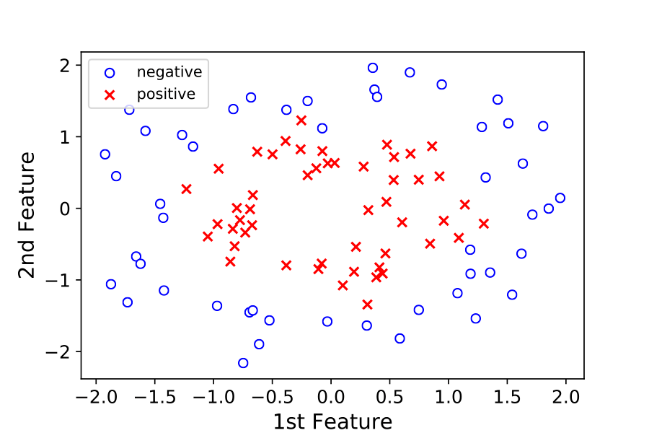
## Week 1a – Introduction to ML

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

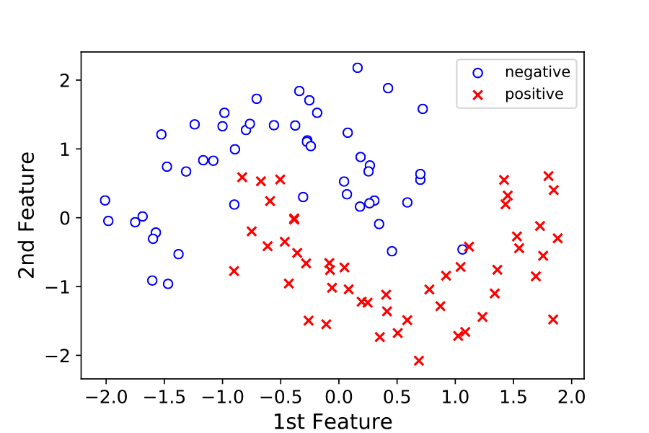
1. What is a difference between logistic regression and multi-layer perceptron models?
   1. Logistic regression maps the feature vector to a single real-valued number, while multilayer perceptron maps the feature vector to multiple real-valued numbers.
2. What is the best synonym for “latest processes”?
   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.

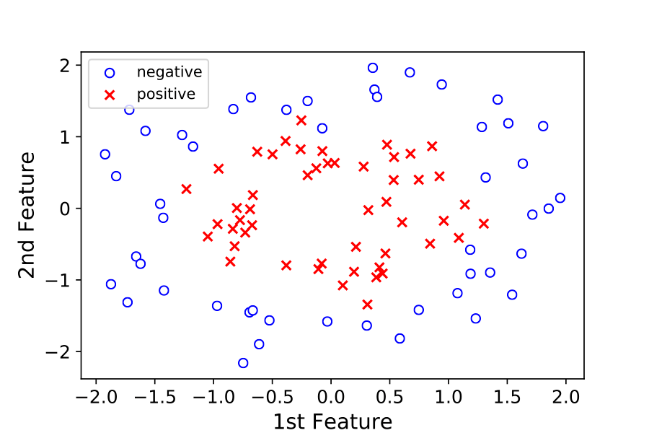
## Week 1e – Model Selection

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

## Week 2b – Learning via Gradient Descent

1. What best describes one iteration of gradient descent?
   1. Find the slope at the current point. Step in the direction that is “downhill”.
2. What is the name for multi-dimensional slope?
   1. Gradient.
3. What is the primary difference between stochastic gradient descent and gradient descent?
   1. Stochastic gradient descent estimates the gradient by choosing a data point at random.
4. What is a minibatch?
   1. Running a few data examples rather than all of the data.
5. What is early stopping?
   1. During an optimisation loop check the validation loss and stop when it stops moving.

## Week 3a – CNN Basics

1. What are sensitivity and specificity in the context of medical image diagnosis?
   1. Sensitivity is the ratio of true positives found to positives in the dataset, and specificity is the ratio of true negatives found to negatives in the dataset.
2. In the figure from Gulshan et. al., how did the machine perform relative to the opthamologists for the task of diagnosing diabetic retinopathy?
   1. The machine outperformed 80% of ophthalmologists.
3. A high value in the convolution indicates a **matched** feature.
4. What is the result of convolving a filter with a two-dimensional image?
   1. A heat map, where bright spots indicate feature detection.

## Week 3b – Core Components of the Network

1. What is meant by “volume” of feature maps?
   1. Combining all of the feature maps for different features into a stack.
2. How is color accounted for in a filter?
   1. Three color channels are separated to create a volume filter.
3. Which of the following are example activation functions?
   1. Rectified linear unit
   2. Sigmoid function
   3. Hyperbolic tangent function
4. Which of the following operations are used in pooling?
   1. Maximum
   2. Mean
5. What is a fully connected layer?
   1. A layer in which each data point is connected via a weight to the upstream elements.

## Week 3c – CNN Implementation

1. What is the goal of training the network?
   1. To learn filter parameters and the final readout weights that minimize the loss function.
2. Which of the following is an example of leveraging all training data given in the lesson?
   1. Learning about the digit 9 by looking at the digit 7.
3. What is transfer learning?
   1. Taking a network trained on a large database to do a classification task, the doing additional training in the domain of interest.

## Week 4a – Word Embeddings

1. What analogy highlights the purpose of word vectors?
   1. Two places near each other on a map are likely to have similar characteristics.
2. What dimension is a typical word vector?
   1. More then two dimensions
3. The result of convolving K filters with a text is N K-dimensional vectors. What is the result of a max pooling step?
   1. A K-dimensional vector
4. In the example shown, what would be done after the max pooling step?
   1. Use existing methods to make a classification decision.

## Week 4b – Sentiment Analysis

1. Which model is used in the example?
   1. Multilayer perceptron
2. What is the softmax function?
   1. A generalization of the logistic function.
3. When does the softmax function reduce to the logistic function?
   1. When V = 2
4. What is the goal of both of CBOW and Skip-Gram?
   1. To learn the model parameters from large corpus of text without human labelling.
5. What function is being maximised to learn the parameters of the model?
   1. The sum of the logs of probabilities for each word.

## Week 4c – Introduction to Modern NLP Methods and Reinforcement Learning

1. What two entities are concatenated and sent into a neural network?
   1. The vector associated with the previous word and the hidden vector of the previous network
2. How many control neural networks are used in LSTM?
   1. 3
3. What best describes what the f\_n neural network does?
   1. Controls the degree of “forgetting” of the old memory cell.
4. In the example of figure captioning, how are the parameters of the LSTM model initialised?
   1. From the features at the top of a CNN that analysed the image.