CL exercise for
Tutorial 1

Introduction

Objectives

In this tutorial you will:

- meet your tutorial group colleagues, and discuss and work with them;
- learn how to submit your tutorial solutions for marking;
- learn more about *Venn diagrams*, *decision trees*, and *encodings*.

Tasks

Usually, this exercise sheet will contain a number of tasks to be completed before the tutorial. **This tutorial is different.**

Exercise 1 must be completed, and submitted, before the submission deadline. (It is basically 3 free marks!)

The remainder of the exercises are intended to be done during the tutorial. It is (very) unlikely you will have time to get through them all – work through with your group in the time available, concentrating on doing early parts fully rather than getting as far as possible.

You should read through the *entire* tutorial sheet before the tutorial, since you will get on much better if you know what you are trying to do with the information you will gather in the tutorial.

Deadline

16:00 Tuesday 28 September

Reminder

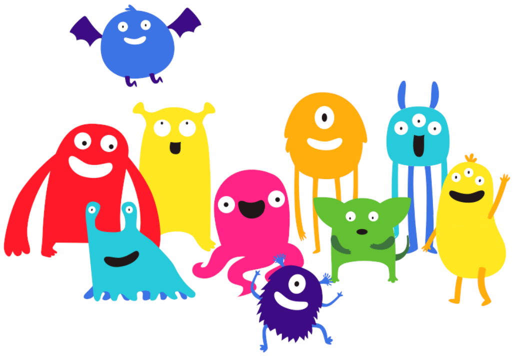
Good Scholarly Practice

Please remember the good scholarly practice requirements of the University regarding work for credit.

You can find guidance at the School page

https://web.inf.ed.ac.uk/infweb/admin/policies/academic-misconduct.

This also has links to the relevant University pages. Please do not publish solutions to these exercises on the internet or elsewhere, to avoid others copying your solutions.



I. We're all different.

In the tutorial, we'll explore diversity and how it can be discovered with logical queries.

You'll start getting to know your colleagues by focusing on qualities which make you uniquely you. You'll use Venn diagrams and decision trees to help you visualize and reason about differences between the members of your group.

There are (about) 12 students in your tutorial group. For this CL part of the tutorial, it will be easier if you split into two subgroups. Try to get a mix of backgrounds in each subgroup.

Exercise 1 –mandatory—marked–

Before the session:

1. Think of 3 questions for your colleagues to help you get to know them.
   The questions should help you identify each student in your group, should admit only true/false answers, and should be sensible – you don't want to be intrusive.
2. Write down your questions: you can either type them and save them in a file, or write them on a piece of paper and then scan or take a picture of the paper.
3. **Upload** the file in the tutorial assignment for week 1 on Learn.

Solution to Exercise 1 So many options here... Take, for example, the following:

Q1 Do you think jaffa cakes are biscuits?
Q2 Have you ever seen a tapir?
Q3 Did you set up properly your GitHub account?

[*Chances are only half of your group will answer yes to Q3. :-)*]

Exercise 2

At the beginning of the session:

Talk to your group colleagues and your tutor. Introduce yourselves: What's your name? Where are you from? Which degree are you on? Is there something you're an expert on?

Your tutor will do the same.

Exercise 3

1. Check on Learn the feedback you received for Exercise 1. Ask your tutor for explanations if something's not clear.

2. Each of you should read out the questions you submitted, and everybody should answer them. Write down everybody's answers to **your** questions, including your own. Then draw the corresponding Venn diagram.

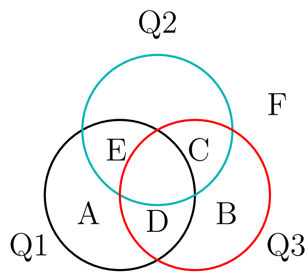
   You might wish to note people's answers to your questions in a text file for later emailing – the 'at home' followup part of this sheet needs you all to have all the answers.

3. Using the diagram, see if it's the case that none of your 3 questions singles out (identifies) a student. That is, for each of the 3 questions you submitted, are there at least 2 students who replied *true* and at least 2 students who replied *false*?

4. Let your colleagues know if your 3 questions verify the constraint above.
   If so, you're a winner! Your questions can be used by your group to solve Exercises 4 and 5. If none of you submitted 3 questions that satisfy the constraint, choose together 3 suitable questions to use next.

Solution to Exercise 3 Let's call the students in our group A, B, C, D, E, and F.
The table below contains all their answers to the questions on slide 5.
We write Y for yes (or true), and N for no (or false).

