



Machine Learning Emerging & Trends

Name : Haoyi Sun

Student Number : 1904101

Module: Emerging and future trends

24/05/2020

Contents

1. Background	3
2. Machine Learning.....	4
2.1 The scope of machine learning.....	4
3. Machine Learning Approaches	7
3.1 Regression algorithm	7
3.2 Neural networks	7
3.3 SVM (Support Vector Machine).....	8
3.4 Clustering algorithm.....	10
3.5 Downside algorithm.....	10
3.6 Recommendation algorithm	10
3.7 Other algorithms.....	11
4. Classification of machine learning	12
5. Evaluation of machine learning models	13
6. Emerging & Future Trends	14
6.1 Deeping Learning	14
6.2 Weight Agnostic Neural Networks	15
7. References.....	18

Figure 1 Venn diagram

Figure 2 The scope of machine learning

Figure 3 Single-hidden Layer Feedforward Neural Networks (SLFNs) Structure

Figure 4 Support Vector Machine Structure

Figure 5 SVM Knowledge Framework Thinking Map

Figure 6 General Neural Networks & Weight Agnostic Neural Networks

Figure 7 Examples of WANN

1. Background

The first thing is to briefly distinguish a few concepts: artificial intelligence, machine learning, deep learning, neural networks. These are the words that should come up the most often, but what difference do they make?

Artificial Intelligence: problems that humans can solve through intuition, such as natural language understanding, image recognition, speech recognition, etc., are difficult for computers to solve, and AI is all about solving such problems.

Machine learning: Machine learning is a method that empowers a machine to perform functions that cannot be done by direct programming. But in a practical sense, machine learning is a way to use data, train a model, and then use the model to make predictions.

Deep learning: at its core, it's about automatically combining simple features into more complex features and using those features to solve problems.

Neural networks: originally a biological concept, generally referring to the network of neurons, contacts, cells, etc. in the brain that are used to generate consciousness and help organisms think and act, later artificial intelligence was inspired by neural networks and developed artificial neural networks.

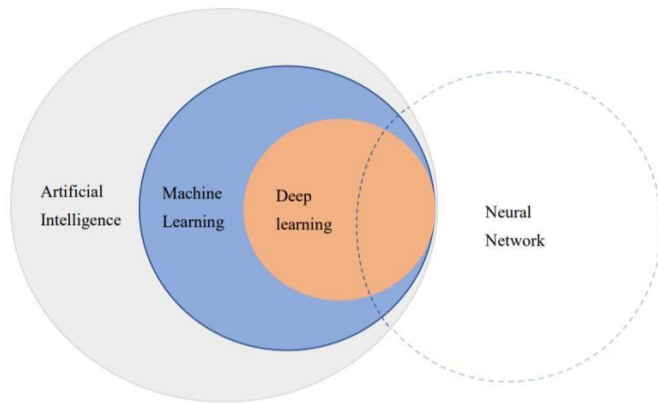


Figure 1 Venn diagram

2. Machine Learning

2.1 The scope of machine learning

Machine learning is deeply connected to the fields of pattern recognition, statistical learning, data mining, computer vision, speech recognition, natural language processing, etc.

Commented [GD1]: USE REFERENCES!!!!

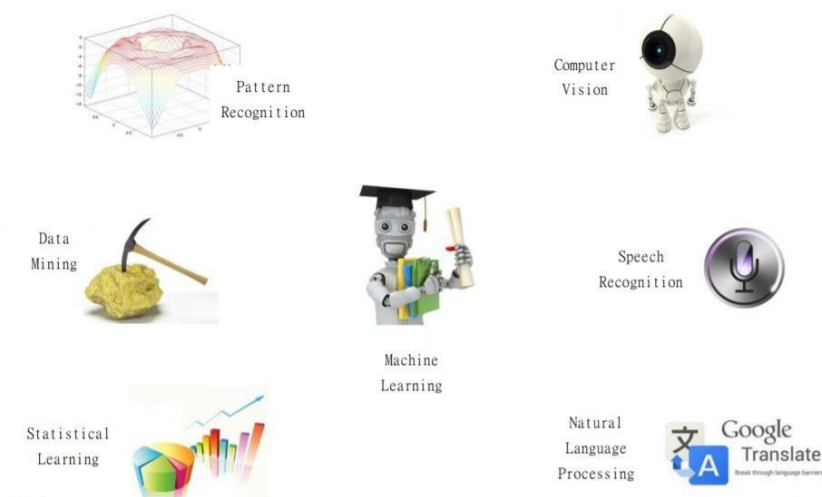


Figure 2 The scope of machine learning

①Pattern Recognition

Pattern recognition is the process of making various analytical judgments based on the alignment of the input raw data to obtain its category attributes, characteristics judgments. In order to have this ability, humans have gradually evolved a highly complex neural and cognitive system over the past tens of millions of years through the knowledge and understanding of a vast number of things. For example, we can easily tell which is a key, which is a lock, which is a bike, which is a motorcycle; and behind these seemingly simple processes, there are actually very complex processing mechanisms. And figuring out the mechanism by which these mechanisms act is the basic task of pattern recognition.

