IOC Topic 5.1 - Introduction to Machine Learning

Transcript & Notes: PART 1

Author: Dr. Robert Lyon
Contact: robert.lyon@edgehill.ac.uk (www.scienceguyrob.com)
Institution: Edge Hill University
Version: 1.0

Topic 5, Module 1, Part 1

Introduction Slide

Hello and welcome to Topic 5, Module 1, Introduction to machine learning. This is part 1 of the module. It will gently introduce ideas and concepts that may unfamiliar to you. Part 1 will take approximately 1 hour to complete. My name is Dr. Robert Lyon, and I'll be your guide during this module. Notes will be provided that accompany this content. I advise that you keep those nearby as we move through the slides. Each slide is numbered, and this number corresponds to a location in the notes.

In the notes I'll sometimes encourage you to undertake some optional self-study. These self-study opportunities will help you understand the content presented. I'll also provide links and references in the notes that enrich the content being discussed – follow those up at your leisure.

Slide 1

We hear a lot about artificial intelligence these days. There's a great deal of hype, and we often hear exaggerations in the media, about what such technologies can do at present.

- But what exactly is Artificial Intelligence (A. I.)? Well, it is a field of study concerned with reproducing, or replicating as closely as possible, human intelligence.
- A.I. is actually an umbrella term – it is comprised of many sub-fields of study. These focus upon replicating,
    - Our capacity for speech.
    - Our ability to see. Humans have a complex visual cortex that provides us with exceptional vision. Despite this complexity, we are able to process visual information with ease, which is remarkable.
    - Our ability to process complex auditory information.
    - Our innate reasoning abilities, that is, our capacity to infer and deduce knowledge from available information.

- Our ability to learn from information provided to us, to make decisions.
- Our natural ability to interpret language, whether written, spoken, or understood via touch (braille).
- Our ability to move our sophisticated bodies around with precision, lift objects, exert fine motor control.
- Our consciousness. This is somewhat contentious as it is difficult to define conscious. Nonetheless, some are trying to replicate this in machines.

- Here we'll talk about replicating the human capacity for learning.
- This field is known more generally as "Machine Learning".

Additional Notes:

To learn more about the visual cortex, this is a good place to start: https://en.wikipedia.org/wiki/Visual_cortex.
For those curious about consciousness, it's definitions and interpretations, this 2018 article in Scientific American may be of interest: https://www.scientificamerican.com/article/what-is-consciousness/

Slide 3

In this module we'll focus on replicating our capacity for learning. This means we'll be introducing a field of study known more generally as Machine Learning (M.L.). Machine Learning is a fascinating subject. It combines insights from Statistics, Logic, Psychology and even Neuroscience. These are brought together to create automated systems capable of "learning" by themselves.

Machine Learning isn't concerned with replicating exactly how humans learn. This is because our learning capabilities are subject to some problems, bias for instance. This makes human decision making flawed. There are countless examples of human decision making leading to bad outcomes.

Thus M.L. attempts to replicate the best aspects of human learning, improving upon them to yield better decision making. To do this machine learning focuses on using available information to make optimal decisions/predictions. What this means in practice will become clear as we progress.

As we progress, you'll acquire an understanding of M.L. that will allow you to apply it for yourself moving forward.

Additional Notes:
Some examples of poor decision making summarised here: https://www.forbes.com/sites/erikaandersen/2013/10/04/it-seemed-like-a-good-idea-at-the-time-7-of-the-worst-business-decisions-ever-made/

Slide 4
You may wonder about the types of problems you can apply machine learning to. Well, here are some examples:

- In my work I apply machine learning algorithms to help look for rare types of star, using large radio telescopes. Here M.L. algorithms look for patterns in astronomical data that are unusual in some way.
- We can use M.L. to try and predict weather patterns, very important for various industries.
- M.L. is now being adopted by medical professionals, who aim to solve a variety of problems. From identifying regions characteristic of disease in medical images, to personalising radiotherapy doses for individual patients.
- M.L. is applied to solve various voice recognition tasks. Such systems can help those with physical impairments, but also can make life easier for all of us via use of tools such as Apple's Siri or Amazon's Alexa.
- M.L. is being widely applied for face recognition tasks. You're perhaps most familiar with the capabilities of Apple Phones, which can be unlocked upon seeing your face without the need for a passcode.

