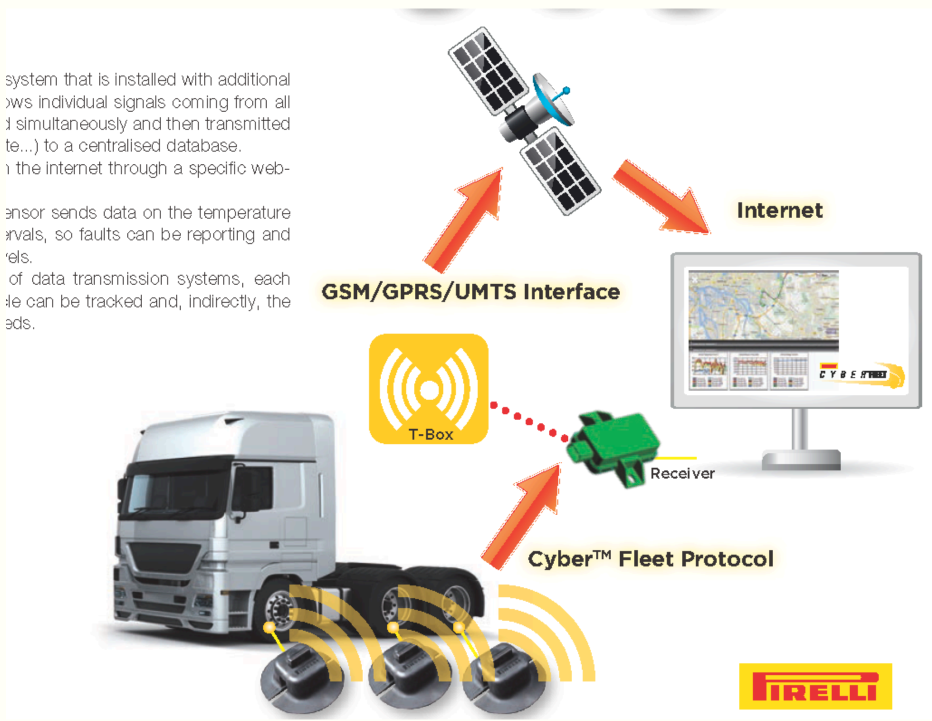
ENGG1811 Computing for Engineers

Week 9C
Machine Learning

We are collecting a lot of data

- Example: Tyre pressure is important for vehicle fuel efficiency
  - A company puts sensors inside the tyres of its fleet and transmits the pressure data via the Internet



Data versus information

- Data: Raw records of facts / measurements
- Information: **Patterns** in the underlying data
  - We also call the pattern a **model**

- Example:
  - In lab07, you worked on the data on the extent of sea ice
    - Data: raw measurements
    - Information: The extent of sea ice in the 1ast 10 years was lowest in record, …
  - What are needed to feel fulfilled in life?
    - Data: People around you, life stories, biographies
    - Information to discover: What are the attitudes, mentalities, ways of thinking, actions etc. that can lead to fulfillment in life?

Machine learning

- Given the huge amount of data generated, can data analysis be automated?

- Machine learning
  - Algorithms to automatically extract patterns from the data

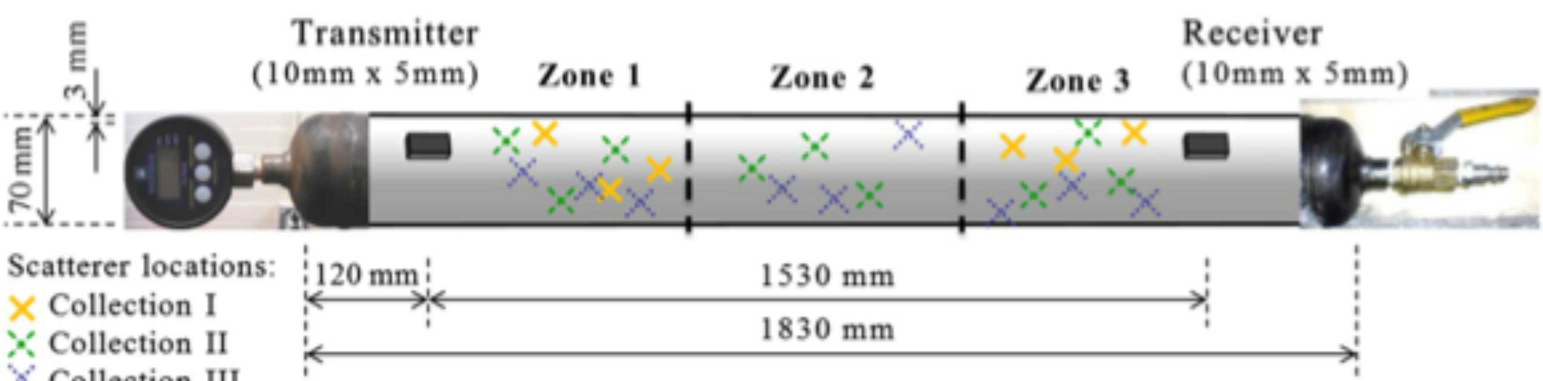
- Let us see what machine learning can do

Damage detection

Toward Data-Driven Structural Health Monitoring: Application of Machine Learning and Signal Processing to Damage Detection

Yujie Ying[1]; James H. Garrett Jr., F.ASCE[2]; Irving J. Oppenheim, M.ASCE[3]; Lucio Soibelman, F.ASCE[4]; Joel B. Harley[5]; Jun Shi[6]; and Yuanwei Jin[7]

JOURNAL OF COMPUTING IN CIVIL ENGINEERING © ASCE / NOVEMBER/DECEMBER 2013 / 667



(a)

Detection problem

- Detection problem: A yes or no answer
  - Is it damaged? Is it not damaged?
  - Is it present? Is it absent?

164 IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 60, NO. 1, JANUARY 2013

Machine Learning-Based Method for Personalized and Cost-Effective Detection of Alzheimer's Disease

Javier Escudero*, *Member, IEEE*, Emmanuel Ifeachor, *Member, IEEE*, John P. Zajicek, Colin Green, James Shearer, and Stephen Pearson, for the Alzheimer's Disease Neuroimaging Initiative

Damage classification (1)

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A
MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

The application of machine learning to structural health monitoring

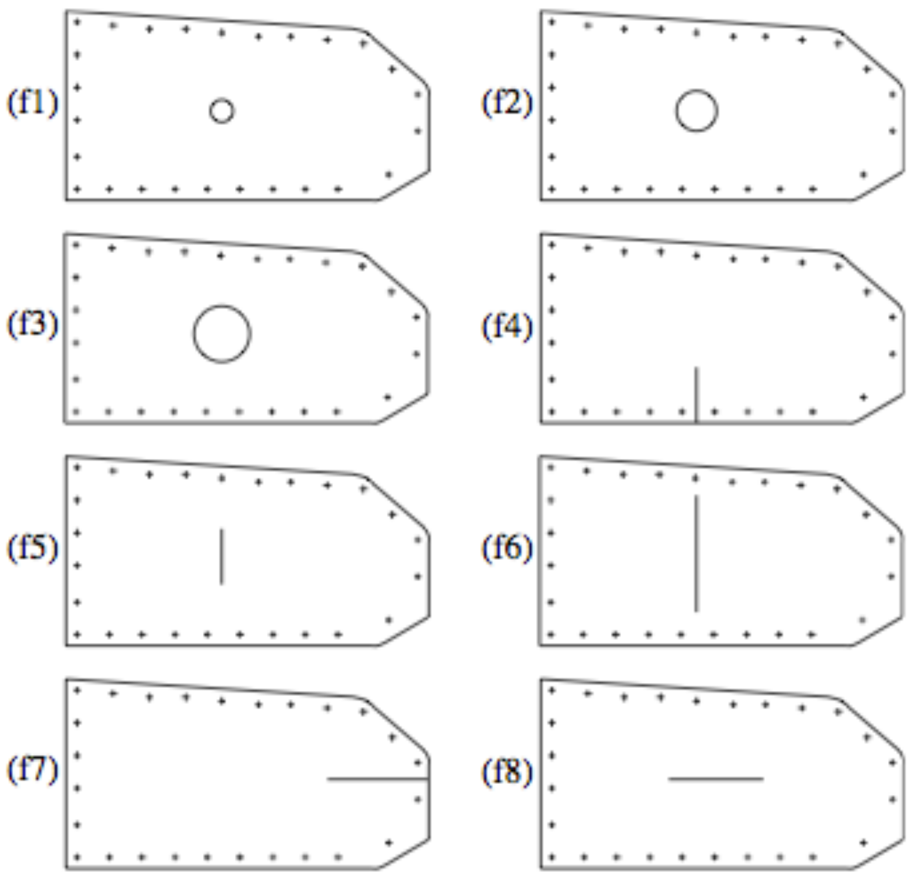
Keith Worden and Graeme Manson

Phil. Trans. R. Soc. A 2007 **365**, doi: 10.1098/rsta.2006.1938, published 15 February 2007

Damage classification (2)

Figure 2. Gnat aircraft and acquisition system.

Damage classification (3)



- These are the possible locations that the wing may be damaged
- The aim is to determine from the measured data which of these possibilities it is

Classification problem

- There are a number of classes and the problem is to determine which class the data is coming from
  - Example: The wing is damaged and each class is a location at which the wing is damaged.
  - The classification problem is to determine which class it is. This is the same as determining where the damage is located.
  - Detection problem is a special case of classification problem with two classes. For example, the damage detection problem is equivalent to a classification problem with 2 classes: damaged or not damaged.

Some other classification problems

- Google driverless cars
  - Classify road sign. Classes: Give way, Stop, speed limit, …
  - Classify junction type. Classes: Traffic lights, roundabout, …

- http://googleblog.blogspot.com.au/2014/05/just-press-go-designing-self-driving.html

http://www.gereports.com/post/123572457345/deep-machine-learning-ge-and-bp-will-connect/

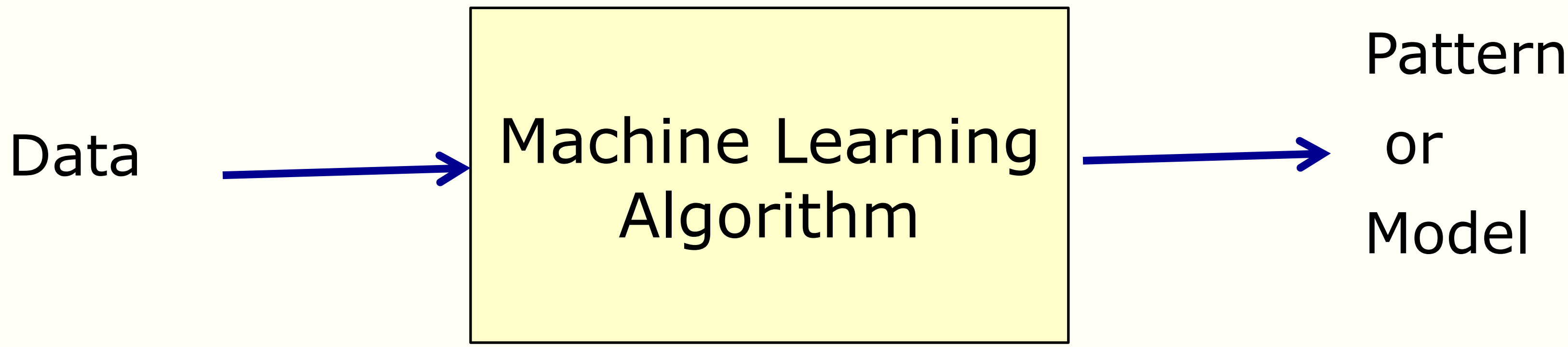
Machine learning for prediction

The industry estimates that operators lose as much as $3 million in revenue per week when a well goes out of commission.

No wonder all kinds of companies operating in the energy space have started looking for ways to reduce unplanned downtime as close to zero as possible. “Telling a customer what to fix after it has failed is relatively easy,” said Bob Judge, director of product management at GE Oil & Gas. “Telling them to fix something before it costs them money is the magic.”

Our agenda on machine learning

- We will not examine the technical details of machine learning algorithms
- We take a black box approach



- We will give you some intuition of how machines can learn
- We will start off with a small example so that we can visualise what’s happening

Fault detection – Problem setting

- You want to detect whether a device is faulty or not
- You have conducted a lot of measurements on both working and faulty devices
- For each device:
  - You measure two physical properties of the device. We denote these two physical properties by x and y
  - A label which says whether the measurements come from a working or faulty device
- It's not important to know what type of physical measurements x and y are in this discussion

Fault detection – the data

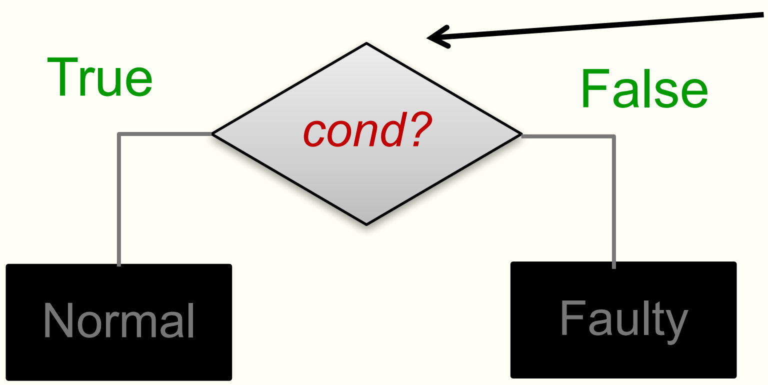
x	y	label
5.1	3.5	**Normal**
4.9	3.0	**Normal**
…	…	**…**
5.0	3.3	**Normal**
7.0	3.2	**Faulty**
6.2	3.4	**Faulty**
5.9	3.0	**Faulty**

50 sets of measurements

100 sets of measurements

Fault detection = Finding Boolean condition

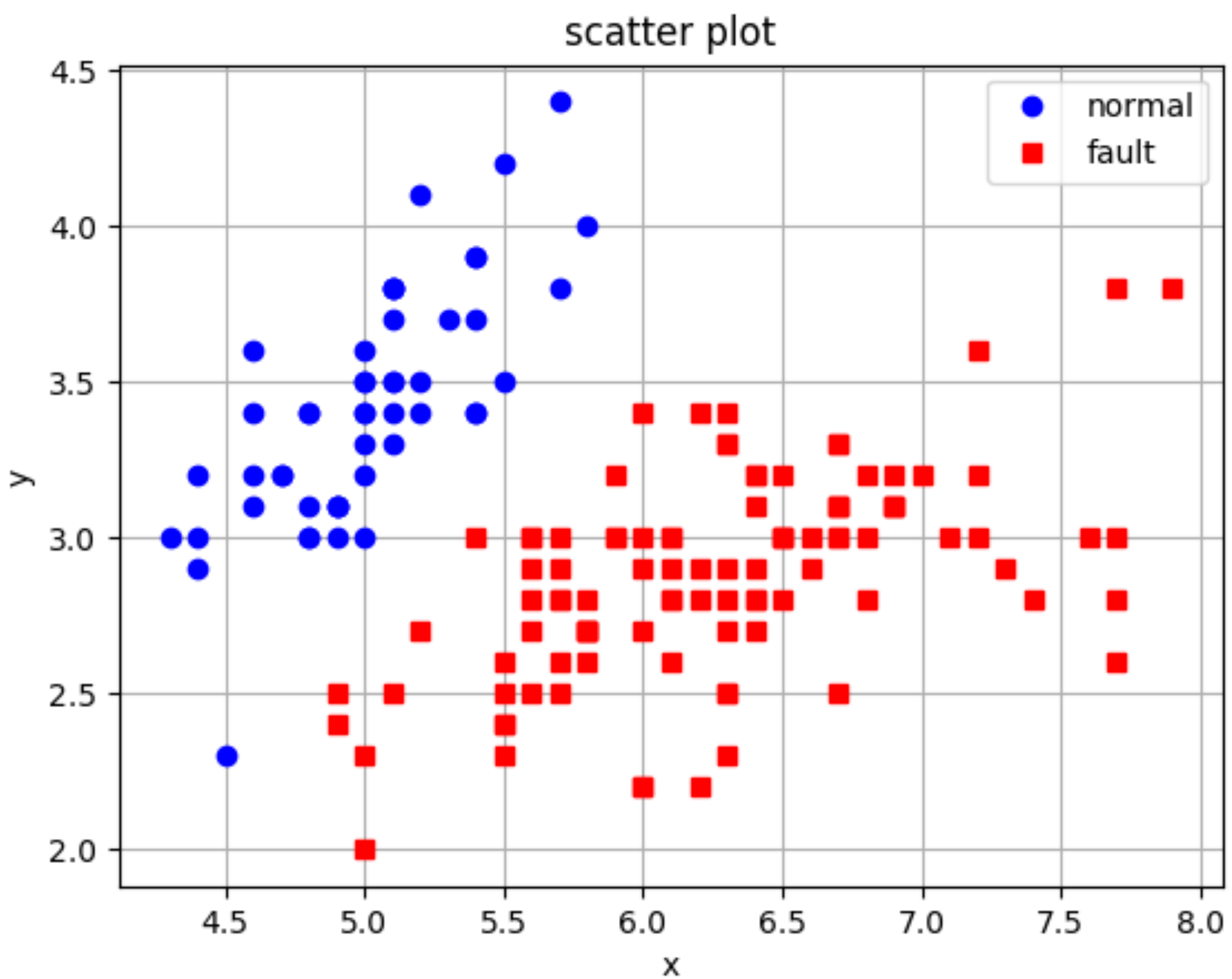
x	y	label
5.1	3.5	**Normal**
4.9	3.0	**Normal**
…	…	**…**
5.0	3.3	**Normal**
7.0	3.2	**Faulty**
6.2	3.4	**Faulty**
5.9	3.0	**Faulty**



This condition depends on x, y

Plotting the data

What pattern do you observe?

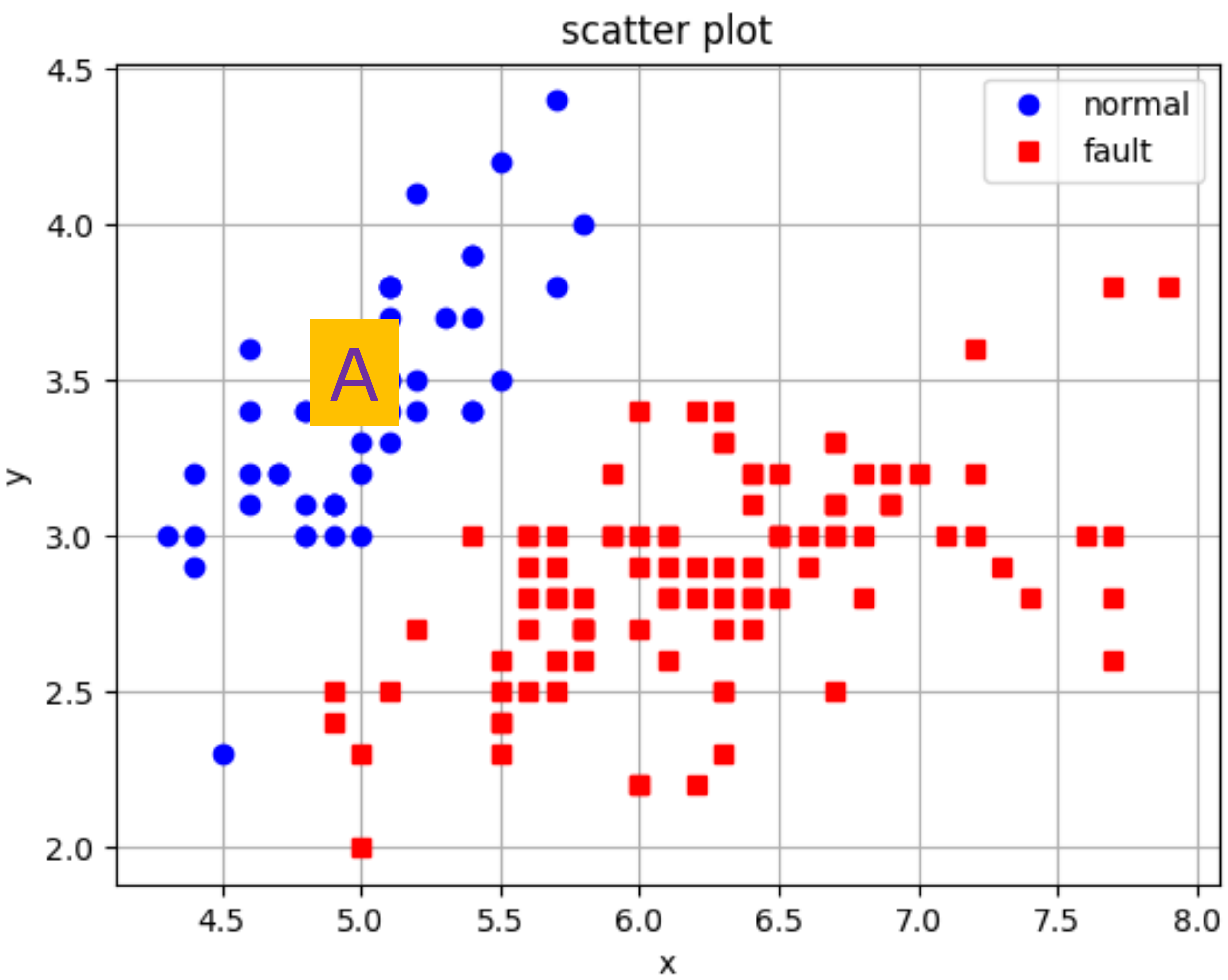


Python file:
plot_data.py

Data files:
data_values.npy
data_labels.npy

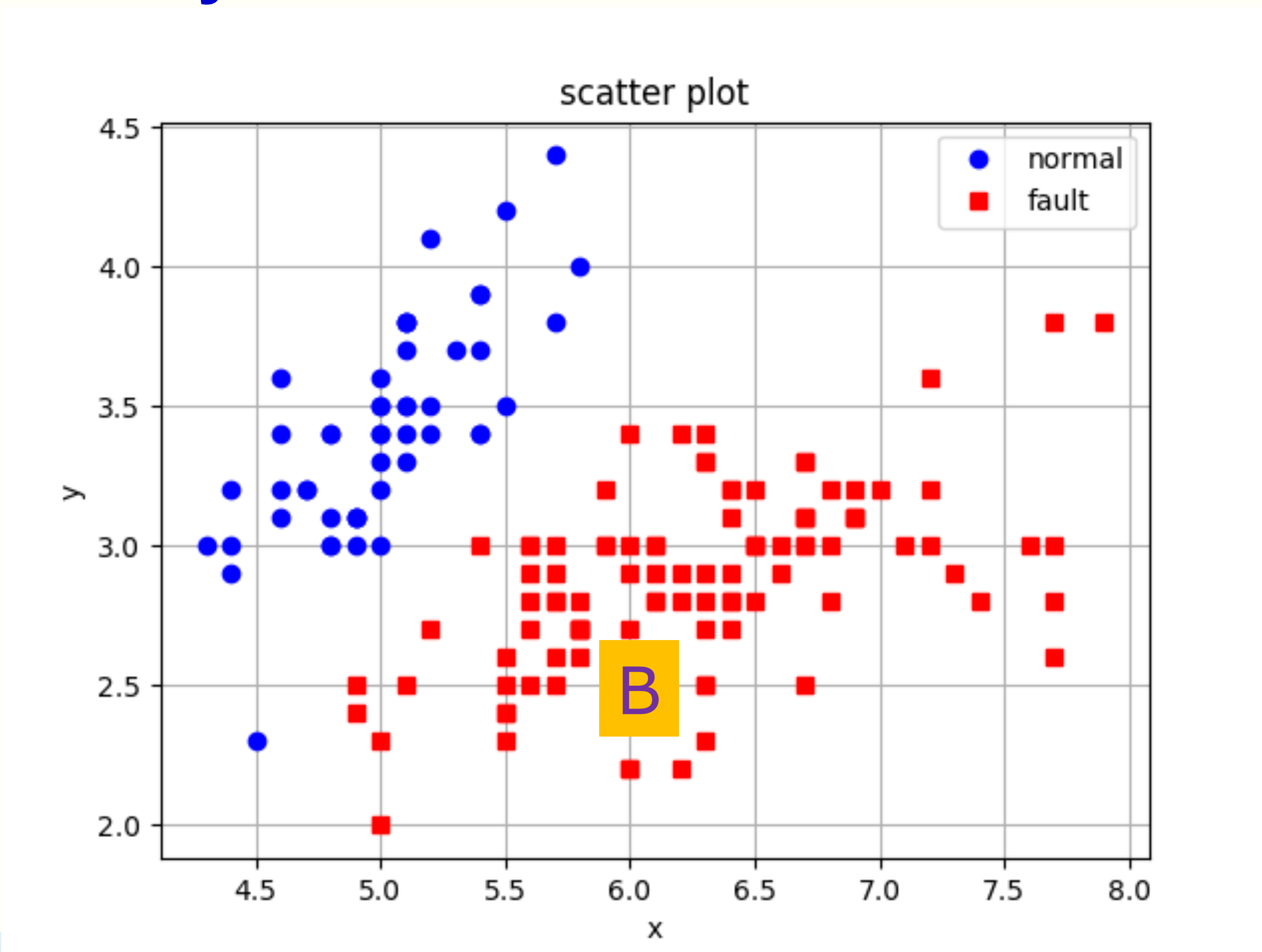
Question: Normal or Faulty?

Question: If you take a new device and measure its x and y. You find that x = 5.0 and y = 3.5. This is marked as Point A in the diagram. If you are to make a guess whether this device is faulty or not, what is your guess?



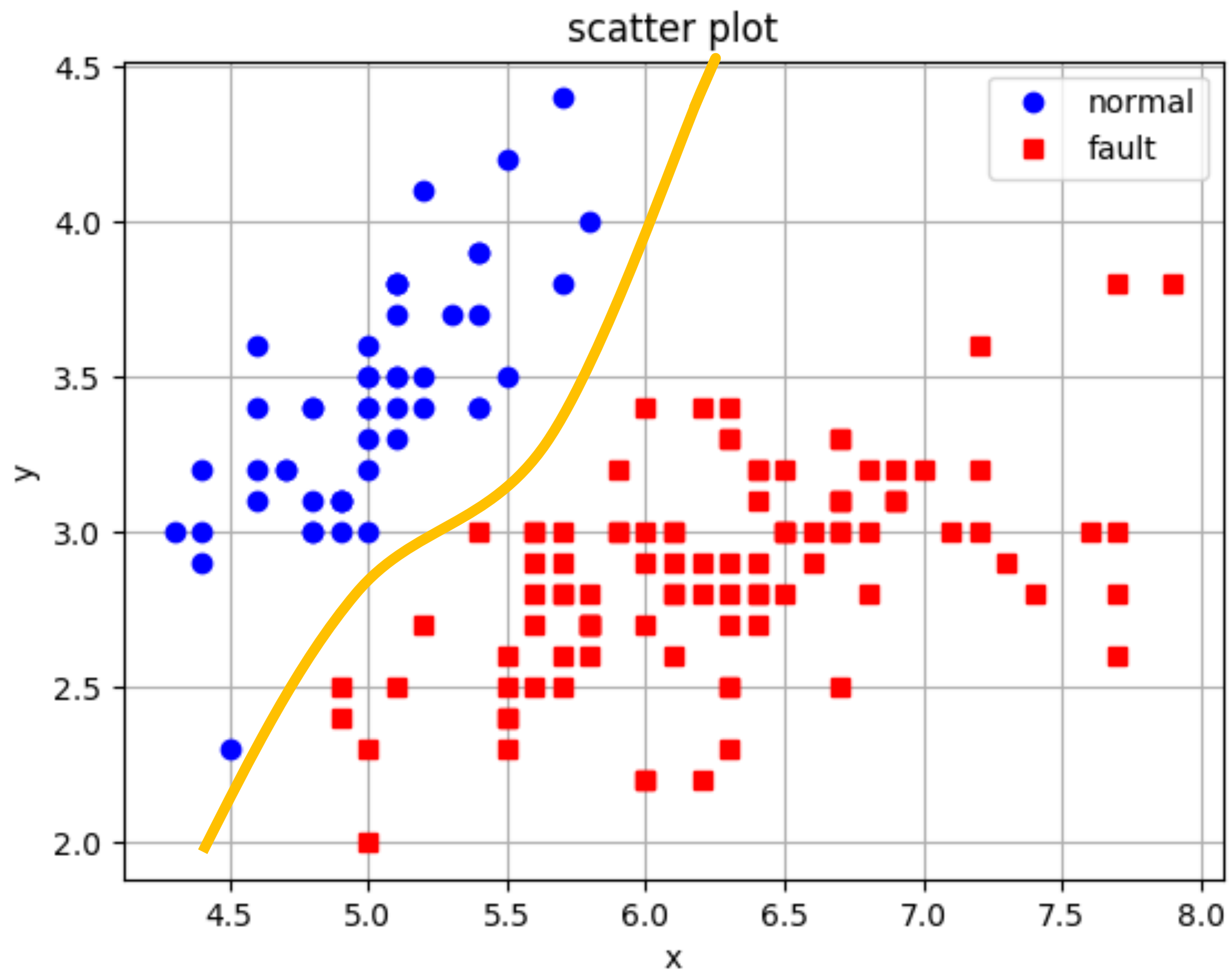
Question: Normal or Faulty?

Question: If you take another device and measure its x and y. You find that x = 6 and y = 2.5. This is marked as Point B in the diagram. If you are to make a guess whether this device is faulty or not, what is your guess?



Question: How to separate?

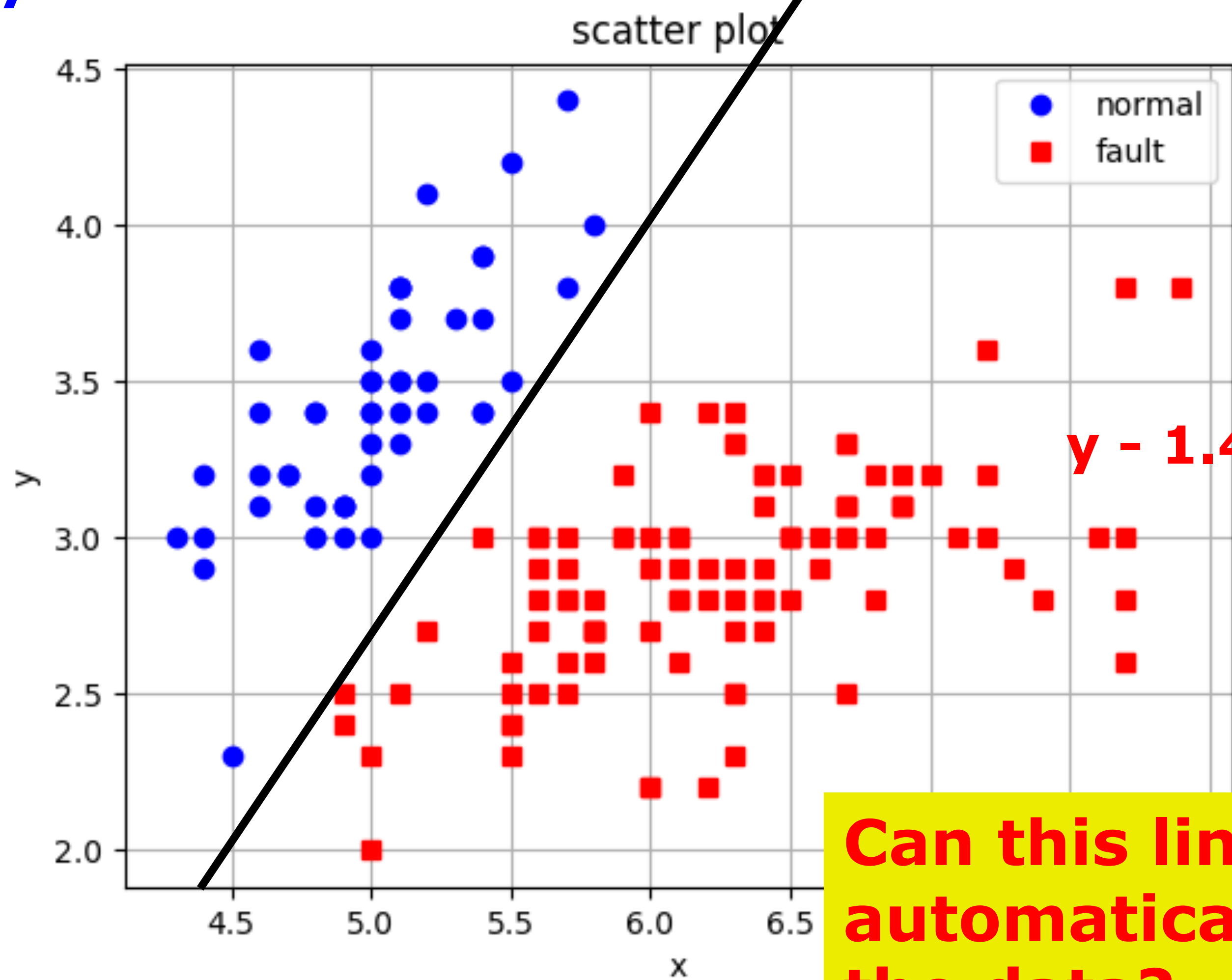
Question: Can you suggest a method to separate normal devices from the faulty devices?



Manual detection

$y - 1.4x + 4.3 = 0$

$y - 1.4x + 4.3 > 0$

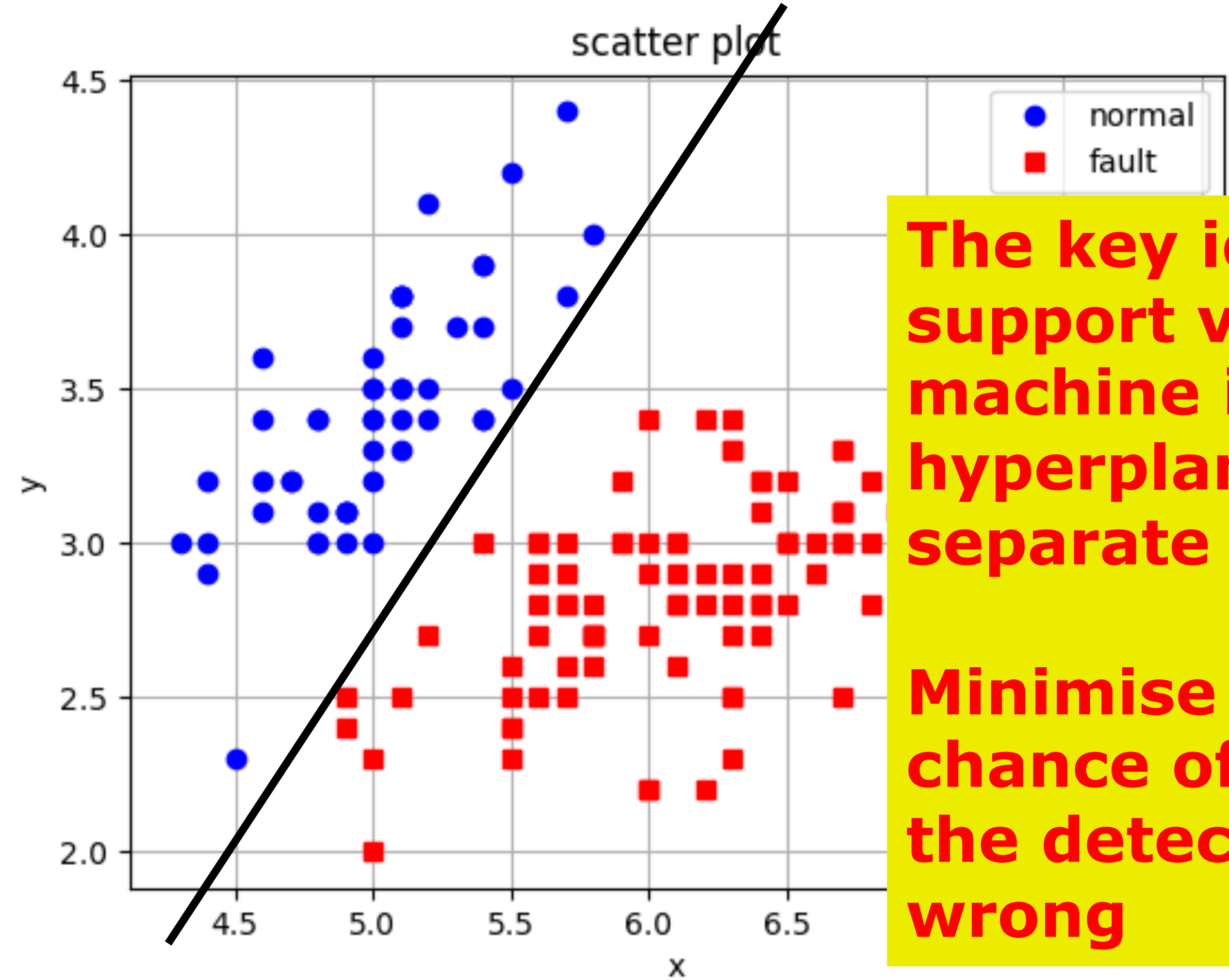


$y - 1.4x + 4.3 < 0$

Can this line be found automatically from the data?

Machine learning method: Support vector machines

Python file: classify_2classes.py



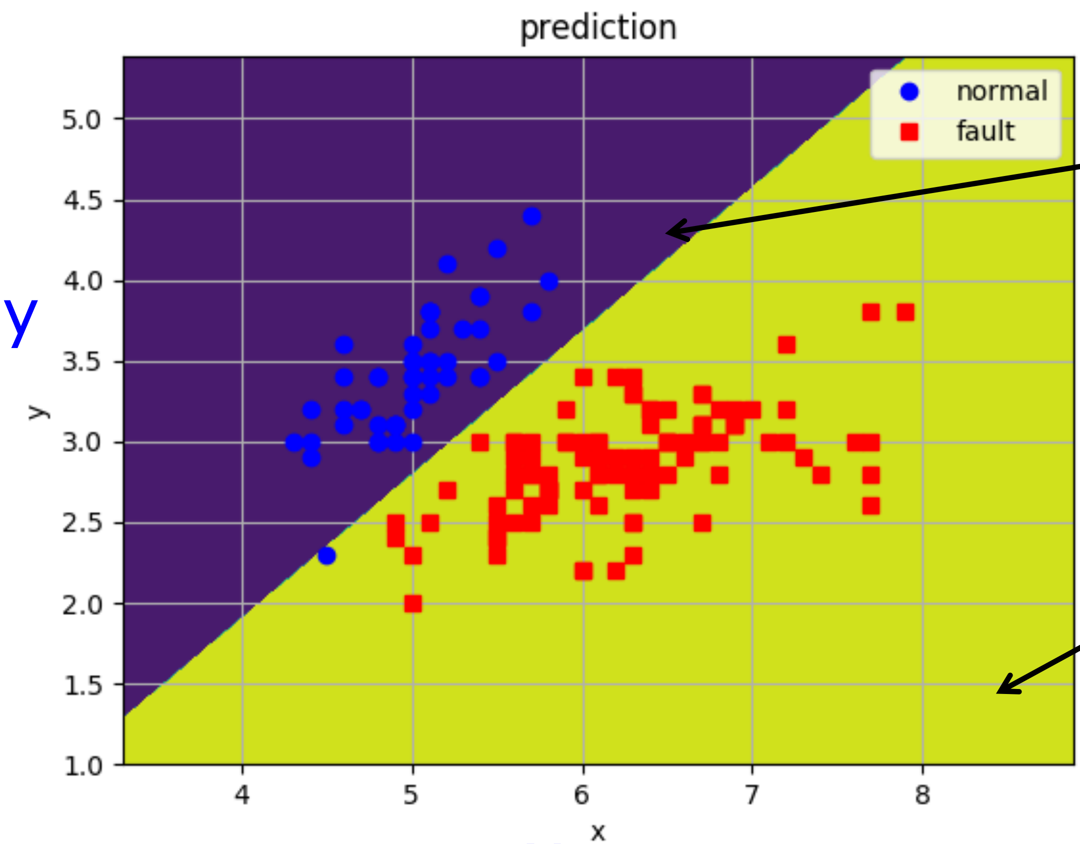
The key idea of support vector machine is to find hyperplane to separate the data

Minimise the chance of getting the detection wrong

Prediction from Support Vector Machines

Python file: classify_2classes.py

- For each pair of (x,y), the model from support vector machine makes a prediction on whether the device is normal or faulty



y

x

If (x,y) is in the purple region, the prediction is normal

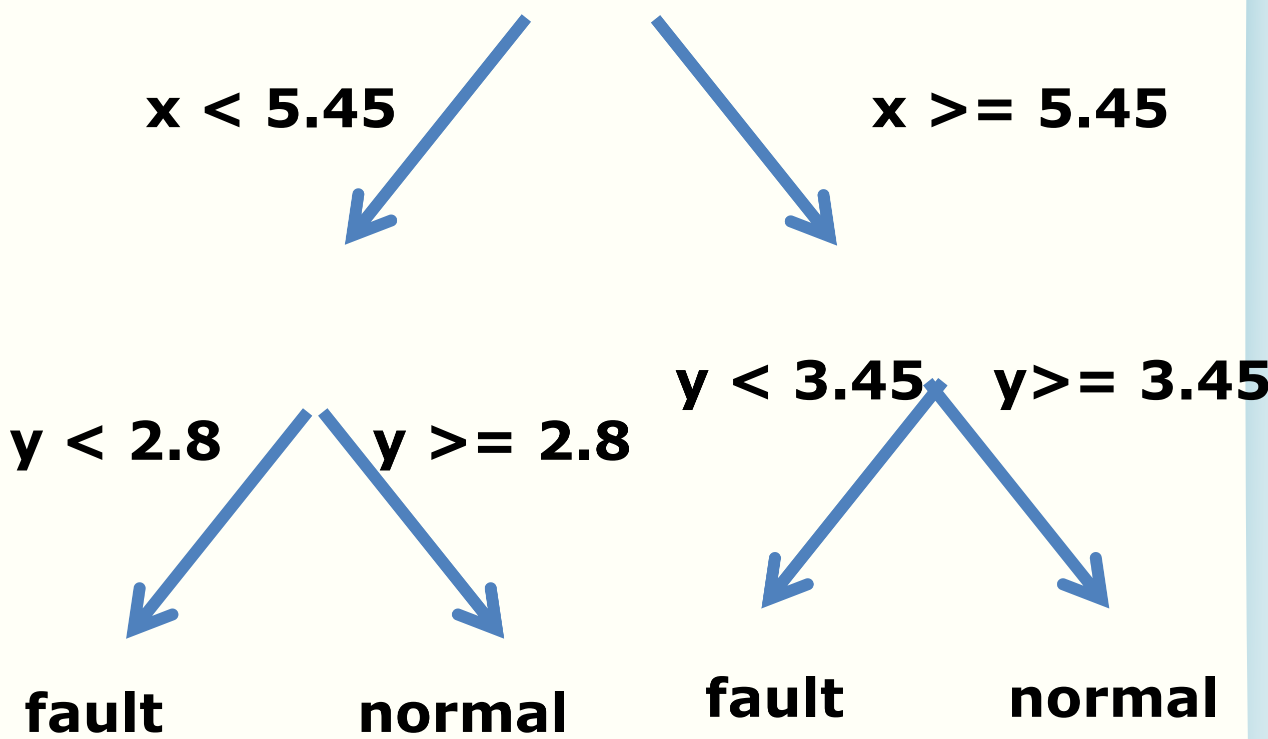
If (x,y) is in the green region, the prediction is faulty

Machine learning method: Classification tree

Python file: classify_2classes.py

- This method comes out with a set of rules automatically

```
if x >= 5.45:
   if y >= 3.45:
     # pred: normal
    else:
      # pred: faulty
else:
    if y >= 2.8:
      # pred: normal
    else:
      # pred: faulty
```



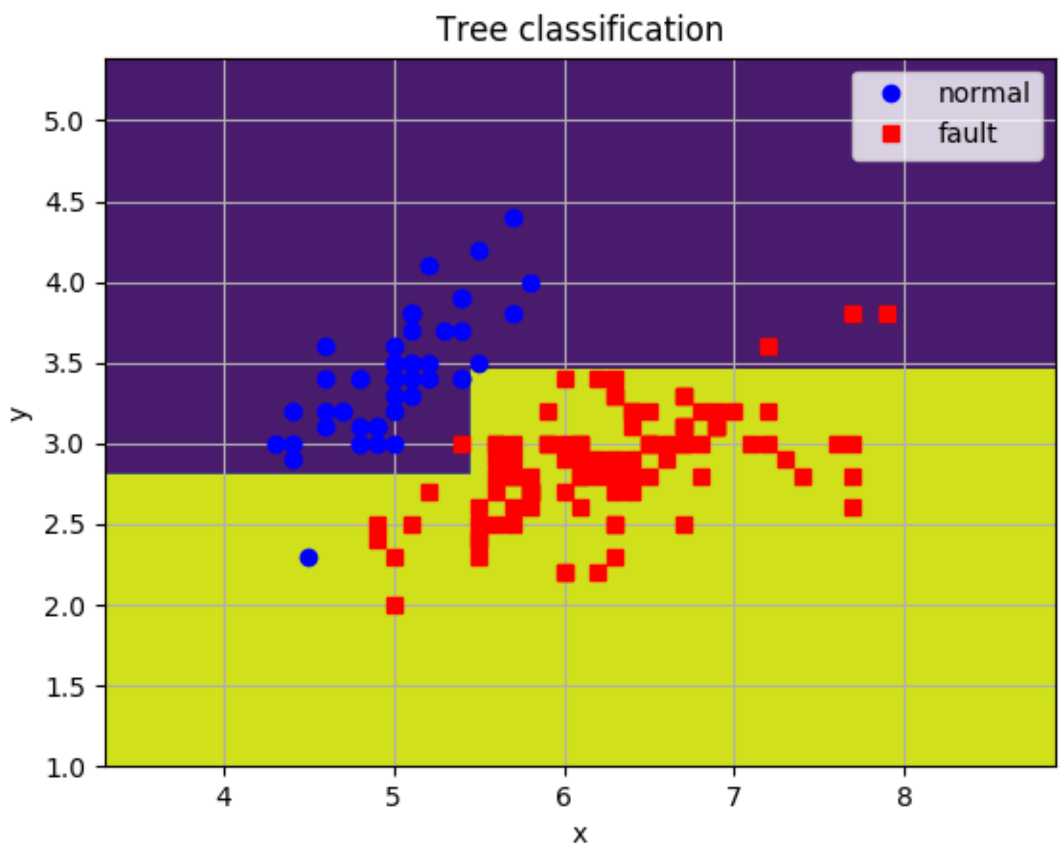
- Nested for-loop with a depth of 2

Prediction from decision tree

Python file: classify_2classes.py

- For a tree with a depth of 2

```
if x >= 5.45:
    if y >= 3.45:
        # pred: normal
    else:
        # pred: faulty
else:
    if y >= 2.8:
        # pred: normal
    else:
        # pred: faulty
```



Compare the two methods

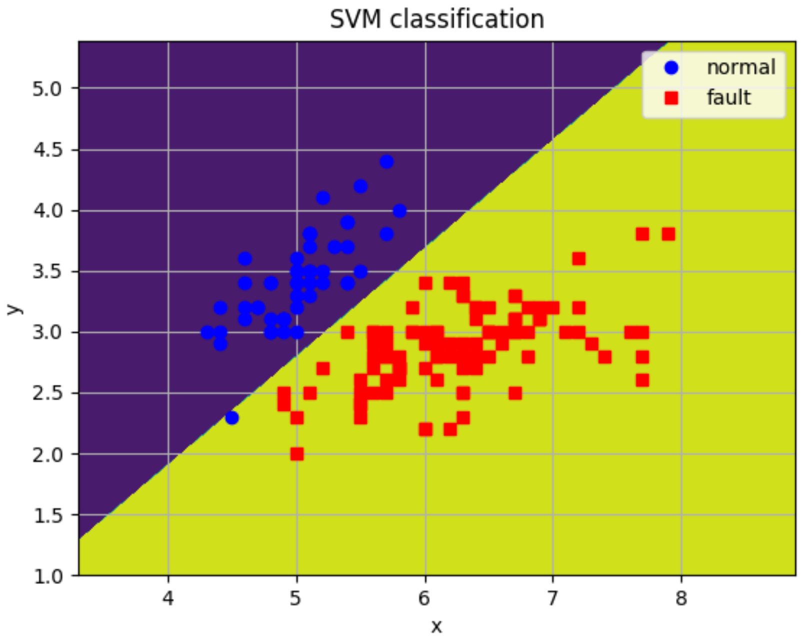
- Python file classify_2classes.py

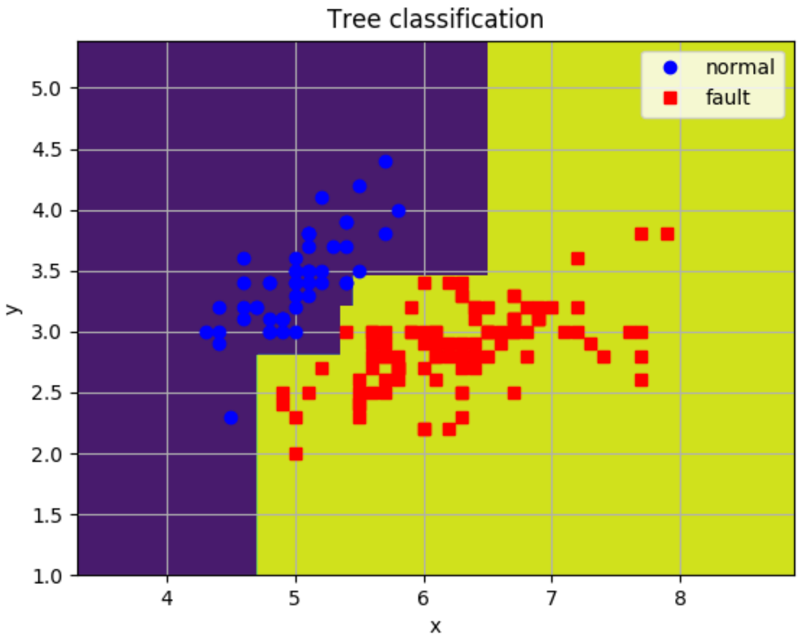
- Using SVM and classification trees
  - Use the maximum possible depth

- Display the classification regions for the two methods
  - Key: Purple means normal, green means faulty
  - Are the classification regions the same for the two methods?

Oops! Two algorithms say different things

Python file: classify_2classes.py





- Can we tell which one is better?
- How do we tell?
- We will answer these questions for classification problem

Fault classification

- The same set of data as before
- The device can have two types of faults, which we call fault1 and fault2
- Every pair of (p,q) measured is given a label, which can be
  - Normal
  - Fault1
  - Fault2

Fault classification – the data

x	y	Label
5.1	3.5	**Normal**
4.9	3.0	**Normal**
…	…	**…**
5.0	3.3	**Normal**
7.0	3.2	**Fault1**
5.7	2.8	**Fault1**
6.3	3.3	**Fault2**
6.2	3.4	**Fault2**
5.9	3.0	**Fault2**

50 sets of measurements

50 sets of measurements

50 sets of measurements

What is learning?

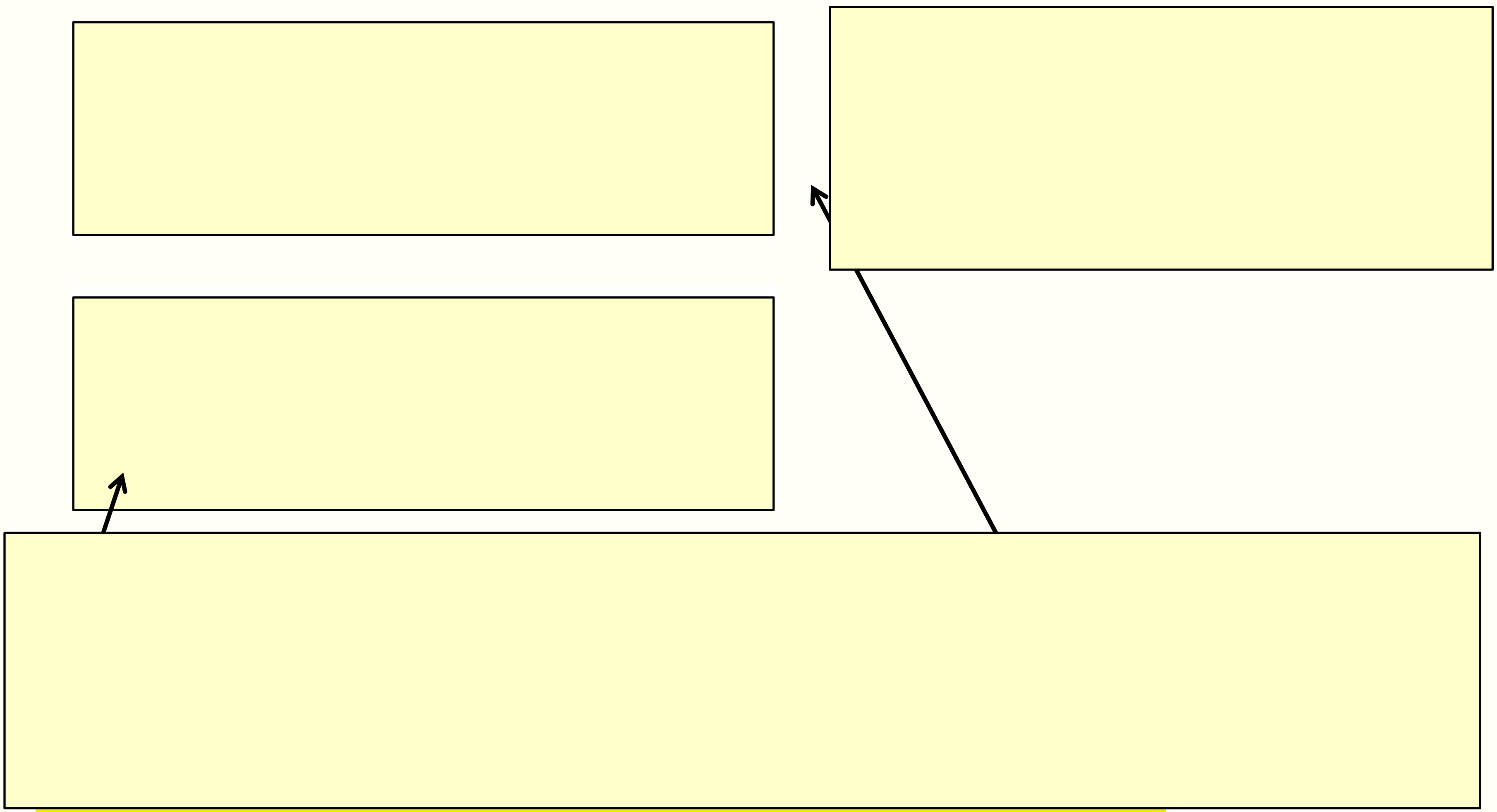
- We learn by memorisation and generalisation

1 2 3 4

http://www.cs.ubc.ca/~murphyk/Teaching/CS340-Fall07/L1_intro.pdf

What is learning?

- We learn by memorisation and generalisation



http://www.cs.ubc.ca/~murphyk/Teaching/CS340-Fall07/L1_intro.pdf

Examples of generalisation

- We are extremely good at generalisation
- When you were a child
  - You were shown a few pictures of giraffe
  - You can recognise any giraffe
  - You did not see and have not seen all the giraffes in this world
- In ENGG1811
  - We showed you examples on how to write code
  - You can now write code for problems that you have never seen before

Two steps of supervised learning

- When you were a child
  - You were shown a few pictures of giraffe/tiger/lion/…
  - You can recognise any giraffe/tiger/lion/…

- Two steps of **supervised learning**
  - Training by examples
  - Generalisation

- Generalisation is the ability to make reasonable guesses when you are given situations that you haven't encountered during training

What are the attributes of good learning algorithms?

Measuring the ability to generalise

- We divide the data into two disjoint sections. We call them training data and test data respectively
- Example: For our data set
  - We have 50 sets of measurements for normal, fault1 and fault2
  - We use 40 sets in each class for training, the rest for testing

Example: Training and Testing data

x	y	Label
5.1	3.5	Normal
4.9	3.0	Normal
…	…	…
		Normal
5.0	3.3	Normal
7.0	3.2	Fault1
5.7	2.8	Fault1
6.3	3.3	Fault2
6.2	3.4	Fault2
5.9	3.0	Fault2

40 sets of measurements for training

10 sets of measurements For testing

40 sets (training)

10 sets (testing)

40 sets (training)

10 sets (testing)

Using training data

- We do not feed all the measurements to the machine learning algorithm
  - We feed only the training data
  - The machine learning algorithms do not know what the test data are

- The machine learning algorithm calculates a model based on the training data

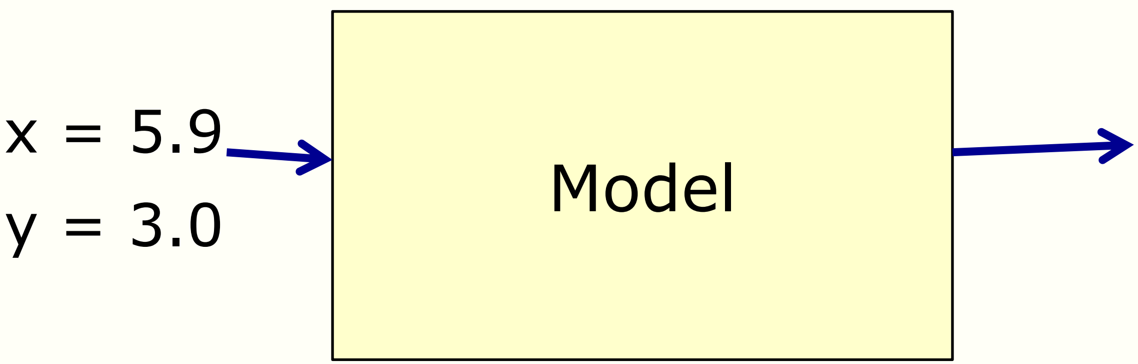
Using test data

x	y	Normal/Fault
5.9	3.0	**Fault2**

1 test sample from fault2

We know what the correct answer is.

- An exam for the model calculated by machine learning
  - We give (x,y) of the test data to the model and see what the model gives

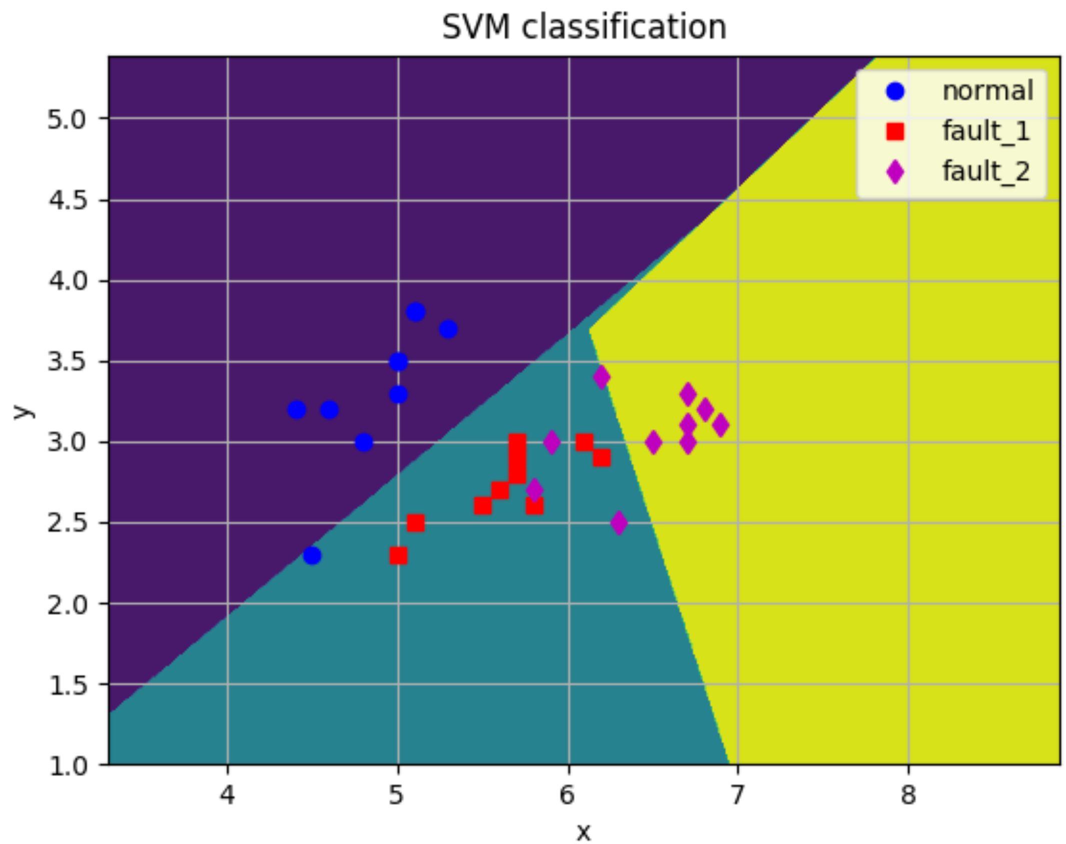


If the model answers "fault2", then it's correct; otherwise wrong.

SVM - Testing data

Python file: classify_3classes.py

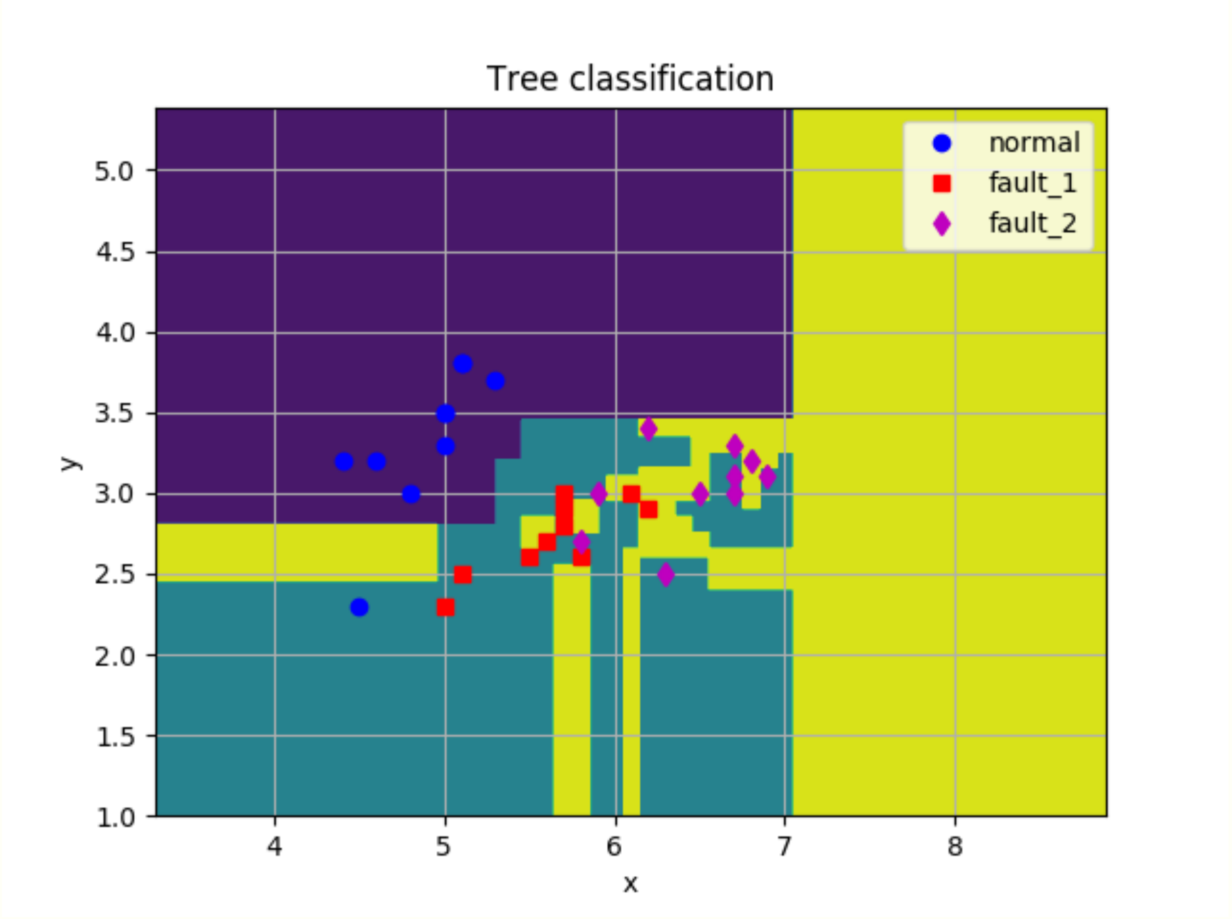
- Purple – Normal; Blue – Fault1; Green – Fault2
- Markers for testing data only



Decision tree - Testing data

Python file: classify_3classes.py

- Purple – Normal; Blue – Fault 1; Green – Fault2
- Markers for testing data only



Confusion matrix

- We present results in the form of confusion matrix:

↓ Correct answers from testing data ↓

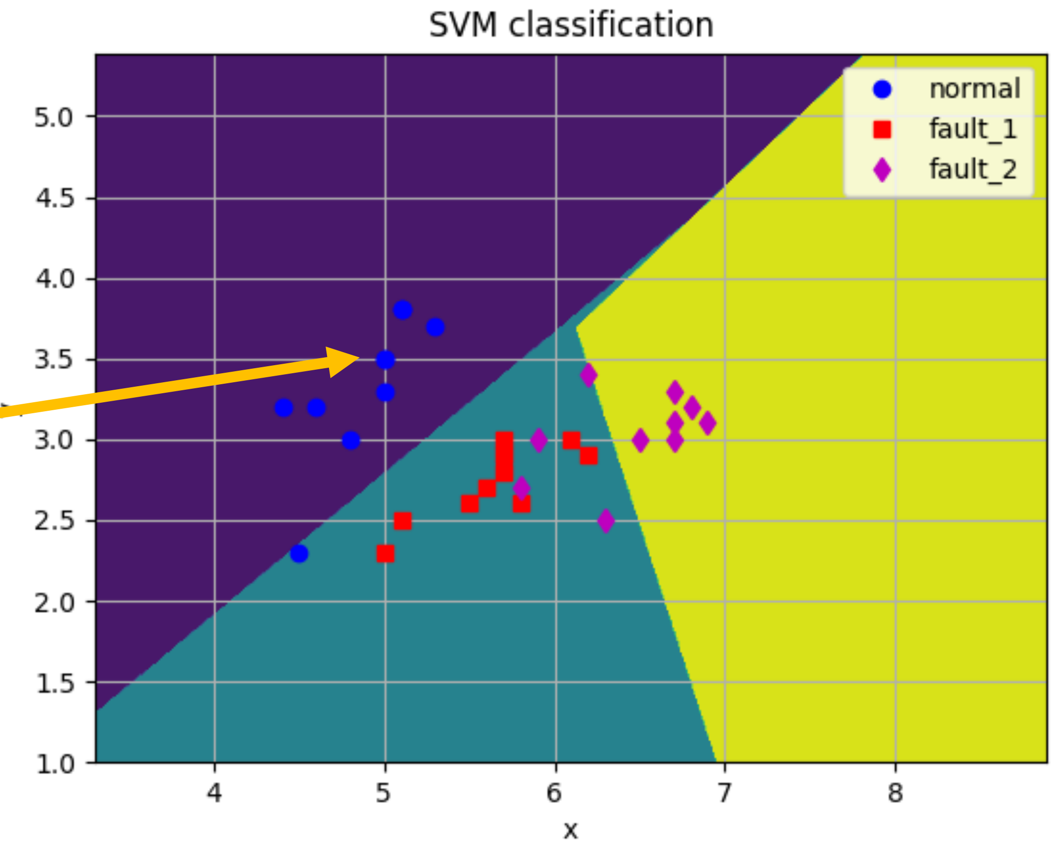
➔ Prediction from model ➔

	Normal	Fault1	Fault2
Normal			
Fault1			
Fault2			

#wrong answers
Expecting: Normal
But model says Fault2

#right answers
Expecting: Fault2
Model also says Fault2

SVM – Working out the confusion matrix



Note: There are 3 overlapping data points here

↓ Correct classes ↓

➔ Prediction ➔

	Normal	Fault1	Fault2
Normal	9	1	0
Fault1	0	10	0
Fault2		4	6

Quiz

Python file: classify_3classes.py

- A better model gives more correct (or fewer wrong) answers for the **test** data
- Which model is better?

SVM

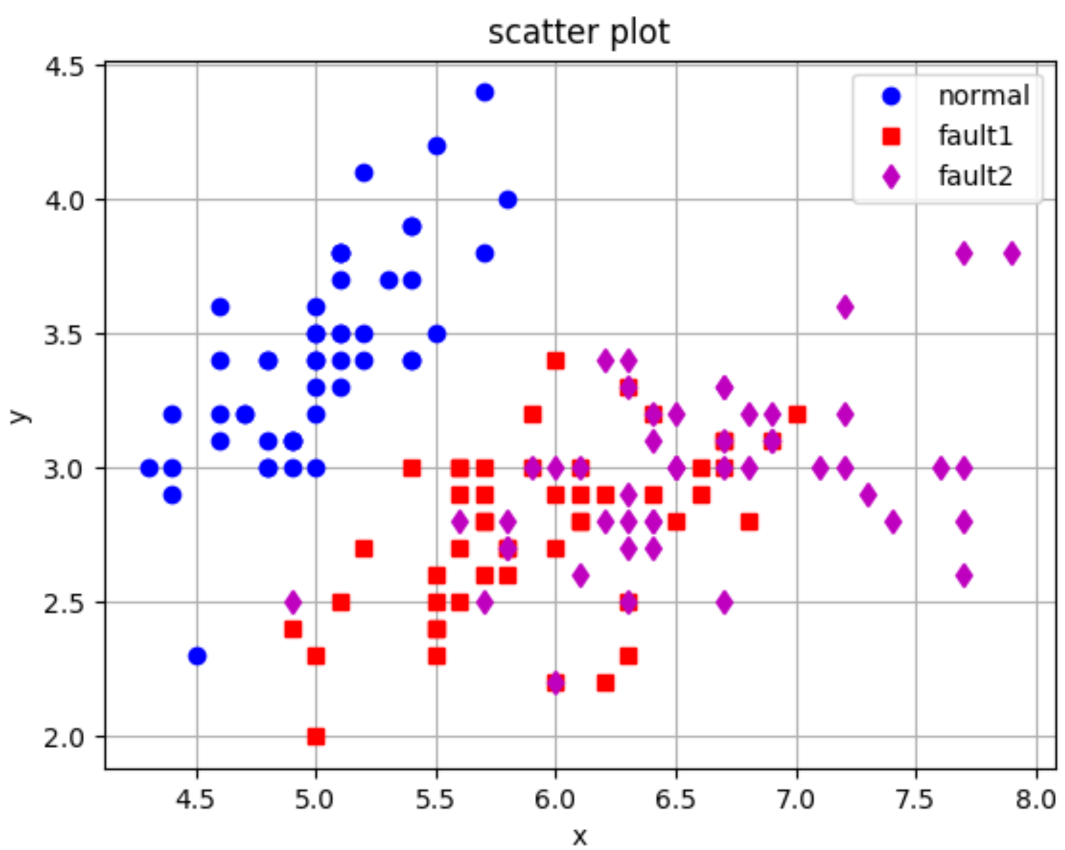
	Normal	Fault1	Fault2
Normal	9	1	0
Fault1	0	10	0
Fault2		4	6

Decision tree

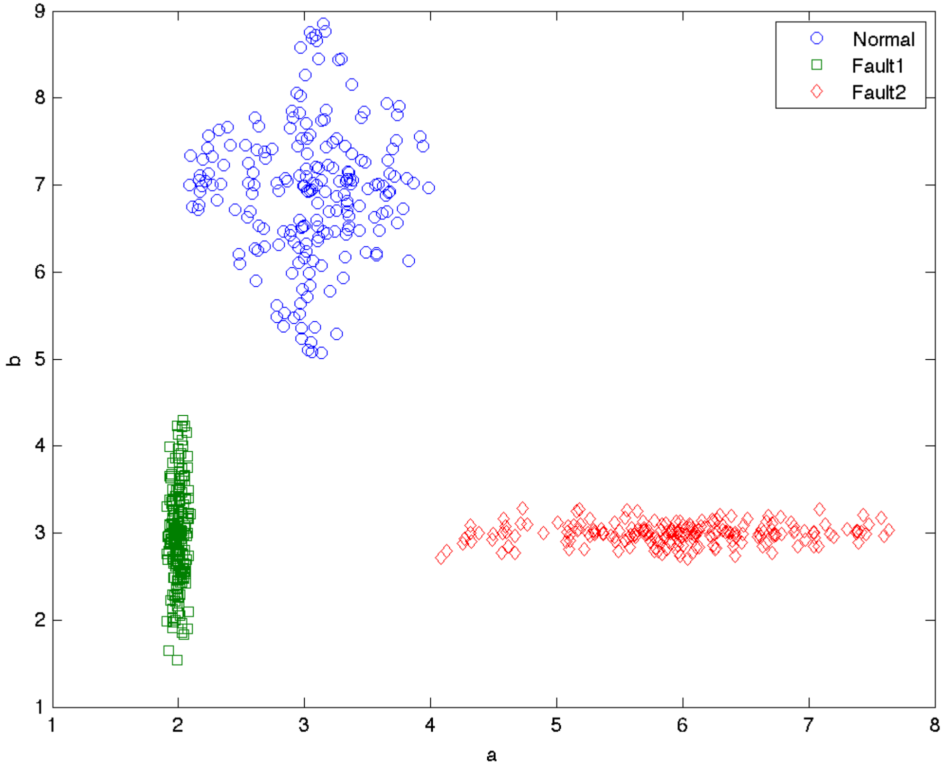
	Normal	Fault1	Fault2
Normal	9	1	0
Fault1	0	7	3
Fault2	0	6	4

Why is it difficult to tell fault1 from fault2

The regions of fault1 and fault2 overlap. The boundary between them is not clear.



What can we do to improve classification results?



