Edge Hill University

</TECHUP_WOMEN>

Introduction to Machine Learning

Topic 5, Module 1, Part 2

Duration: 1 Hour

Durham University · Edge Hill University · The University of Nottingham (United Kingdom · China · Malaysia) · University of York · Institute of Coding · Office for Students (OfS)

1. What we'll cover

This module will introduce,

- Useful terminology
- Key concepts

Part 1

- Some basic mathematical background
- Our first learning system

Part 2

- A number of machine leaning algorithms from first principles
- Examples you can try for yourself

Part 3

Aim: to help you acquire the foundational knowledge required to apply machine learning in practice.

Edge Hill University

2. Making Optimal Decisions

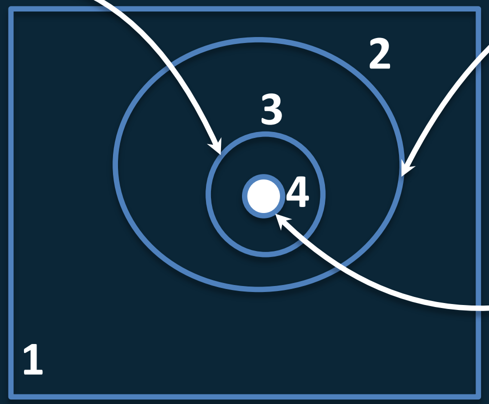
</TECHUP_WOMEN>

- Many have tried to understand how to make optimal decisions.
- We know we should use available evidence at all times.
- Humans are not always so thorough - we do make bad decisions.
- We are biased decision makers - we often use instinct and personal experience to decide.

Personal Knowledge & Experience

E

Available Knowledge & Experience



Knowledge used to make a decision

Set of all knowledge on a subject

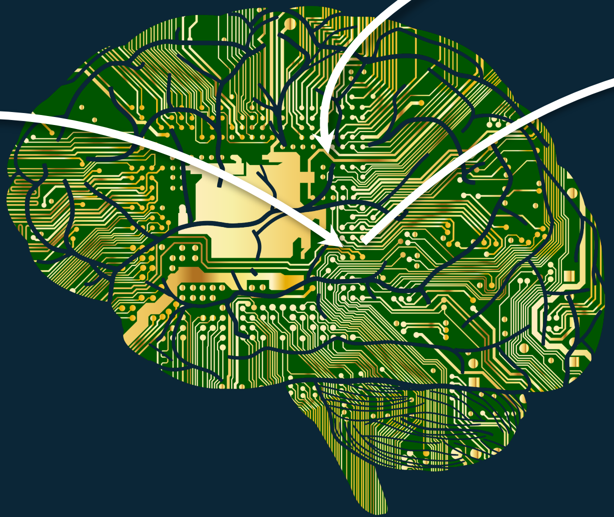
3. Automated Decision Making

- **Machine Learning is concerned with making optimal decisions primarily using the tools of statistics.**

Biased cognitive process

Biased input data

Data (feature values, measures, outcomes…)



Decisions & Predictions

$$E = \{(x_1, y_1), (x_2\,, y_1), \ldots, (x_n, y_1)\}$$

Edge Hill University

4. Automated Decision Making's Flaws

</TECHUP_WOMEN>

- **Biased input data -**
  For example, suppose you want to teach an algorithm to recognize a specific disease in a group of patients.
- You provide the algorithm with data describing patients who are all aged between 60 to 80.
- You then run the algorithm on patients aged 18 to 30. The algorithm performs badly, as it's knowledge is biased toward recognizing disease in much older patients.

- **Biased cognitive process -**
  For instance suppose we try to teach an algorithm to predict when a train on the Tokyo rail network will be late.
- Trains on this network are exceptionally punctual.
- To achieve the best overall performance, just never predict that a train will be late.

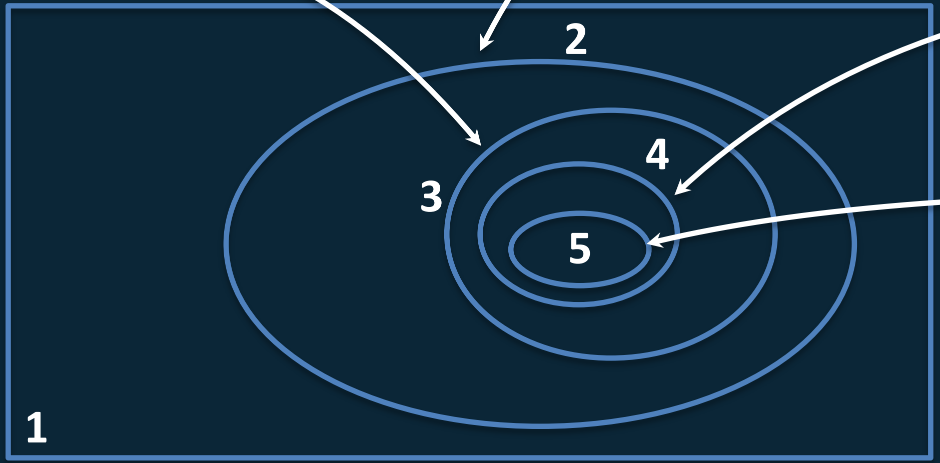
5. Making Optimal Automated Decisions

- Algorithms can process more quantifiable data than humans.
- However the data may be biased/incomplete.
- Individual algorithms also have intrinsic biases.
- Algorithms can therefore be just as fallible as humans!