There are many more domains where machine learning is applied. These examples represent only the tip of a very large proverbial iceberg.

Additional Notes:

Here's a link talking about some of my own work that involves applying machine learning to search for Radio Pulsars
(https://en.wikipedia.org/wiki/Pulsar):
https://www.imeche.org/news/news-article/machine-learning-gives-astronomers-a-hand

Using machine learning we made an important discovery – well, my PhD student did: https://phys.org/news/2018-10-student-slowest-pulsar-star.html

Here's how machine learning is being applied to help improve patient treatment: https://physicsworld.com/a/machine-learning-a-game-changer-for-radiation-therapy/

Slide 5

Let’s take a moment to review the topics this module will introduce…

- Useful terminology and key concepts. This will help you understand more advanced content as we progress.

  This content will be covered right here, in part 1.

- The mathematical background required to understand machine learning. This will be basic – the focus is on fostering an understanding, without worrying about complex equations. Our first automated learning system will follow shortly after.

  This content will be introduced in part 2.

- A number of machine learning algorithms from first principles, supported by examples you can try for yourself.

  This content will be introduced in part 3.

The aim is to help you acquire the foundational knowledge required to apply machine learning in practice.

Let’s begin this journey by first considering how human decision-making works.

Slide 6

Humans are capable of sophisticated learning via trial & error, reinforcement, and by receiving information from some form of “oracle”. An oracle could be a parent, teacher, sibling, friend, book, or an online resource such as this.

The oracle provides a source of information that we process, store in memory, and utilize when required. Our memory is capable of storing a lifetime of knowledge and experience, making it a highly sophisticated analogue of a computer hard drive. Though we store information differently to computer hard drives, the content is the same in both cases – data.

This dataset becomes populated as our senses (sight, smell, sound etc.) start collecting information about the world. When we’re born, this dataset is mostly empty. Over time we collect experiences. Eventually we connect our experiences and the information given to us, with some tangible outcome or label.

So, for example, most babies learn automatically that when they cry, then can get parental attention. Ultimately, we learn to use our experiences to make decisions or take actions. Given the same information, it is possible that two people will take very different actions. This is because knowledge and experience differ from person to person. In other words, the content of our memory datasets is different.

Despite any differences between us, human decision making is geared toward making the best, or "optimal" decisions using past experience and available information.

Additional Notes:

What is memory? This is an active area of study, as it's difficult to quantify. There is an interesting resource available online that explores the nature of memory. You might like to study this content: https://www.fil.ion.ucl.ac.uk/memo/memory.html

You may wonder, what is the storage capacity of the human brain? This is very difficult to estimate, though some have tried. Work done in 2016 indicated that each synapse in the human brain has a storage capacity of roughly 4.7 bits of information.

If one bit is represented by a single 0 or 1, then each synapse can store roughly this much information: 00000.

Consider that there are 1,000 trillion synapses in the average human brain. So that's 4.7 bits of information, multiplied by 1,000 trillion. In other words, that's 4.7 × 10^15 bits. That works out as approximately 0.587 PB (petabytes). That's enough space to store 2,348,000 (two million, three hundred and forty-eight thousand) hour long Netflix shows.

Link to the paper that presented this result available here: https://elifesciences.org/articles/10778 and also summarised here: https://www.scientificamerican.com/article/new-estimate-boosts-the-human-brain-s-memory-capacity-10-fold/

Slide 7

Let's continue thinking about experience. Experience is describable and quantifiable as data.

You already intuitively understand what data is. However, to understand M.L., we must first become familiar with the terminology used in the field. Let's start by considering that data can be quantified and recorded using a simple table structure.

Here we have a table describing the characteristics of animals we've observed at some time during our lives. The animals are described using four characteristics – their mass in Kilograms (Kg), their height in centimetres (cm), the number of legs they possess, and their primary colour when viewed by eye.

Each of these characteristics is represented as a column in the table. We can see there are three rows in the table, representing three unique animals.
In machine learning we refer to the characteristics – mass, height, number of legs, and colour, as features. Sometimes these are also known as variables or attributes, but we'll stick with the term features.

Each row in the table is called an example. That is, the first row represents our first example animal and so on.

Each example may be associated with a "label". In this case the label would correspond to the name of the animal. The label is also sometimes known as the "class label". For instance, the first row of the table represents an animal that belongs to the class "Cat". This class is distinct from all other classes. How a class label is defined impacts it's interpretation. Thus, since we do not clarify here, the label "cat" applies to all species of Feline. But this class label can be broken down into to "sub-classes" such as "Domestic Cat", "Leopard", "Panther" etc.

When the labels for examples are known, we say we have the "ground truth" labels. This means the truth is known for certain – i.e. the label is undeniably correct.

When a dataset is accompanied by ground truth labels, we call this data set "labelled".

Additional Notes:
A link describing labelled data: https://en.wikipedia.org/wiki/Labeled_data

Slide 8

Here we can see the same table as before, describing the animals we've observed at some time in our lives. Consider the equivalent table you hold in your mind – it will have an enormous number of rows, and many features. To make such datasets easier to study, we label each row using some simple notation.

Here we represent examples using the notation x_i. At first this may appear confusing – what do x and i mean?

Read x as meaning "example", and i as a number that uniquely identifies the example. So here we can see there are three examples described using this notation. Now when I say x_2 you know which animal I'm referring to – the animal in the second row of the table.

We can also represent the class label in a similar way. Read y as meaning "label", and i as a number that uniquely identifies the label.

So here we can see there are three labels described using this notation. Now when I say y_3 you know which animal I'm referring to – the mouse in the third row of the table.

This notation allows us to describe examples in a new way, using mathematics that's easier to write down than written language. Here we represent example 1 as the row values stored between two brackets and separated by commas. We can also represent the label similarly. These brackets enclose what we call sets. Sets are simply collections of objects. Thus x_1 is a set containing the characteristics of the animal in row 1, and y_1 is a set containing the labels associated with the same animal.

We can represent entire datasets using this notation.

You may also note that these sets are similar to the Arrays you'll have learned about in the programming module.

Additional Notes:

To gain a better understanding of sets, self-study this intuitive Wikipedia page: https://en.wikipedia.org/wiki/Set_(mathematics)

Slide 9

We now understand how data can be represented using some basic mathematical notation. Using this notation, we can express what we mean by "experience" in a straightforward way.

Experience is simply a set, E, consisting of pairs of features and class labels. In theory it can contain infinitely many pairs, though since our knowledge is imperfect, this is never the case.

These pairs are known more commonly as tuples. A tuple can contain as many features and labels as required. However, for the machine learning we're discussing here, each tuple contains only one label.

Additional Notes:

A description of tuples can be found here: https://en.wikipedia.org/wiki/Tuple

Slide 10

Sometimes we don't know the ground truth. In such cases, we have to collect information, and provide it for ourselves.

This can be costly and time consuming. For instance, suppose we want to learn how to recognise a rare type of disease in medical data. To do this we'll need labelled examples showing us what disease looks like. We'll also need examples of what disease doesn't look like.
To get this data, patients will need to be scanned, and their scans manually labelled by a medical expert as disease or non-disease. Clearly this process costs time and money.

Suppose we're given the following dataset. It has 4 features, External Temperature (℃), Internal Temperature (℃), Time, and Power Consumption.

We're not given labels for this data. This makes the data hard to interpret – it could relate to a variety of scenarios.

We're then told this data represents the records of a heating control system.

If the ground truth is unknown, this is called an "unlabelled" dataset.

This can be represented using the following notation: $(x_1, ?)$

Sometimes some extra information will be provided, which converts an unlabelled data set into a "partially-labelled" dataset.

Slide 11

Features can be categorical or numerical.

The colour feature and the class label are both represented by categorical data types. Here each potential value corresponds to some category. The mass, height and legs features, are all numerical.