	A	B	C	D	E	F
Q1	Y	N	N	Y	Y	N
Q2	N	N	Y	N	Y	N
Q3	N	Y	Y	Y	N	N



As you can see, every row contains at least two Ns and at least two Ys, so the constraint is satisfied. We'll use these questions in the next exercises.

Exercise 4

A **query** is an expression obtained from binary questions and the connectives *and* (intersection, $\wedge$), *or* (union, $\vee$), and *not* (complement, $\neg$). For example, *(isAmerican? and hasBeard?) or (isBritish? and (not hasBeard?))*. In other words, it's a compound proposition viewed as a question.

1. Check together if there exists a student in your group who can be uniquely identified by a query formed using the 3 questions your group has selected in the previous exercise. If there is one, then how many more students can be uniquely identified by a query?

2. Change the 3 questions such that each of you 6 can be uniquely identified. Be sure not to break the constraint in Exercise 3.

3. Draw the decision tree corresponding to your new questions and query, as in the lectures. Using the decision tree, define an encoding for all students. What's your code number?

4. How many questions would you need to ask in order to identify all the students in your group (to visit all the leaves of your tree)?

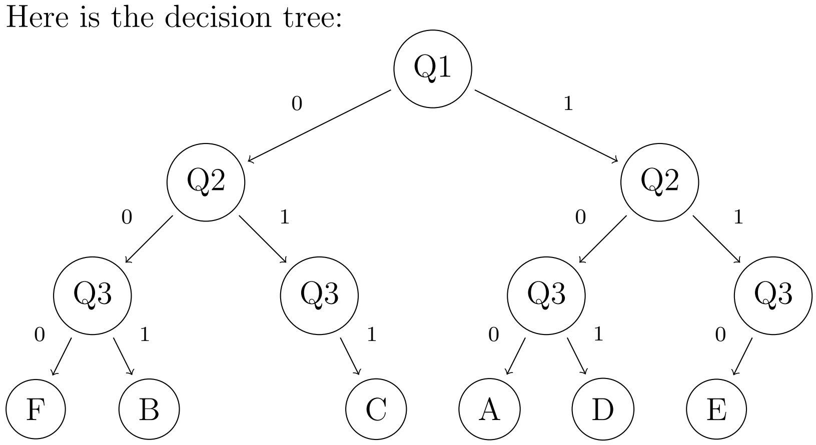
Solution to Exercise 4 Note that a student can be uniquely identified by a query if and only if
they can be uniquely identified by their answers to the 3 questions.

In other words, if the corresponding column in the table on slide 8 is distinct from all other columns, then the student is uniquely identified; otherwise, they are not.

We are lucky: the columns of our table are distinct, so each of the 6 students can be uniquely identified by a query.
For example, A can be identified by Q1 and (not Q2) and (not Q3), so
they are the only student in the group who:

still thinks that jaffa cakes are biscuits,
hasn't seen a tapir yet,
and hasn't set up their GitHub account

Solution to Exercise 4 (cont.)
Here is the decision tree:



Solution to Exercise 4 (cont.) Based on the tree above, my encoding (C) would be 0 1 1.

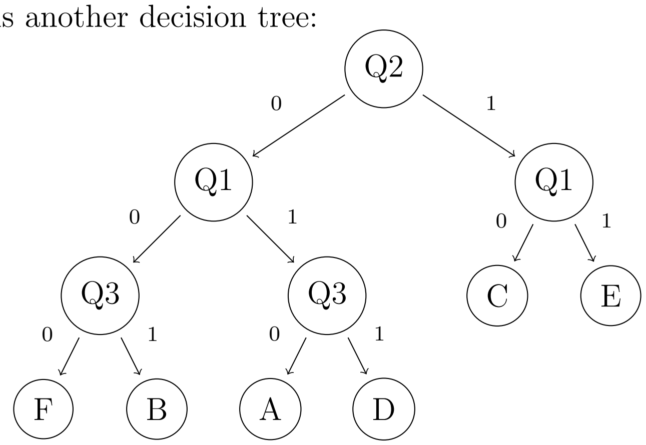
You can obtain this encoding also by substituting 0 for N and 1 for Y in the table on slide 8 – but this is only because the rows in the table and the levels of the tree are the same (and in the same order).

For this decision tree, we would need to check 12 answers to visit all leaves.

We can reduce that number to 10 by changing the order in which we ask the questions (see the next slide).

Solution to Exercise 4 (cont.)

Here is another decision tree:



Exercise 5

1. Can we obtain the same result as in the previous exercise (identify just as many students with a query) by using just 2 base (binary) questions?

2. What is the minimum number of base questions we need in order to identify each of the 6 students with a query?

3. What about the minimum number of base questions for identifying each of the 450 students in your year?

Solution to Exercise 5 With 2 base questions we can uniquely identify at most $2^2 = 4$ students, so you need at least 3 base questions to identify each of the 6 students.

To identify each of the 450 students in our course, we would need at least $\lceil \log_2 450 \rceil = 9$ questions.

This part provides some further work you can do afterwards. You may need, if you were unlucky with your questions, need a question or two from the other members of your subgroup – if so, exchange email (or whatever young people use these days. . .).

II. We have things in common.

Being different is useful when trying to solve complex problems together: every one of us can contribute with things we're experts at. However, it's easy to check that it's often simpler to find things that unite us than things that differentiate us. Let's do the maths!

In the second part of the tutorial, we'll consider the duals of the questions in the first part to reason about things we have in common and how we are alike.

Exercise 6

1. Go back to the questions you submitted for Exercise 1. Check whether it is the case that each question separates at least 1 student from the group – that is, it has both Yes and No answers. If not, then look at your colleagues' questions to find 3 that do this. Draw the associated Venn diagram.

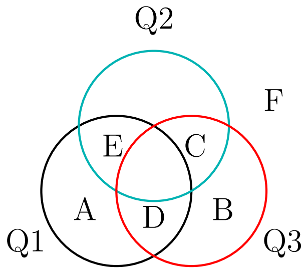
2. Using the diagram, check if any 2 students are united by one of the questions (they give the same answer to the question). What about by a query?

3. Change the 3 questions such that any 2 students are united by a query.

4. What is the minimum number of base questions we need such that any 2 students are united by a question? What about by a query? What if we have 450 students instead of 6?

Solution to Exercise 6 Looking at the table on slide 8, we can see that each question has at least
1 positive answer and at least 1 negative answer.

Looking again at the Venn diagram on slide 8, we can see that A and C (and also B and E) are not united by any of the questions.



However, A and C are united by the query Q1 or Q2 (and B and E by the query Q3 or Q2).

Solution to Exercise 6 (cont.)

No 2 questions that satisfy the constraint (no matter how well crafted they are) can guarantee that any 2 students are united.

On the other hand, provided that we know the students well enough, we can devise two special questions Q1 and Q2 that separate A and B, respectively, from the rest of the group, meaning that:

- only A answers yes to Q1, and
- only B answers yes to Q2.

Now let question Q3 stand for "Q1 or Q2". You can check that any two students give the same answer to either Q1, Q2, or Q3. This shows that,
no matter how many students we have in the group (but at least 4),

- we can find 3 questions such that any 2 students are united by a question, and
- we can find 2 questions such that any 2 students are united by a query.

Exercise 7

What are the odds?

How difficult is it to choose suitable questions for discriminating between or grouping elements of a set?

Obviously, the number of questions is infinite, limited only by one's vocabulary [kudos, mr. Wittgenstein!]. This makes them difficult to analyse. But for the purpose of this tutorial, we can simplify the matter greatly. That is because we are not so much interested in the textual formulation of a question (the words used) as we are in the answers it receives from each of you 6.

So suppose we make no distinction between questions that receive the same answers. For example, if your group comprises students A, B, C, D, E, F, and questions Q1 and Q2 are both answered *true* by A, B, C and *false* by D, E, F, we will consider Q1 and Q2 to be **effectively the same**.

Individual

1. How many effectively different questions can each of you ask?
2. What is the probability of you asking one of those questions?

Team power
Now consider triples of questions.

3. How many triples are there? Note that a question in a triple may be effectively the same as another question in that triple.

4. What is the probability of your group to propose one of these triples?

5. Can you compute the number of triples that satisfy the constraint in Exercise 3 (none of the questions individually identifies a student)?

6. What is the probability of proposing one of those triples?

7. Now think of queries that we can form from triples of questions. Each triple may generate several queries. Can we know exactly how many? Is there a minimum number of such queries? How about a maximum?

8. How many triples can generate a query that identifies a student?

9. What is the probability of identifying a student with a query?

10. Can you think of a similar set of tasks to compute the probability of having any 2 students in your group united by 3 questions that satisfy the constraint in Exercise 6?

By comparing these two last probabilities you will see that there is generally a much greater chance of being alike than different.