The need to rotate training and test data

x	y	Normal/Fault
5.1	3.5	Normal
4.9	3.0	Normal
…	…	…
		Normal
5.0	3.3	Normal

40 sets of measurements for training

10 sets of measurements for testing

- Why should we choose these particular ten measurements as test data?
- In fact, there is no particular reason to do that
- We should rotate training and test data

N-fold validation

- For example, 5-fold validation
- Divide the data into 5 sections
  - Section 1 (Rows 1-10), Section 2 (Rows 11-20), …. Section 5 (Rows 41-50)
- Perform 5 rounds of training and testing

Round	Training	Test
1	Sections 2,3,4,5	**Section 1**
2	Sections 1,3,4,5	**Section 2**
3	Sections 1,2,4,5	**Section 3**
4	Sections 1,2,3,5	**Section 4**
5	Sections 1,2,3,4	**Section 5**

x	y	Normal/Fault
5.1	3.5	**Normal**
4.9	3.0	**Normal**

Features

- For good classification results, we need to find **characteristics** or **features** that can help us to distinguish between different classes

- A key lesson from machine learning: You need **informative features** to get good classification results

- How can you get informative features?
  - Domain knowledge of the problem
  - Trial and error

Feature selection

- Feature selection
  - Give the machine learning algorithms a lot of features and let algorithms rank the features for you

- Feature learning
  - Some modern day machine learning algorithms can learn the features from data

There is a lot more to machine learning …

- Unfortunately we won't have time to talk about it
- I believe machine learning is a useful tool for all engineering disciplines

Big Data Analytics in Chemical Engineering

Leo Chiang, Bo Lu, and Ivan Castillo

The Dow Chemical Company, Freeport, Texas 77541; email: hchiang@dow.com

Annual Review of Chemical and Biomolecular Engineering 2017. 8:63–85

This article highlights recent big data advancements in five industries, including chemicals, energy, semiconductors, pharmaceuticals and food.

Artificial Intelligence & Sequential Decision Problems

(CIV6540 - Machine Learning for Civil Engineers)

Professor: James-A. Goulet

Département des génies civil, géologique et des mines
Polytechnique Montréal

Potential applications of AI in civil engineering

Anomaly detection

Input data: Real-time monitoring of a structure behavior

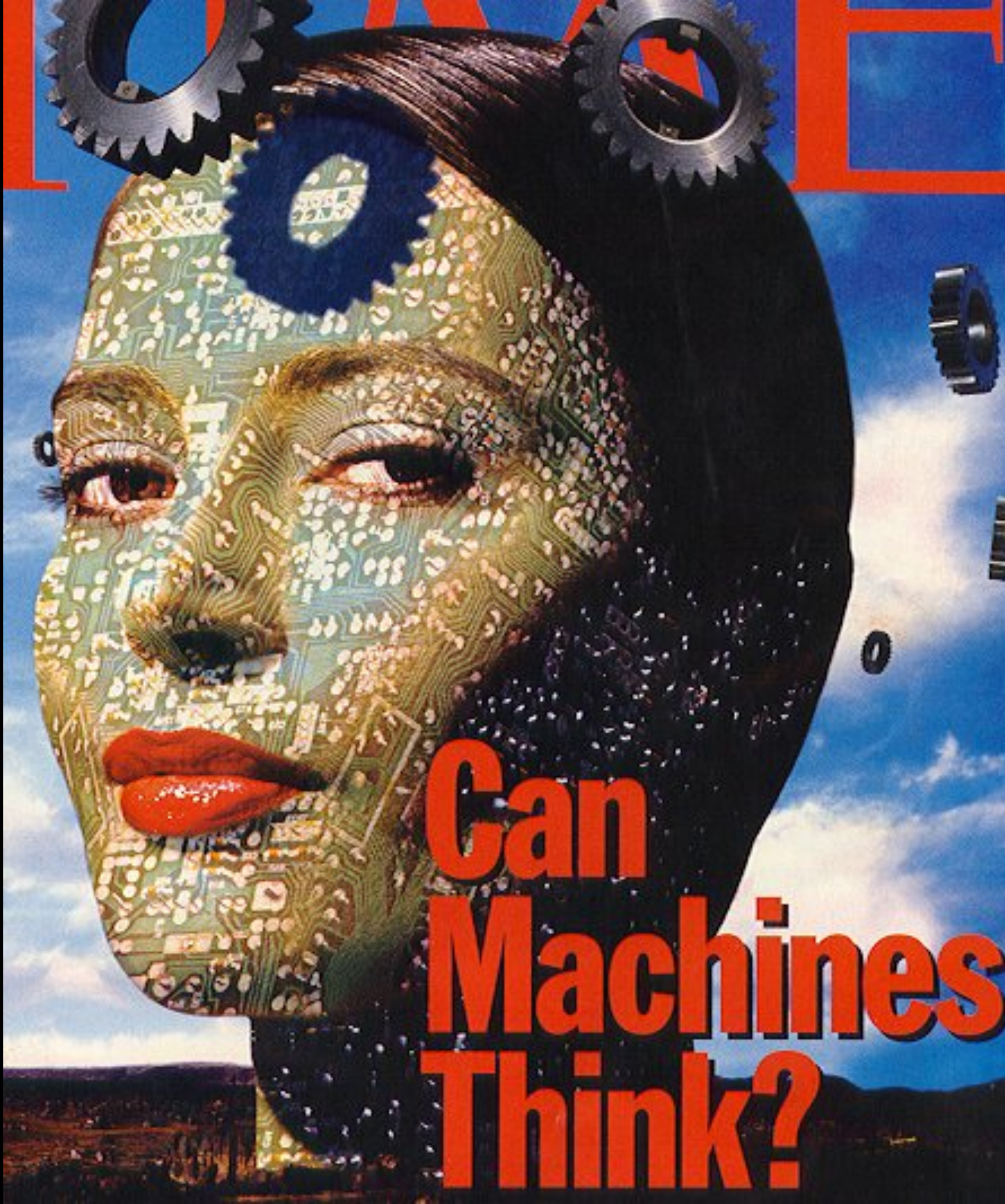
AI role: Decides in real-time whether or not to alert an engineer

http://www.polymtl.ca/cgm/jagoulet/Site/Goulet_web_page_TEACHING.html

Machine learning is everywhere

- Machine learning is everywhere. You have probably used machine learning without knowing it e.g. automatic face detection and recognition

http://support.apple.com/kb/ht4416

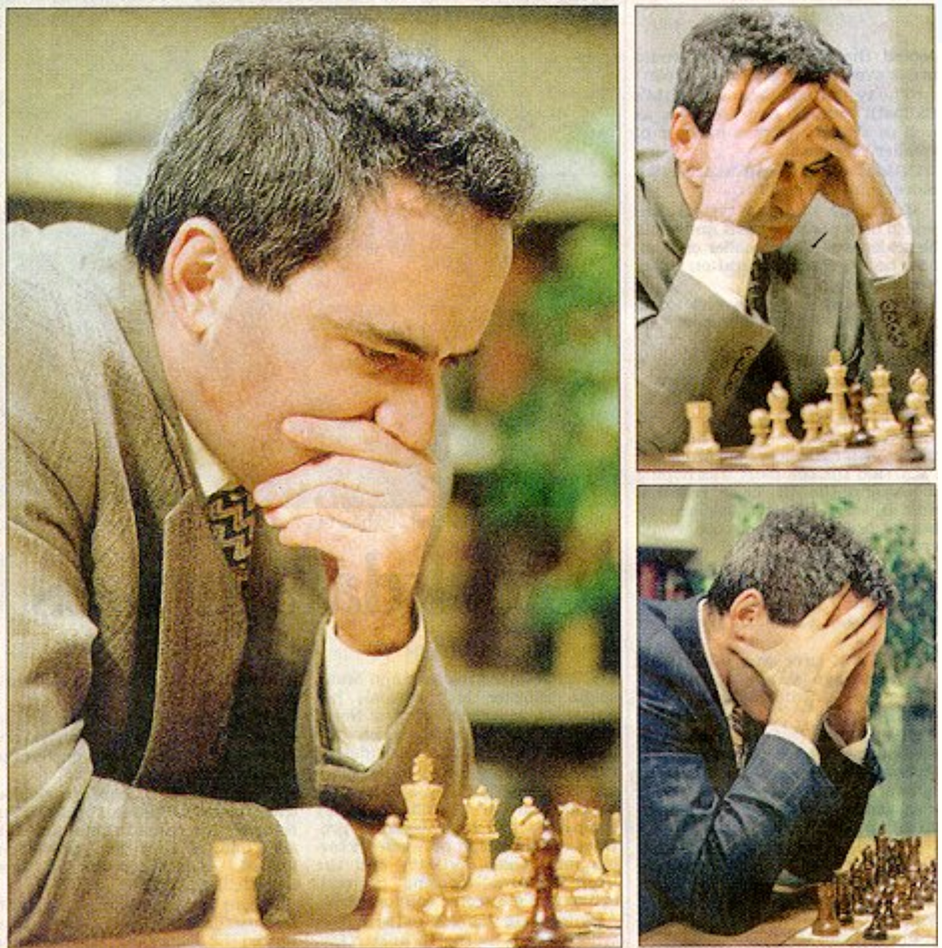

Can
Machines
Think?

Computer defeated Grand Master

On May 12th, 1997, the best chess player in the world, Gary Kasparov, lost a six-game chess match to a computer named “Deep Blue 2”

The press called this “humanity’s endgame” and a “bloody nose for human[s]”

What was so significant about this event?



World champion Kasparov rides an emotional roller-coaster — Pictures: AP (main), Reuters

Being able to program a computer to defeat a Grand Master level chess player had been a long-standing goal of the science of ***artificial intelligence*** - and it was achieved 1.7 decades ago.

Computers understand natural languages

TECHNOLOGY

Computer Thumps 'Jeopardy' Minds

In the end, humans were no match for the machines. In a nationally televised competition, the Watson computer system built by International Business Machines Corp. handily defeated two former "Jeopardy" champions.

How did Watson fare on Jeopardy Monday night? AllThingsD's Arik Hesseldahl and Steve Baker, author of "Final Jeopardy: Man vs. Machine", join digits to discuss Watson's performance on Jeopardy and the future of the supercomputer.

Watson took an early lead and maintained it throughout the last game Wednesday until the final clue. All three contestants correctly guessed the final clue: Who is Bram Stoker?

So Watson came away the winner with a final three-day tally of $77,147. Contestant Ken Jennings came in second with $24,000 and Brad Rutter came in third with $21,600.

http://online.wsj.com/articles/SB10001424052748704171004576148974172060658

The hardest game of all: Go

Artificial Intelligence

Google's DeepMind wins historic Go contest 4-1

DeepMind's AlphaGo artificial intelligence has won the final match of the Go series against world champion Lee Sedol.

http://www.wired.co.uk/article/alphago-deepmind-google-wins-lee-sedol

Computer can drive cars

- Google driverless cars.
- http://googleblog.blogspot.com.au/2014/05/just-press-go-designing-self-driving.html


Are computers that smart?

Summary

- Computers have changed the way how engineers work

- There are plenty of unsolved engineering challenges and computers can help in many of them

- Two key skills in this course
  - Programming
  - Algorithms

- Plus your domain knowledge + ingenuity
  - You will have plenty to offer the world