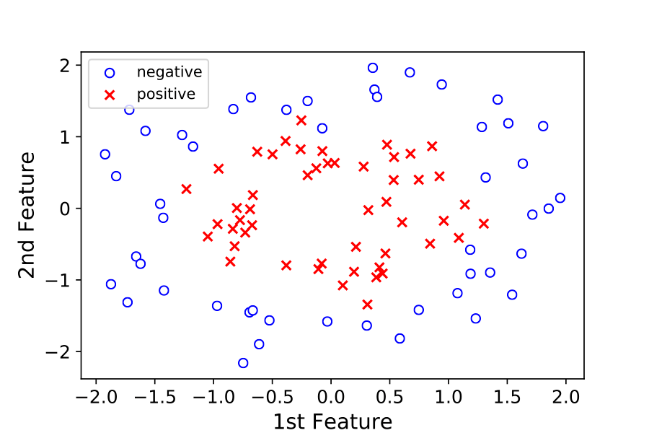
## Week 1a – Introduction to ML

1. Why is ML exciting?
   1. It is involved in many business applications.
   2. It can automate tasks at human-level performance.
2. What are examples of applications?
   1. Playing Go.
   2. Image recognition.
3. Recent ML algorithms can identify classes in ImageNet as accurately as humans.
   1. True

## Week 1b – Logistic Regression

1. What kind of outcomes does logistic regression predict?
   1. Binary.
2. What does the sigmoid function do?
   1. It converts a real-valued number to a probability of a positive outcome.
3. Can logistic regression draw an appropriate decision boundary between the classes of this data?



* 1. No.

## Week 1c – Multilayer Perceptron

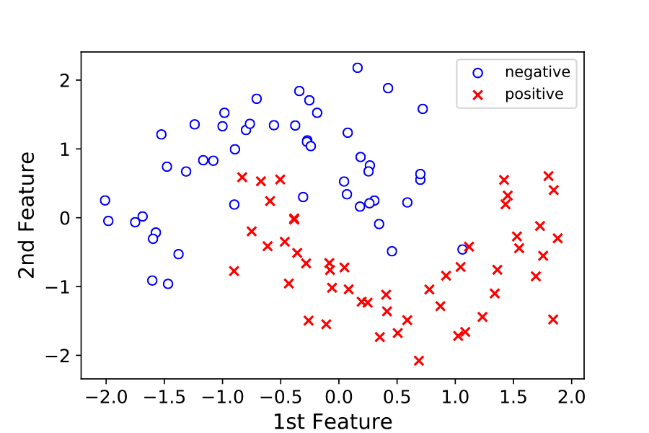
1. What is a difference between logistic regression and multi-layer perceptron models?
   1. Logistic regression maps the feature vector to a single real-valued number, while multilayer perceptron maps the feature vector to multiple real-valued numbers.
2. What is the best synonym for “latest processes”?
   1. Filter
3. What is the best interpretation for this mathematical expression?
   1. Map K latent feature probabilities through a logistic regression.

## Week 1d – Deep Learning

1. What best characterises deep learning?
   1. Multiple layers of latent processes.
2. Using a multi-layer perceptron for document analysis, what is the best interpretation of the layer 1 filters through ?
   1. Topics, such as sports or history.
3. Using a multi-layer perceptron for document analysis, what is the best interpretation of the layer 2 filters through ?
   1. Meta-topics, such as history of sports or politics of sports.

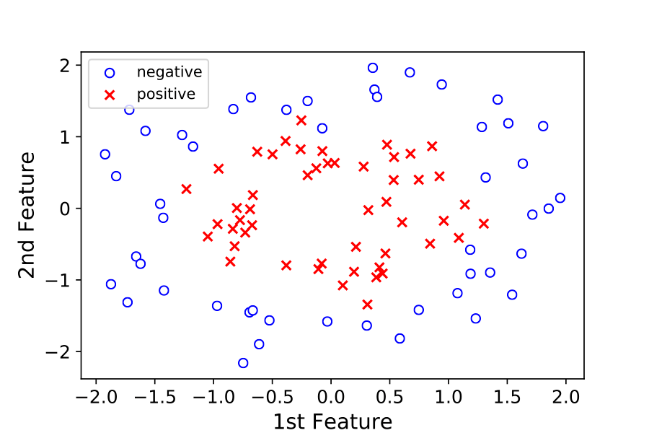
## Week 1e – Model Selection

1. Given the following image of data classifications, what model would you choose?



* 1. Multi-layer perceptron.

1. Given the following image of data classifications, what model would you choose?



* 1. Multi-layer perceptron.

1. Given different sets of training data, in which model would you expect more variance in the learned parameters?
   1. Multi-layer perceptron.

## Week 1f – History of Neural Networks

1. When was the multi-layer perceptron introduced?
   1. 1960.
2. Why didn’t neural networks perform “well” in the early 1990s?
   1. Insufficient training data.
3. What contributed to neural networks outperforming humans in the task of image analysis?
   1. Large set of training data.
   2. New, deeper networks.
   3. GPU computational platform.

## Week 1g – CNN Concepts

1. What model, when used for image classification, can exceed the performance of humans?
   1. Convolutional neural network.
2. What best describes the hierarchical structure of images, listed from most complex to simplest?
   1. High-level motifs, sub-motifs and atomic elements.
3. In the example CNN, what allows a classification decision to be made?
   1. Layer 3 feature map.

## Week 1h – CNN Math Model

1. What is the outcome of convolving the layer 2 filters with the layer 1 feature maps?
   1. Layer 2 feature maps.
2. In general terms, the basic steps to do learning are:
   1. Obtain a large set of labelled data.
   2. Determine the loss function, which computes loss between the true label and model label.
   3. Determine parameters that minimise the sum over loss.
3. Why is learning hard?
   1. The parameter space is large.
   2. There can be many local minima, even for two parameters.
   3. The parameters are computationally expensive to compute.

## Week 1i – Applications in Use and Practice

1. In the experiment where light was shined on mammalian retinas, what were found to be (almost exactly) the fundamental shapes that excited exactly one neuron?
   1. Layer 1 filters that CNN learns
2. In the example network, which layer of filter might have the shape of an eye in a CNN trained on real image data?
   1. Layer 2 filter
3. Which of the following are example of deep CNNs given in the videos?
   1. Digit recognition
   2. Playing Go
   3. Classifying images
   4. Captioning images

## Week 2a – Logistic Regression as a Running Example

1. What is the purpose of a loss function?
   1. To define a penalty for poor predictions
2. How is the loss function defined?
   1. Negative log-likelihood
3. In the polynomial fitting example, which one of the following is an example of overfitting?
   1. Eighth order polynomial
4. What is the “gold standard” validation strategy?
   1. Try on new real-world data
5. When existing data is used to validate performance, into which of the following groups is data split?
   1. Validation set
   2. Test set
   3. Training set