Usually we “discretise” categorical features, as this makes them easier for many algorithms to work with. This is either done automatically using tools, or manually.

For instance, here the class labels have been changed to numerical values. These values have been chosen to map intuitively to the original categories.

This time the colour feature has been changed so that it is numerical.

We may also “normalise” our data from time to time. Normalization is a technique often applied as part of data preparation for machine learning. The goal of normalization is to change the values of numeric columns in the dataset to a common scale, without distorting differences in the ranges of values. For machine learning, not every dataset requires normalization. It is required only when features have different ranges.
We can apply this process to mass feature.

Here we can see that the mass values now all lie in a range from 0 to 1. It isn’t important right now to understand exactly how this process works. Rather, we want you to know that it happens, and why. For those that are interested, the notes for slide 10 show how the normalized values are computed.

Additional Notes:
Consider a data set containing two features, age, and income. Age ranges from 0–100. Income is about 1,000 times larger than age and ranges from 20,000–500,000. Thus, these features are in very different in terms of range. Such differences can cause problems for M.L. algorithms, as they can treat this disparity in data ranges as more meaningful than it actually is. So, we correct for it by scaling the data to a new range.
How were the values in the final table computed via normalisation? We use a formula to normalise our data into a desired range. In this case the desired range is [0,1]. The formula for doing this is follows:

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

Let’s work through it:

$$z_1 = \frac{4.2 - \min(x)}{\max(x) - \min(x)}$$
$$= \frac{4.2 - 0.1}{18 - 0.1}$$
$$= \frac{4.1}{17.9}$$
$$= 0.229$$

You can apply the same formula to the other rows in the table. This normalisation is a little contrived but suits our example. In reality you might apply more complex normalisation formulae to your data.

Slide 12

As humans we are able to use our senses to extract the key features, we can use to make decisions.

So, for instance when we look at an animal, we can almost automatically assess its size, mass, age, species, etc. Humans are gifted feature extractors.

We are also able to assign labels with ease, and extremely quickly.

Slide 12

Using our experience set E, humans are capable of extremely sophisticated learning. For example:

Humans are capable of accurately,

- Clustering. When given a few examples of different types of animal, we are able to group them together in different ways.
- Classifying. We are able to classify unseen animals as for example, a mammal or a bird.
- We can make predictions based on our past experiences, for example, it's unlikely to rain when it's sunny out.

Such sophisticated learning is only possible due to the feedback we receive during our lives. If this feedback is inaccurate, or poor quality, so too will be our decision making. This seems obvious to say, but a good education, which in principle involves receiving high quality feedback, generally makes individuals better decision makers.

Slide 14

There are many different types of learning that enable us humans to perform these tasks,

- Learning via a teacher. This is known as "supervised" learning. It is called supervised learning, because the teacher supervises the student, providing them with valuable feedback. In this example, we're given some labelled examples of birds and mammals in the experience set E. Then we're asked to correctly assign a set of unknown animals to their correct class group –

either mammals or birds. This is a supervised learning task. Supervised learning is only possible when we have labelled datasets.

- It is possible to acquire knowledge from a mix of independent learning and teacher led learning. This is called “semi-supervised” learning. In this example, we’re given some labelled examples of mammals in the experience set E. Then we’re asked to correctly assign a set of unknown animals to their correct class group – either mammals or birds. In this instance we don’t know what birds are, but that’s ok. We can use what we know about mammals to separate the animals into two groups accurately. Semi-supervised learning is possible when we have partially-labelled datasets.
- Then there is “unsupervised” learning. Here learning is done via trial and error, without feedback from any oracle. In this example, we’re given animals in the experience set E, but there are no labels. Therefore, we don’t know what a mammal is, or what a bird is. Nonetheless, we’re again asked to correctly assign a set of unknown animals to their correct class group – either mammals or birds. To solve this problem, we can use the inherent structure of the data, to self-separate it into groups. For instance, most mammals are larger than birds. Thus, mammals generally have higher mass. This means we can separate on a feature like mass in this instance. However, that feature may not work, if all the unknown mammals are field mice, and all the unknown birds are Ostriches or Emus. Unsupervised learning is possible when we have unlabeled datasets.

  It is often the case that the type of data available to us, dictates the learning that is possible.

<u>Slide 15</u>

These sorts of problems are “binary” classification tasks.

- They are called binary, as there are two potential classification options. In this case mammal or bird.
- Such classification tasks are everywhere, and we do them every day. We classify the weather as likely to rain or not and pick up an umbrella if needed. We Classify the faces of people we see into people we know, and those we don’t.
- We complete these tasks effortlessly. Yet to complete them, we make predictions almost without conscious thought.
- Classification problems in the real-world are usually far more complex than those presented here. For example, there may be more than 2 potential classes to predict.
- These are known as multi-class problems. Let’s try one.

Slide 16

Let's test your own decision making. There are five animal families summarised on this slide, using the image of their most famous member.

There is the Feline family, in the top left image, Lutrinae family in the middle left image (this family includes animals like the Weasel, Otter, and badger relative). The Canine family is in the bottom left image, the Mongoose family in the top right image, and finally the Viverridae family (which includes Civets) in the bottom right image.

The question is this:

Can you tell me which family of creatures this animal is most closely related to? This will only work if you haven't seen it before. Is it a,

- Canine?
- Feline?
- Lutrinae (pronounced "Loo-tree-nigh")?
- Viverridae (pronounced "Viv-ver-i-die-)?
- Mongoose?

I imagine your currently searching your past experience and looking for features in the image that will tell you what type of animal it is!

Well, this animal is a Fossa:
https://en.wikipedia.org/wiki/Fossa_(animal)
It is a rare animal endemic to Madagascar. It's most closely related to the Mongoose family, though appears very similar to Felines in most respects. If you answered correctly, then your past experience was sufficient for this task. If not, then your experience was lacking - but now you've learned via trial and error, so you're unlikely to make this mistake again.

Slide 17

In the last slide, we saw an animal you may not be familiar with.

Even if you have extensive knowledge of the animal kingdom, you may not have known the animal was a Fossa.

Perhaps this animal was not described in your personal experience dataset.

Even if you did not know it was a Fossa, you could still tell it was a type of mammal.

This application of your knowledge, from past experience, illustrates your ability to generalize beyond known facts.

Humans are gifted generalizers - but they are susceptible to two problems:

- Overfitting. Overfitting occurs when we learn so well that our knowledge becomes too specialised. Imagine there's a student trying to prepare for an exam. They buy a revision book and learn every factoid it contains. The student fails to realise that the book doesn't contain all the knowledge that could be tested in the exam. They take the exam, and the worst happens – nothing from the revision book is in the exam. In this case, failure happens because the student over trained, that is "overfitted", to the revision book.
- Underfitting. Underfitting occurs when our knowledge becomes too flimsy. The student who failed before now tries to re-sit the exam. This time, they buy 100 revision books to learn from to avoid their last mistake. Except they can't process that much information. They diligently dip into each book lightly, not covering any topic to any great depth. The student takes the resit and the worst happens again – their knowledge is too shallow, and they fail the exam. In this case, failure happens because the student under trained, that is "underfitted", to the revision books.

  It is desirable to achieve a balance between under and over fitting in practice. There is certainly a trade-off to be had.

Slide 18

Before we proceed any further, we recap the notation you've learned so far. If we have the dataset shown in the slide,

x_i represents an example (row) in a dataset (replace $\boldsymbol{i}$ with row number).

Such that we can label each row in the dataset as shown.

x_i can be viewed as an array containing many values (columns).

y_i is a ground truth label associated with x_i.

Such that we can annotate the labels as shown.

A tuple is a finite ordered list.

And finally, E is the set of experience, which is comprised of many tuples.

Slide 19

So far in this module we've introduced,

- Data sets, features, and class labels.
- Labelled and unlabelled datasets.
- Different types of learning that we're capable of.
- Bias.
- The concept of classification.
- Generalisation, and under/over fitting.

Next we develop these ideas further to study automated machine learning. We'll find that M.L. borrows a great deal from how we learn and is just as susceptible as we are to error!