Available Knowledge & Experience

Knowledge / Experience captured by data

Knowledge / Experience captured by features



Knowledge / Experience captured by algorithm used to make a decision

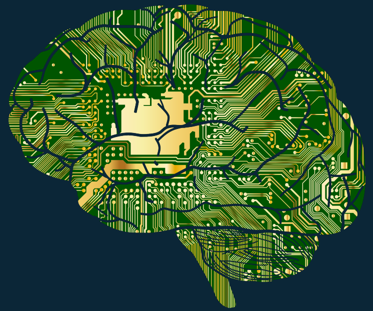
Set of all knowledge on subject *A*

Edge Hill University

6. From Experience to Training Data

</TECHUP_WOMEN>

- Experience is known as "training data".
- An algorithm is ‘taught’ using training data.
- Training data can be labelled or unlabelled.
- We evaluate performance on a disjoint dataset called "test data" - this data must be labelled (ground truth known).



$$Training\ data = \{(x_1, y_1), (x_2\ , y_1), \ldots , (x_n, y_1)\}$$

Durham University · Edge Hill University · The University of Nottingham (United Kingdom · China · Malaysia) · University of York · institute of CODING · Office for Students OfS

Edge Hill University

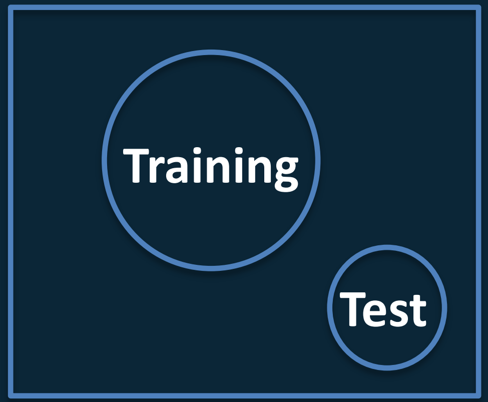
7. Training vs Test Data

</TECHUP_WOMEN>

There is an easy way to understand the difference between training and test data.

- **Training data represents the knowledge given before an exam.**
- **Test data is a disjoint set of information, that represents the exam – we use this to test performance.**

Set of all knowledge on subject *A*



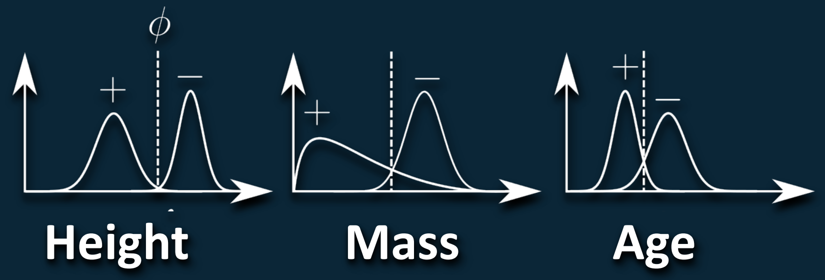
Edge Hill University

8. Features & Labels Revisit

</TECHUP_WOMEN>

Features (Variables)

$$x_i = \{x_i^1 \, , \ldots, x_i^m \}$$



Labels (Feedback)

$$y_i \in \{-1, 0, 1\}$$

e.g. -1 = no disease etc

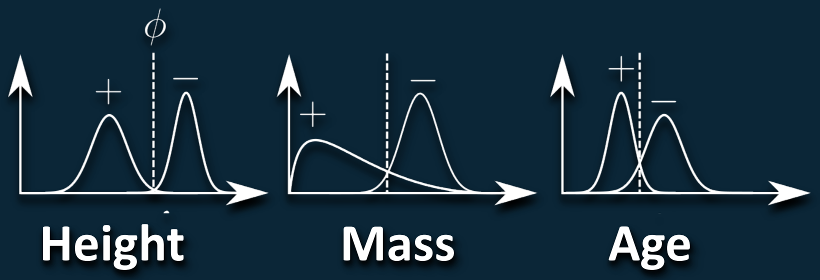
$$Training\ data = \{(x_1, y_1), (x_2\, , y_1), \ldots, (x_n, y_1)\}$$

Edge Hill University

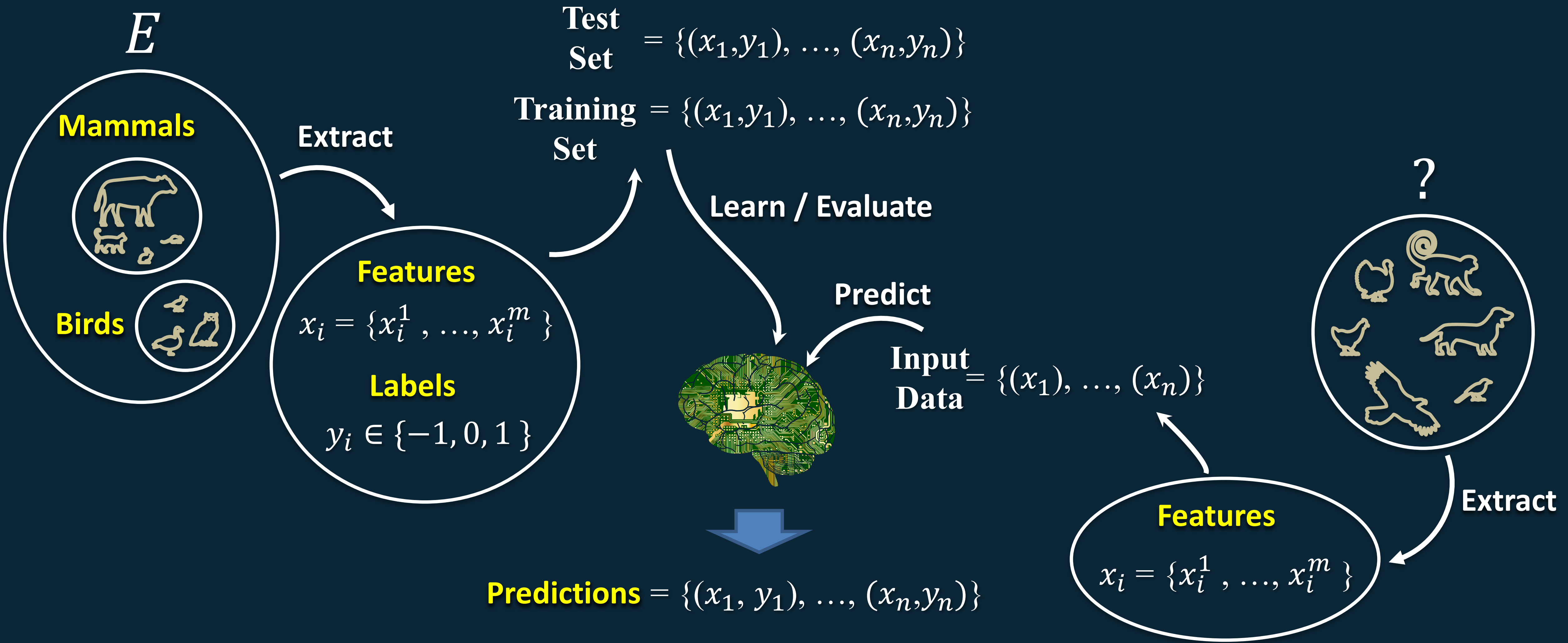
9. Feature Design

</TECHUP_WOMEN>

- In principle feature design involves studying your data.
- Considering it's properties.
- Extracting information you believe will help with class separation.
- In practice this involves,
  1. considering as many candidate features as possible.
  2. Considering their useful ness in turn.
  3. Selecting a sub-set of features you think will be effective.
  4. Testing those, see what happens.
  5. Return to step 1 if results are not as expected.
  6. Collect new data?
  7. Deriving new features?



10. Classification Process



Edge Hill University

11. So How is “Learning” Done?

</TECHUP_WOMEN>

- **Learning via trial and error – learning from mistakes made.**

We need to be at work, but don’t know what time the bus arrives to get us there. Work starts at 09:00, and the journey takes roughly 40 minutes.

We start to wait at the bus stop each day, recording some details.

Day	Arrived at bus stop	Bus arrived
Mon	08:32	08:45
Tue	08:17	08:30
Wed	08:12	08:15

Late for work y
Yes
Yes
No

In this case learning involves finding an arrival time at the bus stop, that lets you make it to work on time. Each time you’re late, you’ve made an error that you learn from.

Edge Hill University

12. So How is “Learning” Done?

</TECHUP_WOMEN>

- In ML error is quantified and minimised using the tools of mathematics.
- This sort of error minimisation is done using functions.

We can reduce finding the best time to wait at the bus stop, to a mathematical problem, i.e. find a value for t that minimizes the number of times you’re late.

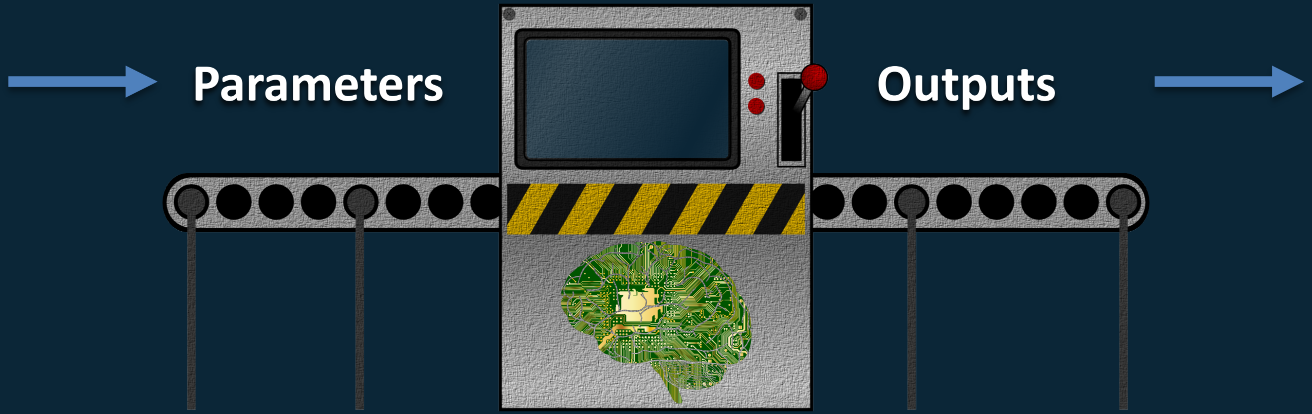
Many potential values for t work, but clearly it must be 08:15 or earlier!

Day	Arrived at bus stop t	Bus arrived
Mon	08:27	08:45
Tue	08:17	08:30
Wed	08:12	08:15

Late for work y
Yes
Yes
No

13. Functions

- Functions can be thought of as simple input/output boxes.
- Machine learning algorithms are functions that ingest data, and produce some output.
- How they do this is hard to convey without some basic mathematics!