②Data Mining

Data Mining refers to the automated process of classifying large data sets to identify trends and patterns and build relationships to solve business problems through data analysis. Data mining is used to address the scalability of data and to increase or change the scalability of data. The second is to address the high dimensionality of the data. Handling heterogeneous and complex data. Address data ownership and distribution issues. A rational treatment of non-traditional analysis. In other words, data mining is the process of extracting information and knowledge implicit in, and unknown to people in advance, but potentially useful, from large, incomplete, noisy, fuzzy, random amounts of data.

Commented [GD2]: USE REFERENCES!!!!

③Statistical Learning

Statistical learning is a discipline that builds probabilistic statistical models based on data and uses the models to make predictions and analyse the data, also known as statistical machine learning. The purpose of statistical learning is to make predictions and analysis of data; statistical learning is method centered, and statistical learning methods build models and apply them to make predictions and analysis; statistical learning is a cross-cutting discipline of probability theory, statistics, information theory, computational theory, optimization theory, and computer science, and it has gradually formed its own theoretical system and methodology during development. The object of statistical learning is data. It starts with the data, extracts the characteristics of the data, abstracts the model of the data, discovers the knowledge in the data, and returns to the analysis and prediction of the data.

④Computer Vision

Computer vision is the study of the organization of image information, identification of objects and scenes, and consequently the interpretation of events, using images (video) as input and with the goal of expressing and understanding the environment. From the current state of research, the focus is still on the organization and identification of image information, and the interpretation of events is still very little involved, at least at a very preliminary stage. At present, the application of machine vision has become more and more mature in recent years, from daily supermarket barcode detection, commuting fingerprint attendance, face recognition attendance, highway license plate recognition, aviation remote sensing detection and control of terrain and landforms, film special effects production, industrial production automation inspection, medical image inspection, and aviation astronomy field, etc., the application has gradually matured.

⑤Speech Recognition

The main method of speech recognition is the pattern matching method. During the training phase, the user says each word in the vocabulary in turn and deposits its feature vector as a template into the template library. In the recognition phase, the feature vectors of the input speech are compared with each template in the template library in order of similarity, and the one with the highest similarity is output as the recognition result. In recent years, deep learning research in the machine learning field has been introduced to the training of acoustic models for speech recognition using multi-layered neural networks with RBM pre-training, greatly improving the accuracy of acoustic models. In this regard, researchers at Microsoft pioneered a breakthrough when they used the Deep Neural Network (DNN) model to reduce speech recognition error rates by 30 percent, the fastest advance in speech recognition technology in nearly 20 years.

Commented [GD3]: USE REFERENCES!!!!

⑥Natural Language Processing

Natural Language Processing explores how to process and use natural language, especially how to program computers to successfully process large amounts of natural language data. Basic tasks include regular expressions, participles, lexical analysis, speech recognition, text classification, information retrieval, question and answer systems - such as answering some questions or interacting with the user - machine translation, etc. Commonly used models are Markov model, Park Bayesian, circulating neural network, etc. The goal of natural language processing is to bridge the gap between human communication (natural language) and computer understanding

(machine language), and ultimately to achieve a computer that is as intelligent as humans in understanding natural language. In the future, the development of natural language processing will enable artificial intelligence to gradually face more complex situations and solve more problems, and will certainly bring us a more intelligent era.

Commented [GD4]:

3. Machine Learning Approaches

3.1 Regression algorithm

In most machine learning courses, the regression algorithm is the first algorithm introduced. There are two reasons for this: one. The regression algorithm is relatively simple, introducing it allows a smooth migration from statistics to machine learning. ii. Regression algorithms are the cornerstone of several powerful algorithms behind them, and it is impossible to learn those powerful algorithms without understanding them. There are two important subcategories of regression algorithms: namely linear and logistic regression.

In terms of implementation, logistic regression is just the addition of a Sigmoid function to the linear regression result, which translates the numerical result into a probability between 0 and 1 (the image of the Sigmoid function is generally not intuitive, you just need to understand that the larger the log value, the closer the function is to 1, the smaller the value, the closer the function is to 0), then we can make predictions based on this probability, for example, if the probability is greater than 0.5, then the email is spam, or whether the tumour is malignant, etc. Intuitively, logistic regression draws a line of classification. The logistic regression algorithm draws essentially linear classification lines (there are also logistic regressions that draw non-linear classification lines, but such models are inefficient when dealing with large data volumes), which means that when the boundary between the two categories is not linear, the logistic regression is less expressive.

3.2 Neural networks

Let's look at the logical architecture of a simple neural network. In this network, it is divided into an input layer, a hidden layer, and an output layer. The input layer is responsible for receiving signals, the hidden layer is responsible for decomposing and processing the data, and the result is integrated into the output layer. A circle in each

layer represents a processing unit, which can be thought of as simulating a neuron, with several processing units forming a layer and several layers forming a network, or "neural network".

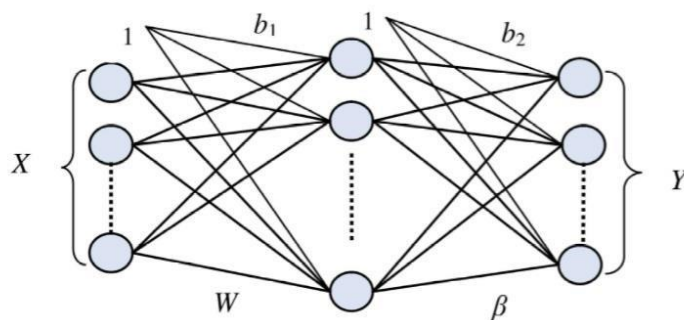


Figure 3 Single-hidden Layer Feedforward Neural Networks (SLFNs) Structure

In a neural network, each processing unit is in fact a logistic regression model, which receives inputs from the upper level and transmits the predicted results of the model as outputs to the next level. Through this process, the neural network can complete very complex nonlinear classification.

3.3 SVM (Support Vector Machine)

Support for vector machine algorithms is in a sense an enhancement of logistic regression algorithms: by giving more stringent optimization conditions to logistic regression algorithms, support for vector machine algorithms can obtain better classification boundaries than logistic regression. But without some specific function technique, the support vector machine algorithm is at best a better linear classification technique.

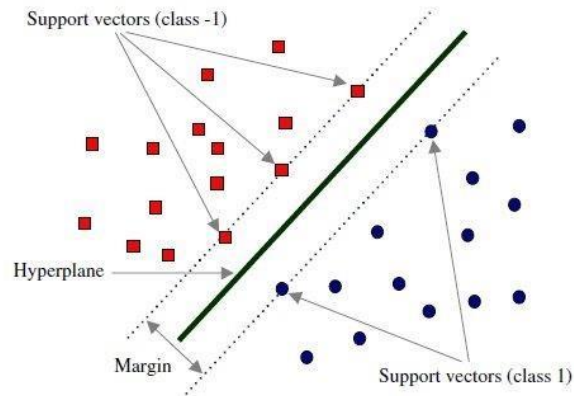


Figure 4 Support Vector Machine Structure

The basic idea of SVM learning is to solve for separate hyperplanes that correctly divide the training data set with the largest geometric interval. As shown in the figure, the middle partition line is the separated hyperplane, and for a linearly separable data set, there are an infinite number of such hyperplanes (i.e., perceptual machines), but the separated hyperplane with the largest geometric interval is unique.

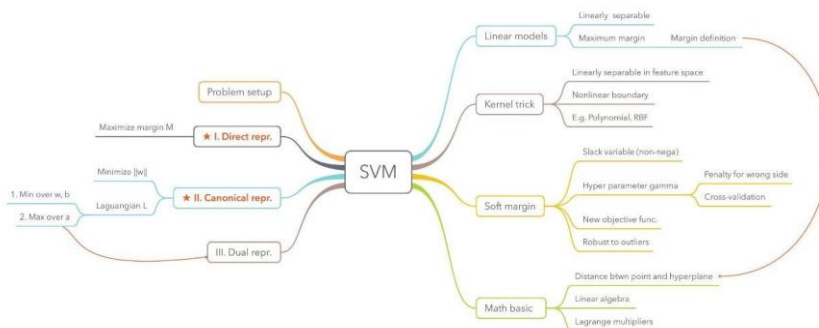


Figure 5 SVM Knowledge Framework Thinking Map

It is called a support vector machine because the core idea is that the support vector sample will play a critical role in the problem identified. So what is a support vector? The support vector is also the nearest sample point to the classification hyperplane.

The optimal partition plane (hyperplane) in the SVM is the maximum value that allows the minimum distance between the support vector and the hyperplane; the goal is to find a hyperplane so that points closer to the hyperplane can have a greater distance.

That is, regardless of the fact that all points must be away from the hyperplane, the point is that the resulting hyperplane allows the nearest of all points to have the maximum distance between them. The role of the kernel function is actually very simple to map the low dimension to the high dimension for easy classification.

3.4 Clustering algorithm

The most typical representation of unsupervised algorithms is the clustering algorithm. Let's still take a two-dimensional data, where a certain data contains two features. I'd like to use the clustering algorithm to label the different species in them, how do I do that? In simple terms, the clustering algorithm calculates the distances in a population and divides the data into multiple populations based on how close they are. The most typical representative of clustering algorithms is the K-Means algorithm.

3.5 Downside algorithm

A reduced dimensional algorithm is also an unsupervised learning algorithm whose main feature is the reduction of data from a higher dimension to a lower dimensional level.

In machine learning, it is common to encounter high-dimensional data sets, and in highdimensional data scenarios there are difficulties such as sparse data samples, distance computation, and so on, which are serious problems common to all machine learning methods, called "dimensional disasters". In high-dimensional features there is a tendency for linear correlation between features, which means that some features are redundantly present. Based on these issues, the idea of descending dimensions emerged. There are many methods of downscaling, and they are divided into linear downscaling and nonlinear downscaling, the common methods being singularity decomposition and principal component analysis methods.

3.6 Recommendation algorithm

The recommendation algorithm is a very popular algorithm in the industry and is widely used in the e-commerce industry, such as Amazon, Tesco, Adidas, etc. The main feature of the recommendation algorithm is that it can automatically recommend to the user

what they are most interested in, thus increasing the purchase rate and enhancing the benefits. There are two main categories of recommendation algorithms.

A category is based on the content of the item recommended, is to recommend to the user with the user to buy similar items, such as the premise that each item has to have a number of labels, so that the user can find items similar to the purchase of items, so that the recommended benefit is a greater degree of association, but because each item needs to be labelled, so the workload is greater.

Another type of recommendation is based on user similarity, which is to recommend to the target user what other users who have the same interest as the target user have purchased, for example, Xiao A has historically purchased items B and C. After the algorithm analysis, it is found that another user who is similar to Xiao A, Xiao D, has purchased item E, and therefore recommends item E to Xiao A.

Both types of recommendations have their own advantages and disadvantages, and in general it's a mix of the two in an e-commerce application. The most famous of the recommended algorithms is the collaborative filtering algorithm.

3.7 Other algorithms

In addition to the above algorithms, there are other algorithms in the machine learning community such as Gaussian discriminant, Park Bayesian, decision tree, etc. But the six algorithms listed above are typical of the most used, most influential, and most varied. One of the characteristics of the machine learning world is that there are many algorithms, and a hundred different developments.

The following is a summary, according to the training data with or without labels, the above algorithm can be divided into supervised learning algorithm and unsupervised learning algorithm, but the recommended algorithm is more special, neither supervised learning nor unsupervised learning, is a separate category.

Supervised learning algorithms: Linear regression, logistic regression, neural networks, SVM. Unsupervised learning algorithms: Clustering algorithms, downscaling algorithms. Special algorithms: recommendation algorithm.

In addition to these algorithms, there are some algorithms whose names appear frequently in the machine learning field. But they aren't really a machine learning algorithm per se, they were born to solve some subproblem. You can understand them as sub-algorithms of the above algorithm, used to drastically improve the training process. Some of the representatives are: gradient descent method, mainly used in linear regression, logistic regression, neural network, recommended algorithm; Newton's method, mainly used in linear regression; BP algorithm, mainly used in neural network; SMO algorithm, mainly used in SVM.

4. Classification of machine learning

The current mainstream of machine learning is divided into: supervised learning, unsupervised learning, and reinforcement learning.

a) Supervised machine learning is also known as "learning with a teacher" and the so-called teacher is the label. The supervised process is to first train through known training samples (e.g., known inputs and corresponding outputs) to obtain an optimal model, which is then applied to the new data and mapped to the output. After going through this process again, the model has predictive power. Typical supervised learning is: in a classification problem, input a picture and output the classification of the picture; in target detection, input a picture and output the border of the target object; in semantic segmentation, label each pixel.

b) The data for unsupervised learning is unlabelled. Typical examples of unsupervised learning are: clustering - for a large number of unknown labelled datasets, dividing the dataset into categories based on the intrinsic similarity of the data, so that the data within categories are more similar and the data between categories are less similar. Downscaling - find some axes on which the variance of the training data is greatest. These axes are the underlying structures within the data. We can use these axes to reduce the data dimensions, and the data has a large variance in each preserved dimension. