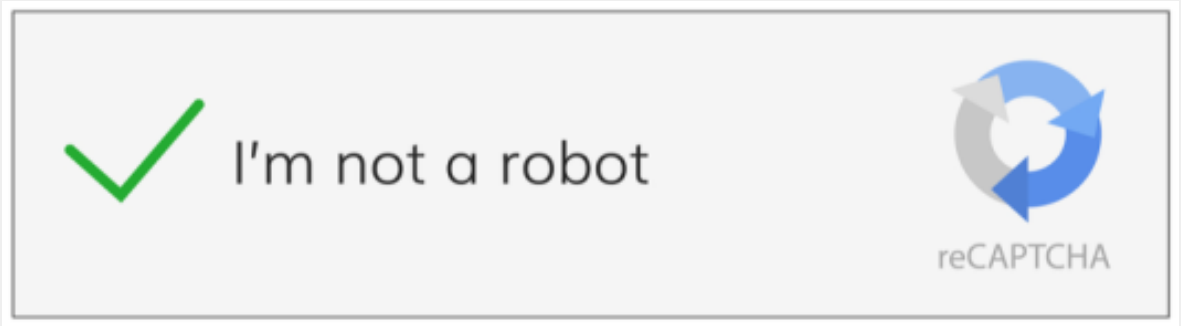
- This is Recaptcha.
  - Recaptcha helps stop millions of spam a day.

What is Machine Learning?



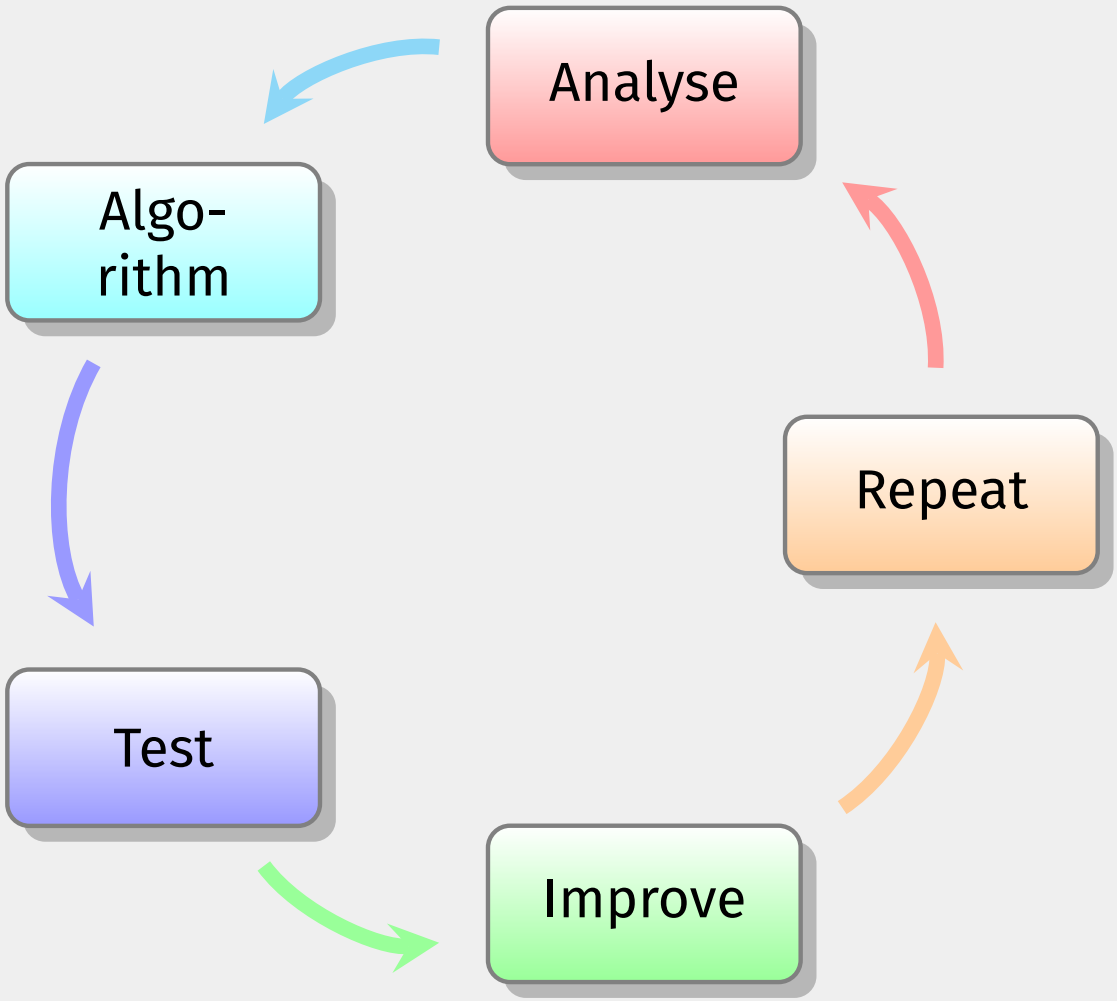
- This is Recaptcha.
  - Recaptcha helps stop millions of spam a day.
  - In some old days, we have to type Captcha texts to distinguish ourself from bots.