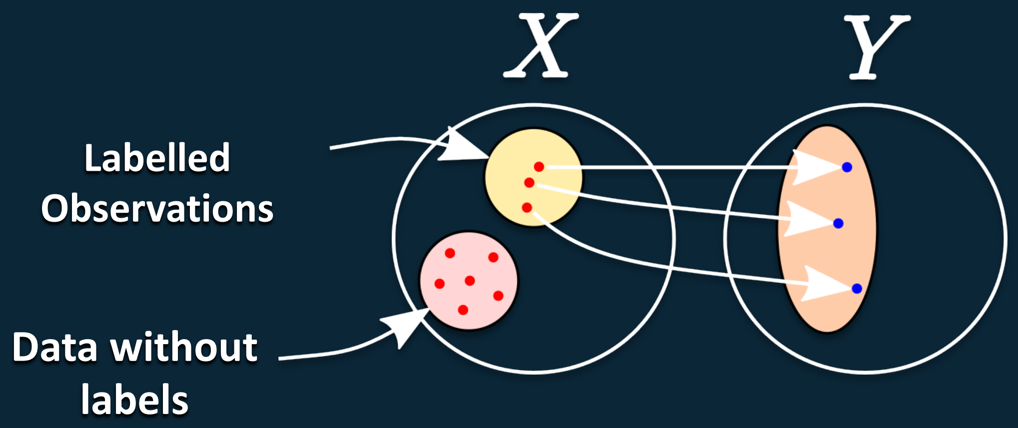


$f(x, y) = x + y$ e.g. $f(x = 2, y = 4) = 2 + 4 = 6$

Edge Hill University

14. Functions That "Map"

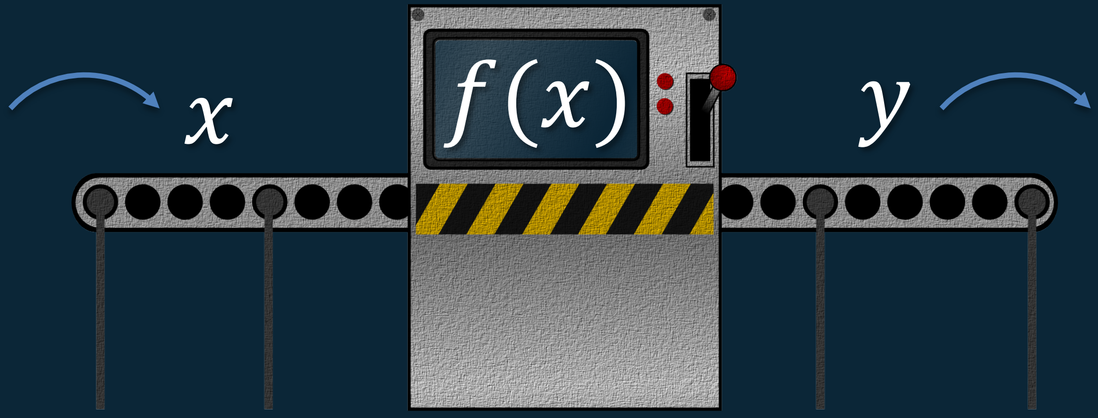
</TECHUP_WOMEN>



E.g.

$$f(x) = \begin{cases} 0, & if\ feature_1 \leq 0 \\ 1, & otherwise \end{cases}$$

$feature_1$	$feature_2$
7.4	29
6.6	24
3.1	11
…	…
0.5	2



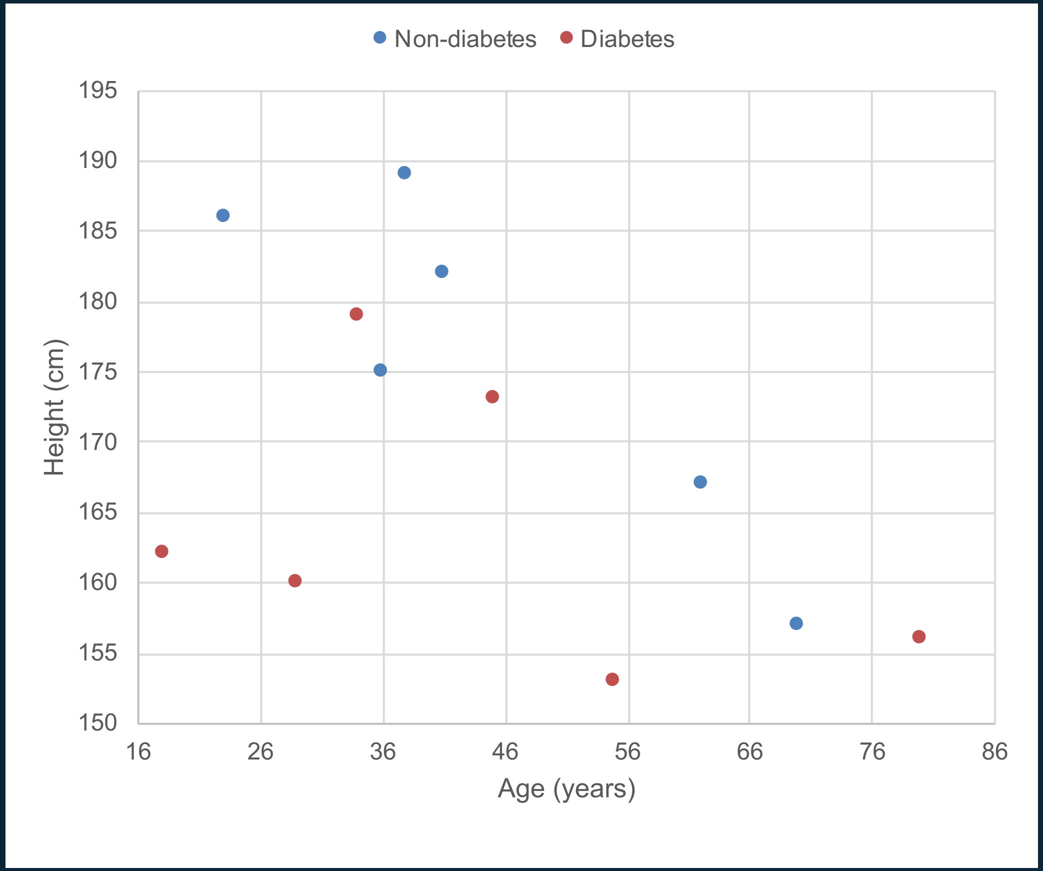
y
1= Diabetic
1= Diabetic
0 = Non-diabetic
…
Unknown

15.Concrete Example

</TECHUP_WOMEN>

ID	Gender	Age	Weight (Kg)	Height (cm)	2 hour glucose	Fasting glucose	Diabetes
x_1	Male	18	52	162	8.2	7.7	1
x_2	Male	23	75	186	7.2	5.1	0
x_3	Female	29	47	160	15.2	13.5	1
x_4	Male	34	80	179	13.1	12.8	1
x_5	Female	36	60	175	7.4	7.5	0
x_6	Male	38	80	189	7.8	7.4	0
x_7	Male	41	94	182	6.2	5.7	0
x_8	Female	45	52	173	12.5	9.7	1
x_9	Female	55	69	153	9.2	9.1	1
x_{10}	Male	62	75	167	5.6	6.6	0
x_{11}	Female	70	50	157	6.2	7	0
x_{12}	Female	80	45	156	10.2	7.9	1

Table 3. Patient data



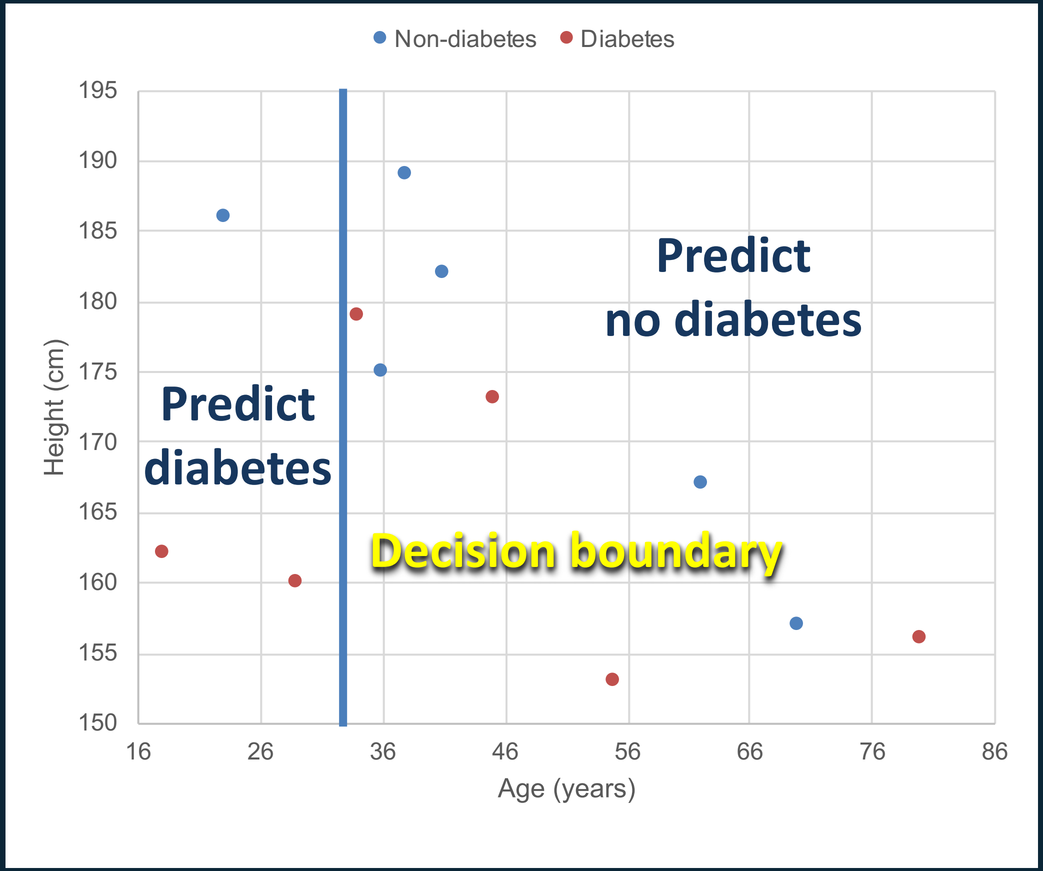
Plot of height vs. age

16. Concrete Example

</TECHUP_WOMEN>

ID	Gender	Age	Weight (kg)	Height (cm)	2 hour glucose	Fasting glucose	Diabetes
x_1	Male	18	52	162	8.2	7.7	1
x_2	Male	23	75	186	7.2	5.1	0
x_3	Female	29	47	160	15.2	13.5	1
x_4	Male	34	80	179	13.1	12.8	1
x_5	Female	36	60	175	7.4	7.5	0
x_6	Male	38	80	189	7.8	7.4	0
x_7	Male	41	94	182	6.2	5.7	0
x_8	Female	45	52	173	12.5	9.7	1
x_9	Female	55	69	153	9.2	9.1	1
x_{10}	Male	62	75	167	5.6	6.6	0
x_{11}	Female	70	50	157	6.2	7	0
x_{12}	Female	80	45	156	10.2	7.9	1

Table 3. Patient data

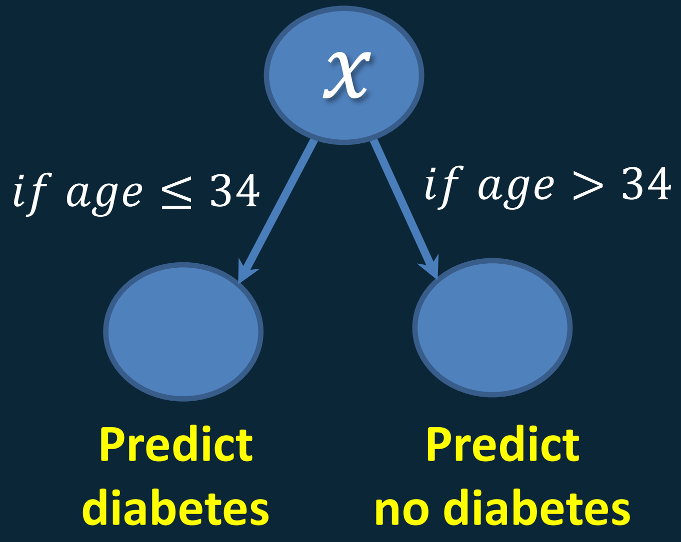


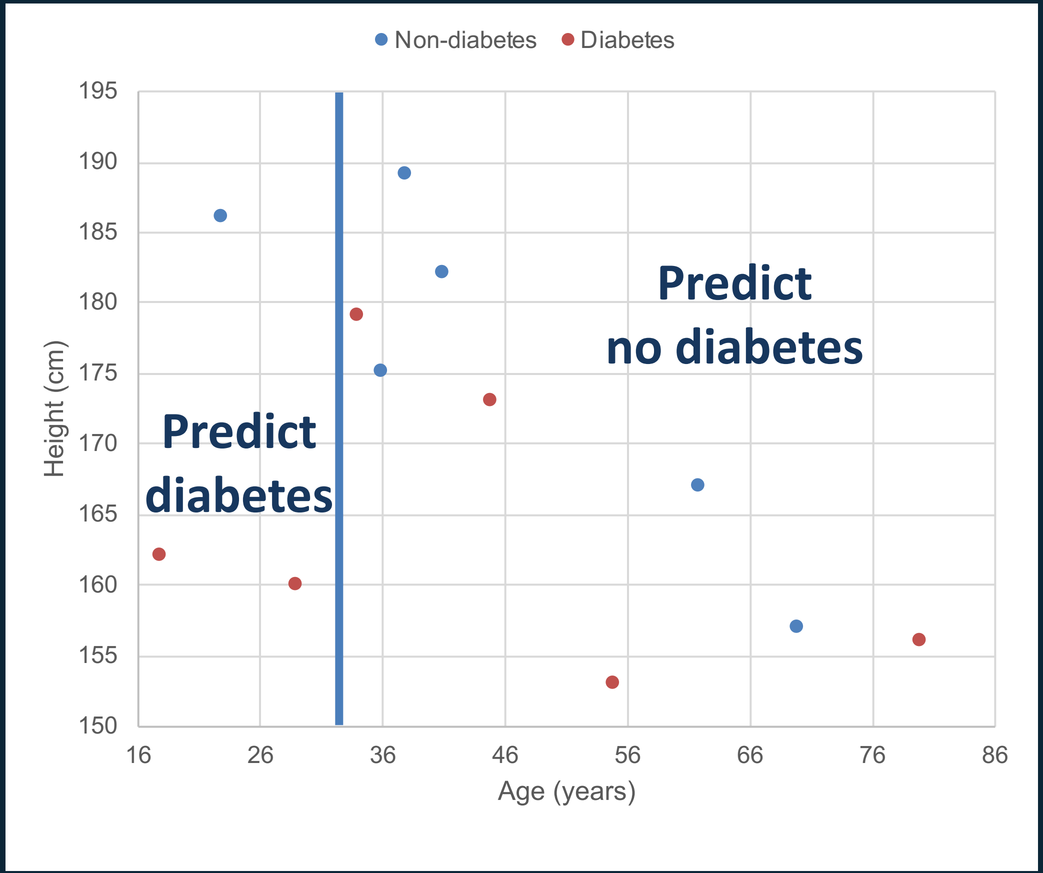
Plot of height vs. age

Edge Hill University

17. Decision Stump

</TECHUP_WOMEN>





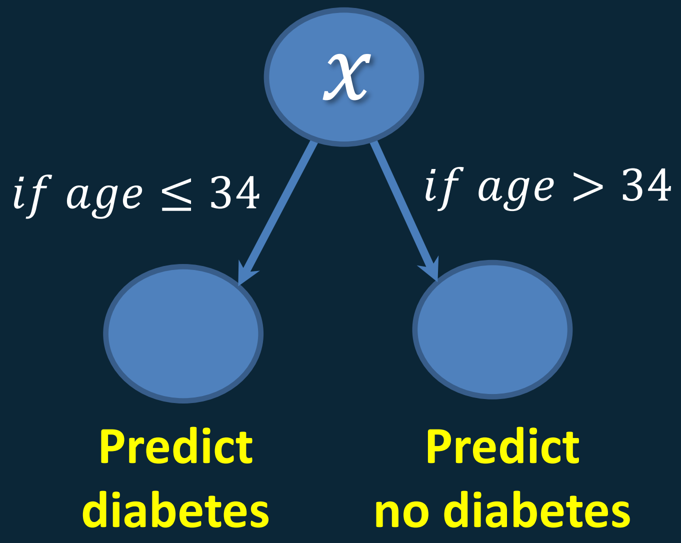
- We’ve just covered our first ML algorithm - the “Decision stump”.
- It is a linear separator, or a linear model.
- It simply looks for a feature to split on, in a optimal way.
- It produces a linear separator between two classes.

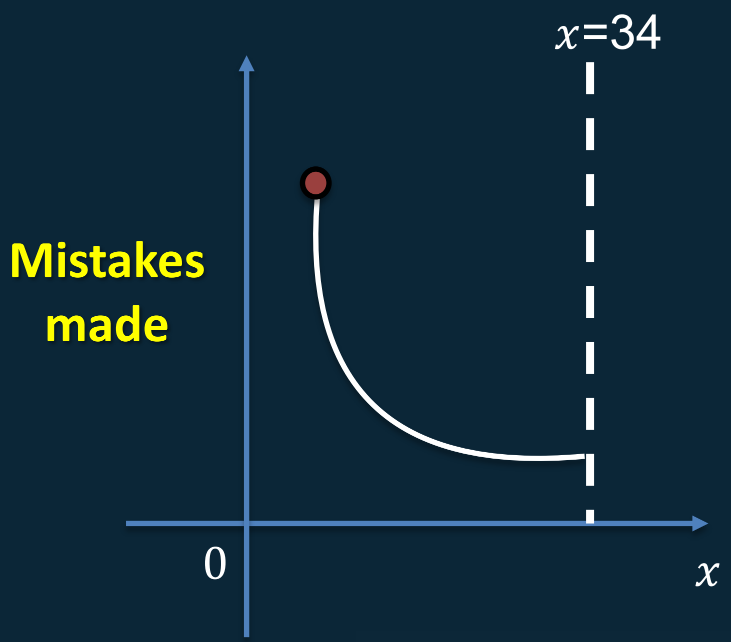
Edge Hill University

18. Decision Stump

</TECHUP_WOMEN>

- **For each feature in the data, the stump will search for a split-point, that minimises the error rate.**
- **In our example we used age as a feature, and our threshold = 34.**



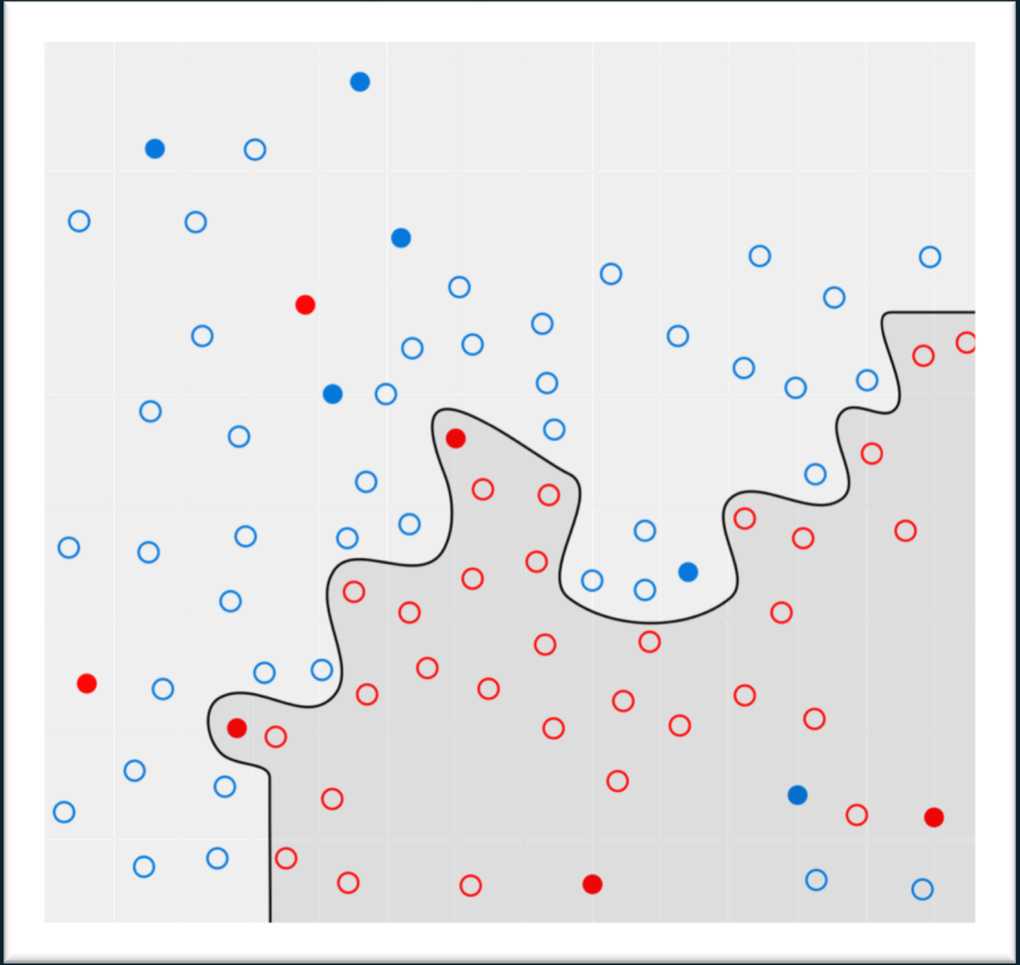


Edge Hill University

19. Decision Boundary

</TECHUP_WOMEN>

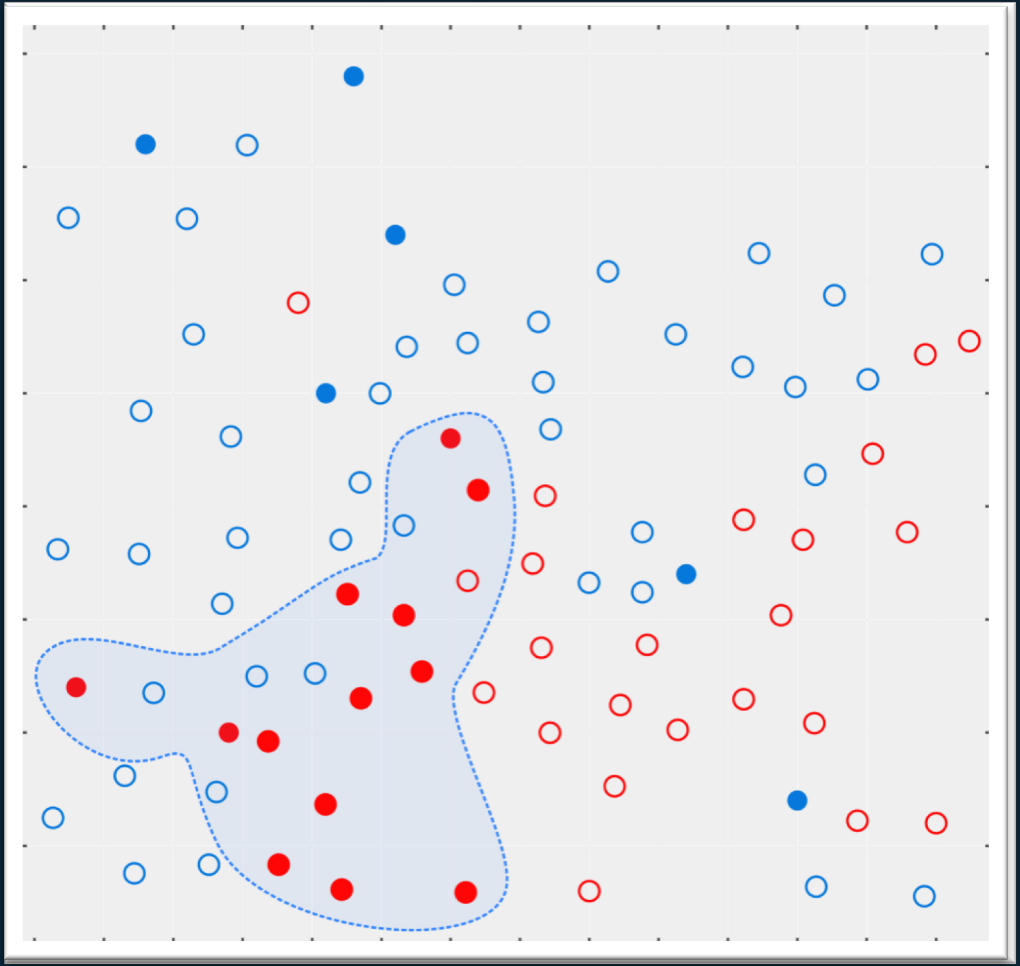
Shaded region predicts diabetes



- You've now seen a linear model in action, and learned what a decision boundary is.
- Real-world problems are complex - so too are the decision boundaries needed to accurately separate data.
- This means linear models don't always work well. More complex ML algorithms are required.
- When we generalise well we can produce effective decision boundaries.
- However sometimes we don't perform well in practice.

19. Decision Boundary

Shaded region predicts diabetes



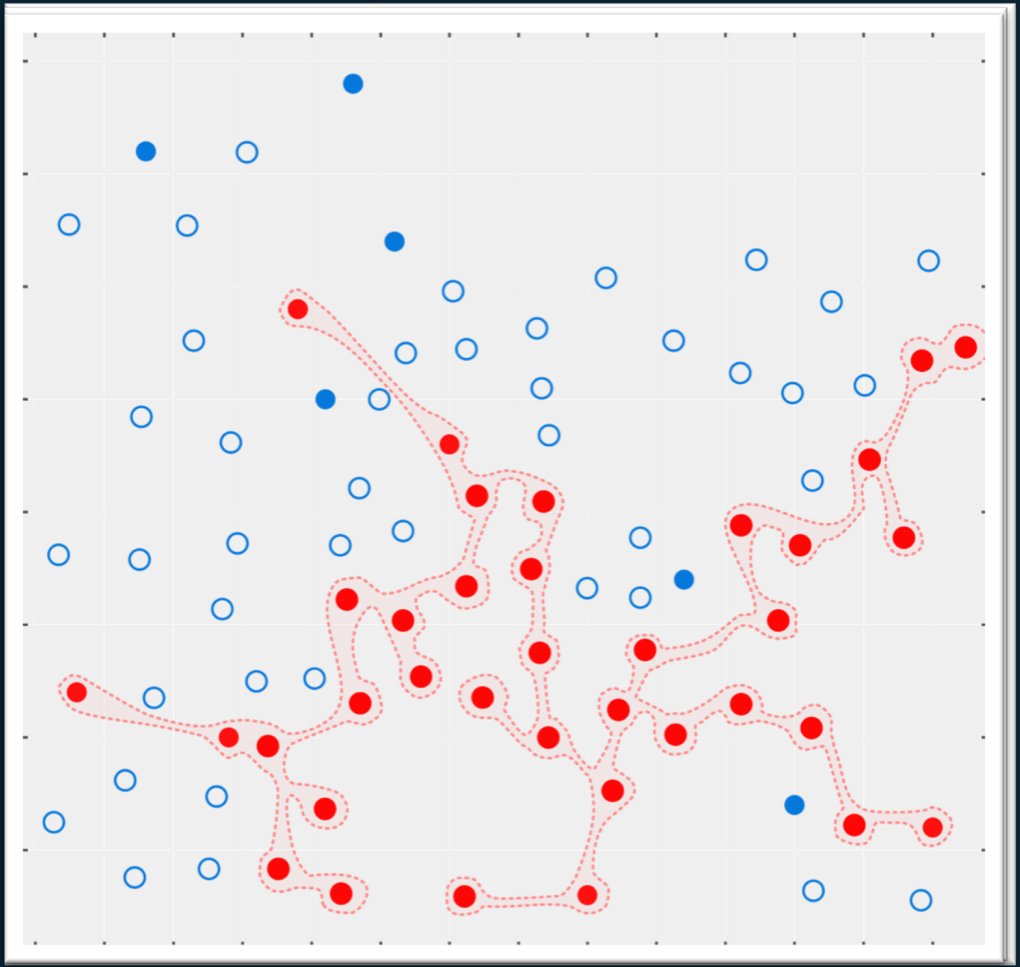
- You've now seen a linear model in action, and learned what a decision boundary is.
- Real-world problems are complex - so too are the decision boundaries needed to accurately separate data.
- This means linear models don't always work well. More complex ML algorithms are required.
- When we generalise well we can produce effective decision boundaries.
- However sometimes we don't perform well in practice.
- We can underfit to our data.

Edge Hill University

19. Decision Boundary

</TECHUP_WOMEN>

Shaded region predicts diabetes



- You've now seen a linear model in action, and learned what a decision boundary is.
- Real-world problems are complex - so too are the decision boundaries needed to accurately separate data.
- This means linear models don't always work well. More complex ML algorithms are required.
- When we generalise well we can produce effective decision boundaries.
- However sometimes we don't perform well in practice.
- We can underfit to our data.
- We overfit to our data.
- Decision boundaries allow us to visualise this happening.

20. Checkpoint

</TECHUP_WOMEN>

We've covered these topics in the last few slides:

- Learning from a training set.
- Using a test set.
- That humans and algorithms suffer from bias, inherent to the data available to them.
- The classification process.
- Functions, and mapping functions in particular.
- The concept of error minimization.
- Decision boundaries.
- Our first ML algorithm, the decision stump.