What is Machine Learning?

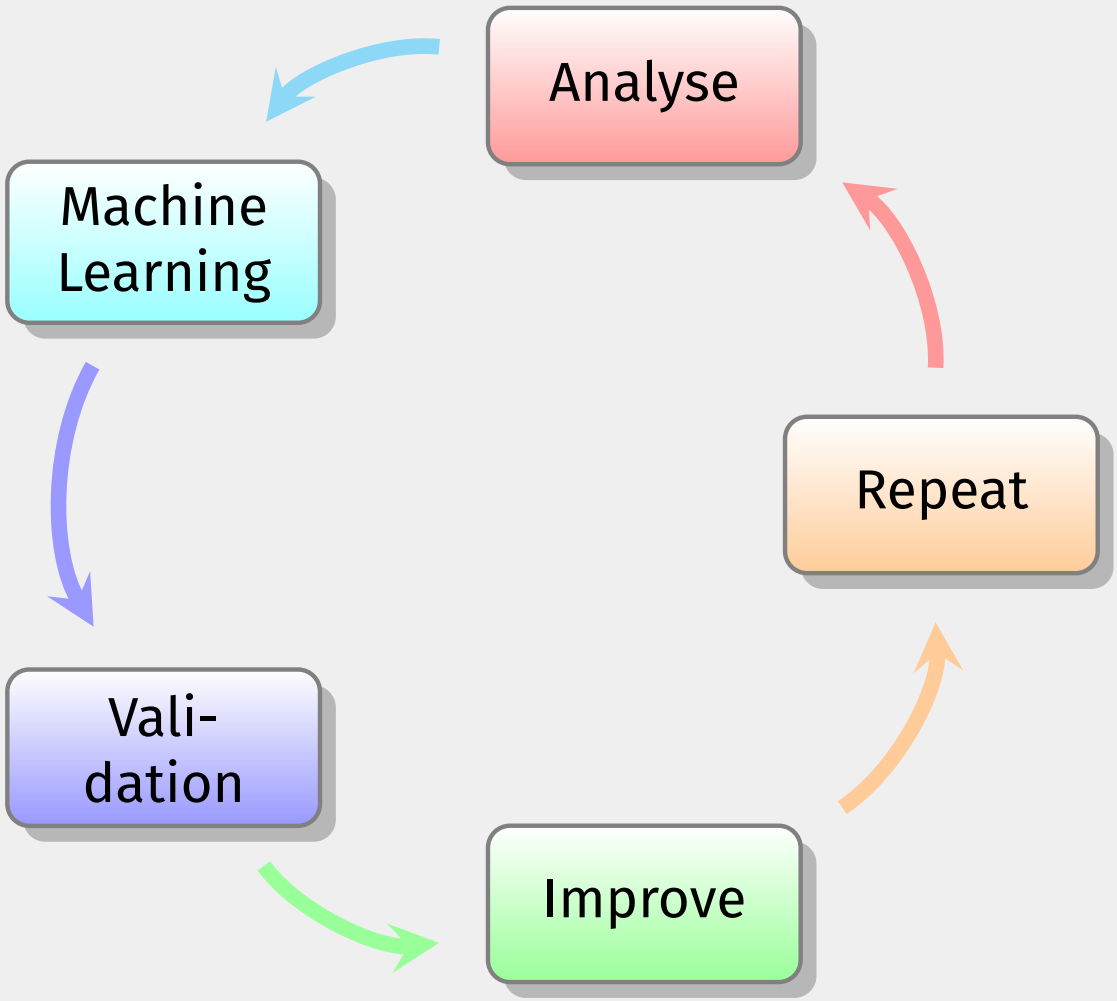


- This is Recaptcha.
  - Recaptcha helps stop millions of spam a day.
  - In some old days, we have to type Captcha texts to distinguish ourself from bots.
  - How is it possible that with a single click, an automated system can distinguish bots from humans?

Traditional programming approach



Machine learning approach



In other words...

Machine Learning

In other words...

Machine Learning

= Data + Data analysis algorithm

In other words...

Machine Learning

= Data + Data analysis algorithm

= Adapt to change

Machine Learning Problems

Types of Machine Learning problems

Types of Machine Learning problems

1. Supervised learning

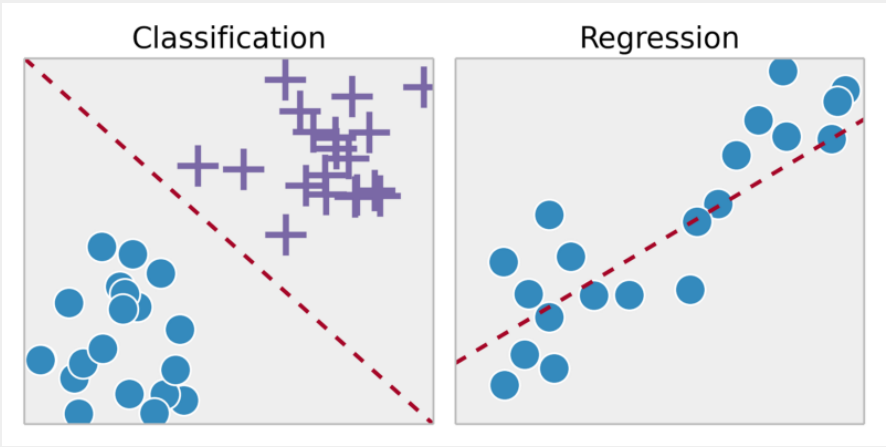
Types of Machine Learning problems

1. Supervised learning
2. Unsupervised learning

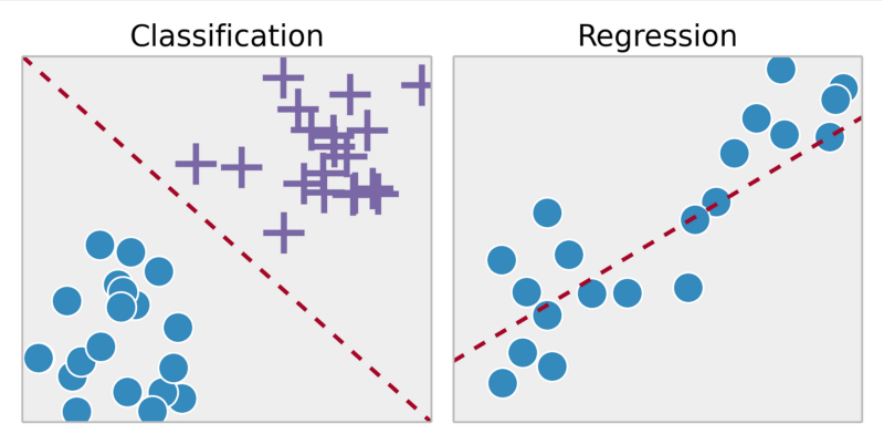
Types of Machine Learning problems

1. Supervised learning
2. Unsupervised learning
3. Reinforcement learning

Supervised learning

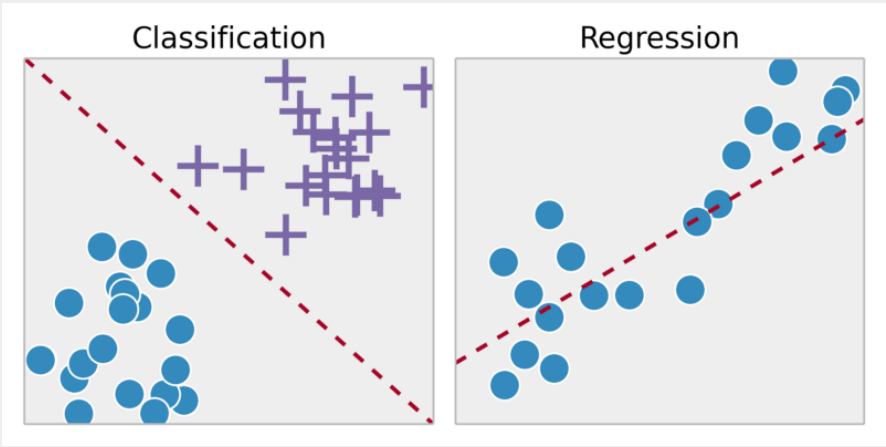


Supervised learning



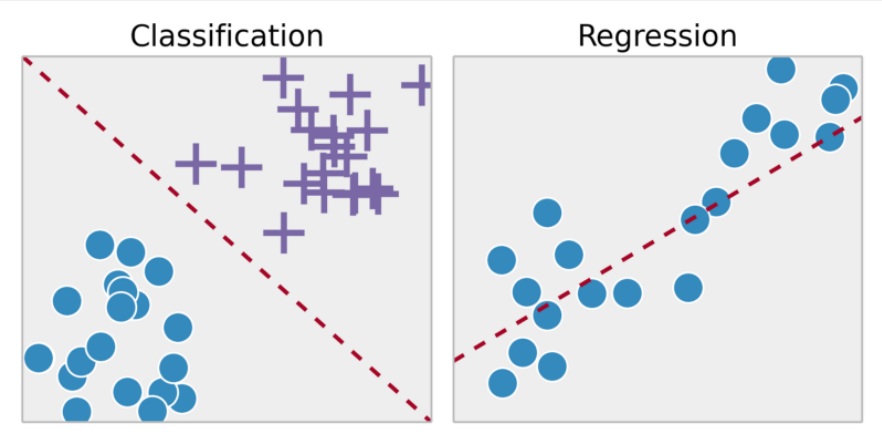
- Given a **training set** for the data, find a **model** to **generalise** well to **unseen** data.

Supervised learning



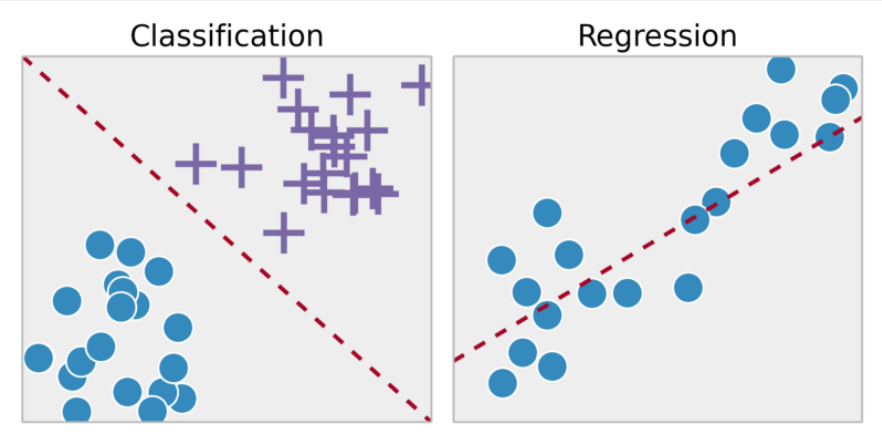
- Given a **training set** for the data, find a **model** to **generalise** well to **unseen** data.
- Two main supervised learning problems

Supervised learning



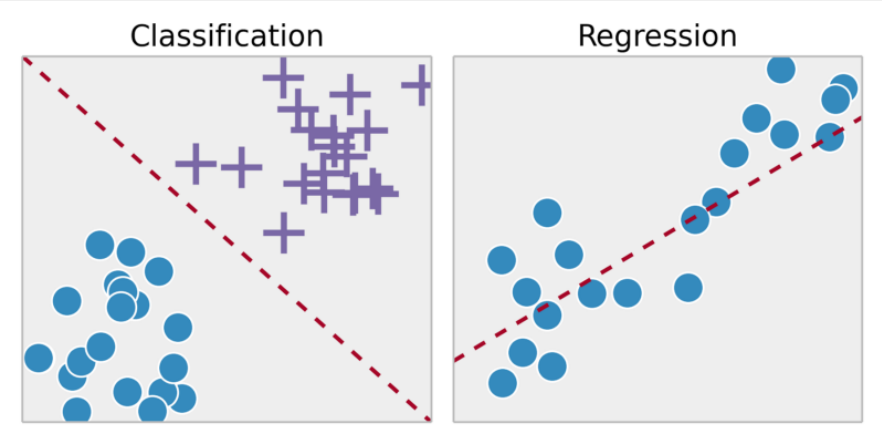
- Given a **training set** for the data, find a **model** to **generalise** well to **unseen** data.
- Two main supervised learning problems
  - Classification: On the discrete data

Supervised learning



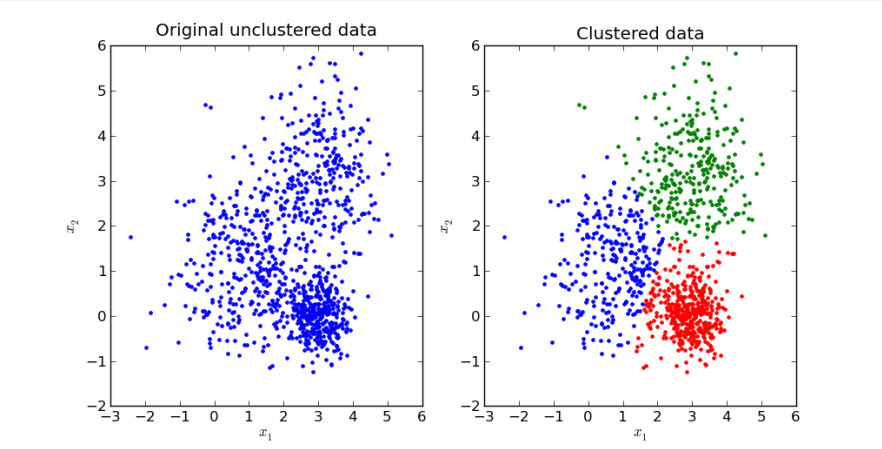
- Given a **training set** for the data, find a **model** to **generalise** well to **unseen** data.
- Two main supervised learning problems
  - Classification: On the discrete data
  - Regression: On the continuous data

Supervised learning

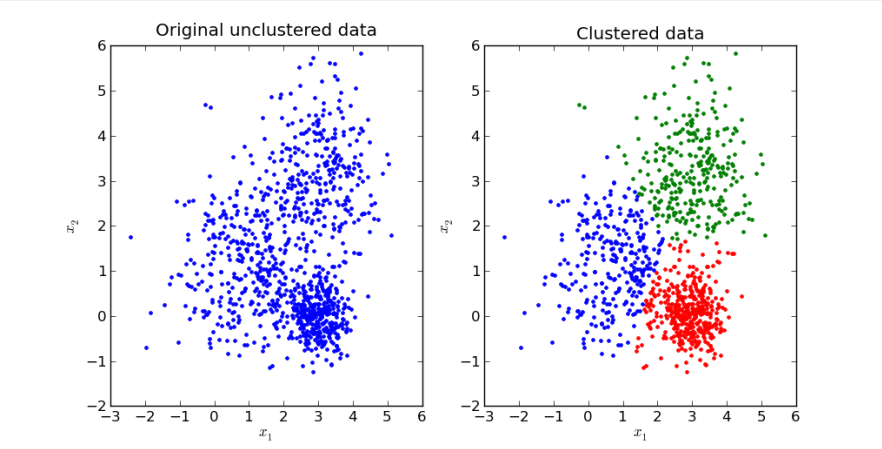


- Given a **training set** for the data, find a **model** to **generalise** well to **unseen** data.
- Two main supervised learning problems
  - Classification: On the discrete data
  - Regression: On the continuous data
- Example problems: Spam E-mail detection, Facial recognition

Unsupervised learning

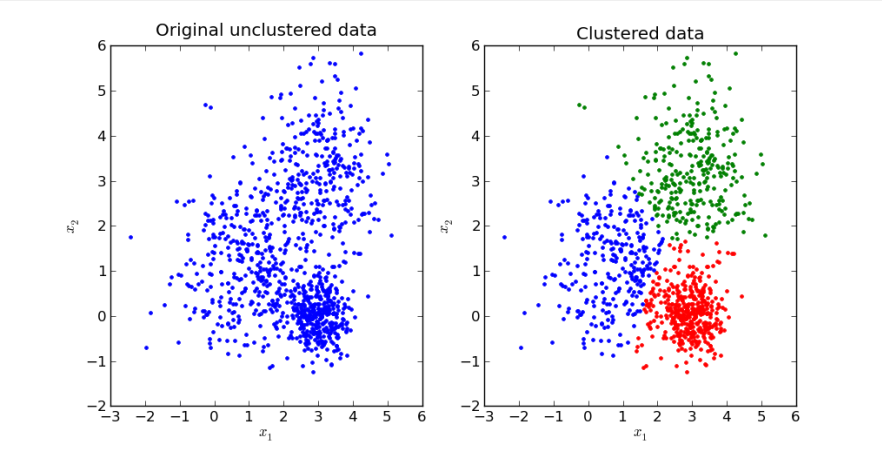


Unsupervised learning



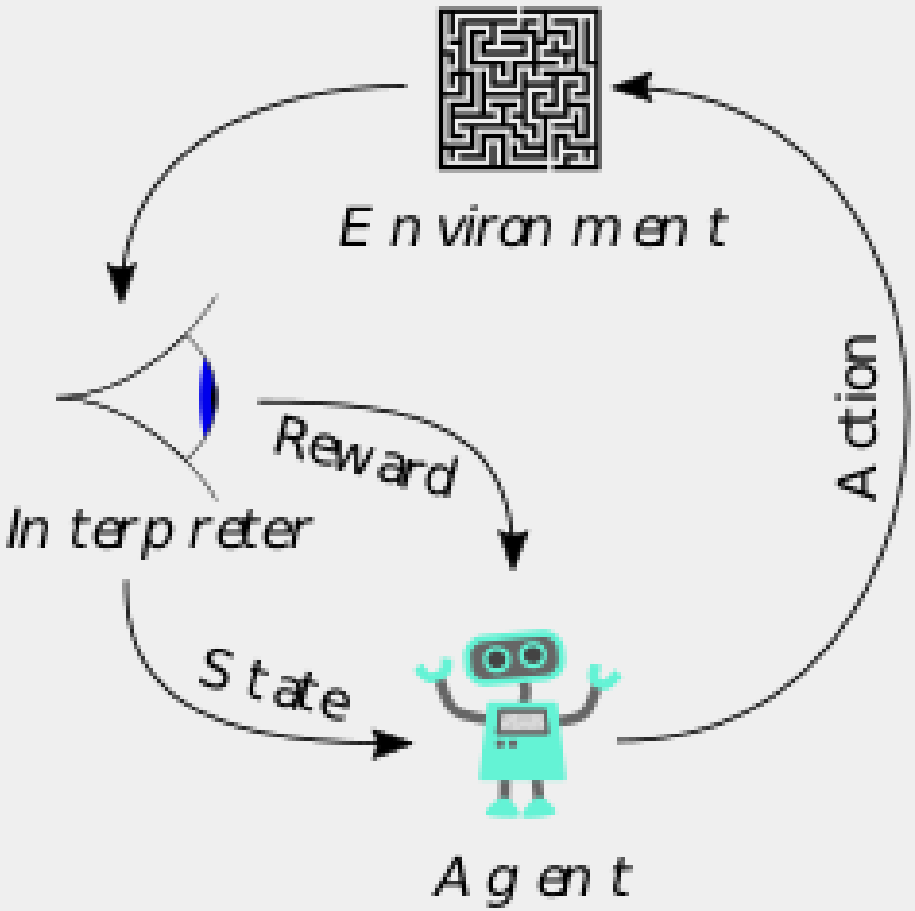
- Discover **hidden** structure in **non-labelled** data.

Unsupervised learning



- Discover **hidden** structure in **non-labelled** data.
- Example: Clustering, Generative models

Reinforcement learning



Model

Model

- A result of the combination between...

Model

- A result of the combination between...
  - a **method** to recognise the data, and

Model

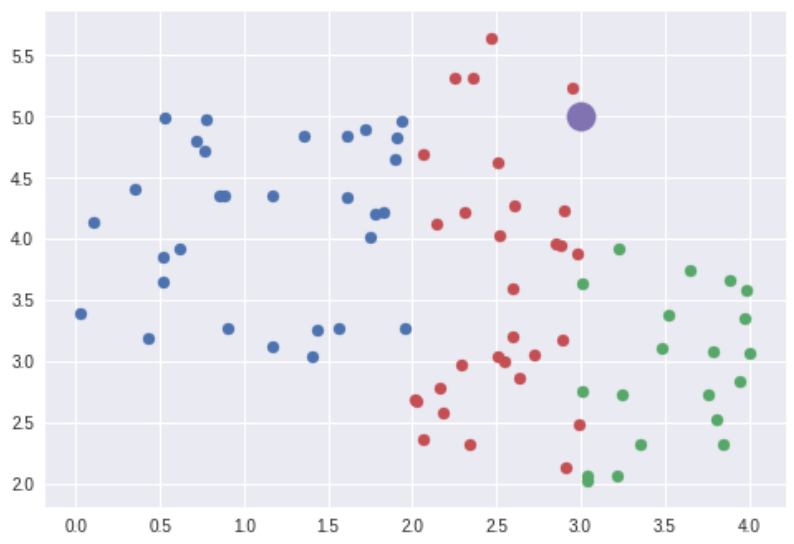
- A result of the combination between...
  - a **method** to recognise the data, and
  - **sample datas** for such the method

Model

Data

Method

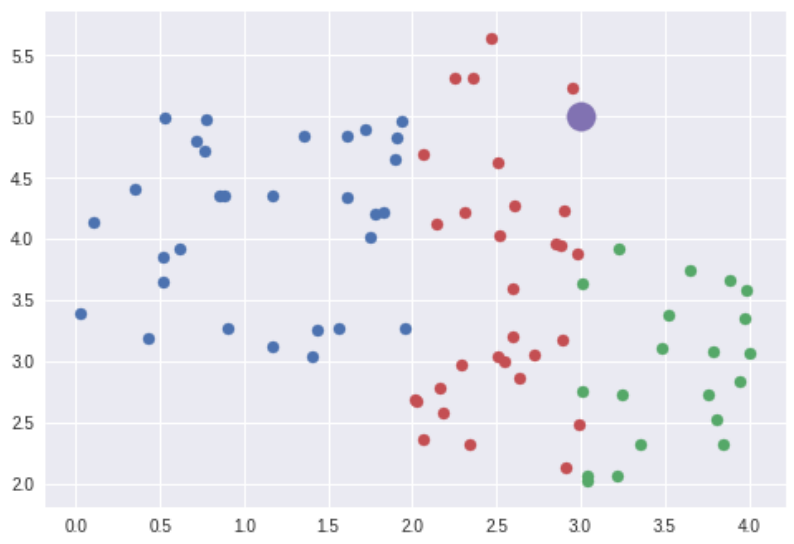
Model



Data

Method

Model



Data

Determine which group should the purple dot be in (red/green/blue) by **checking the colour of its nearest dot.**

Method

Beginning with our first model

Beginning with our first model

- We're going to write our **first own** machine learning algorithm called *k*-**Nearest Neighbour** (*k*NN)

Beginning with our first model

- We're going to write our **first own** machine learning algorithm called *k*-**Nearest Neighbour** (*k*NN)
  - *k*-NN is known to be very simple, with its concept as

Beginning with our first model

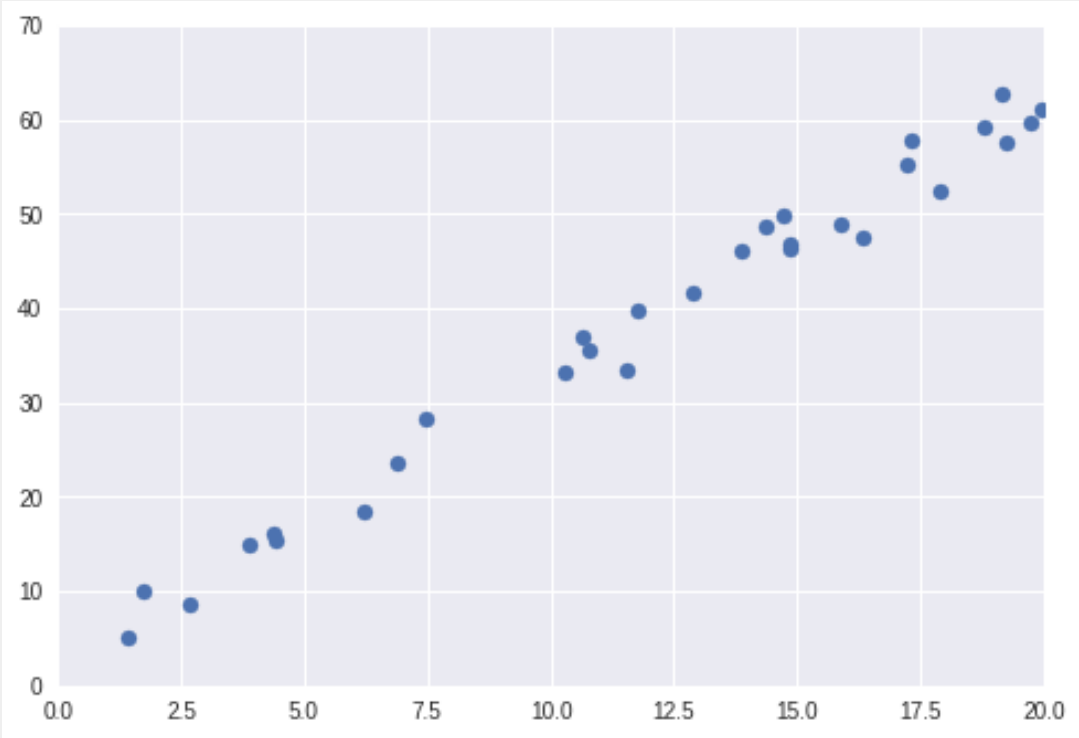
- We're going to write our **first own** machine learning algorithm called *k*-**Nearest Neighbour** (*k*NN)
  - *k*-NN is known to be very simple, with its concept as

_k_-NN algorithm

To classify label of a data point, get *k* nearest data points to the data point, and select the major label among those data points.

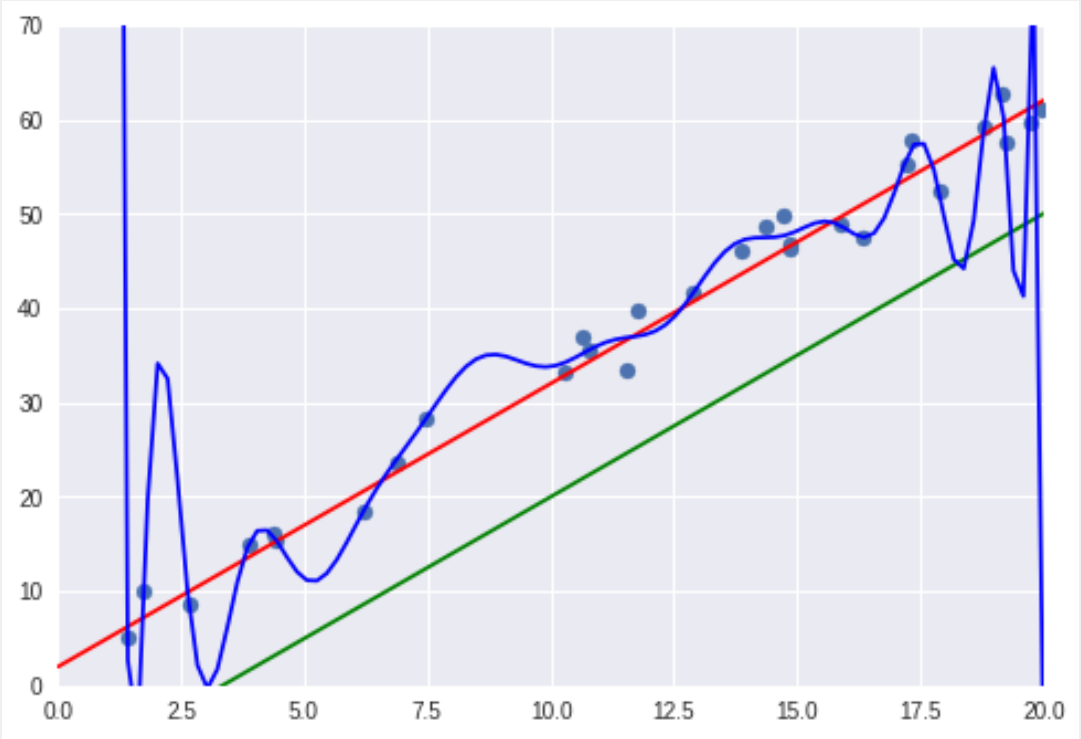
Good model?

Good model



How should we *draw* the line to predict this data?

Good model



Blue, red, or green line?

Overfitting and underfitting

Overfitting and underfitting

1. Underfitting

Overfitting and underfitting

1. Underfitting
   - Our model **fails to know the data's trends**

Overfitting and underfitting

1. Underfitting
   - Our model **fails to know the data's trends**
   - Resulting in failure to predict further data

Overfitting and underfitting

1. Underfitting
   - Our model **fails to know the data's trends**
   - Resulting in failure to predict further data
2. Overfitting

Overfitting and underfitting

1. Underfitting
   - Our model **fails to know the data's trends**
   - Resulting in failure to predict further data
2. Overfitting
   - Our model **memorise instead of generalise**

Overfitting and underfitting

1. Underfitting
   - Our model **fails to know the data's trends**
   - Resulting in failure to predict further data
2. Overfitting
   - Our model **memorise instead of generalise**
   - Resulting in failure to catch the trend

Good model must **generalise**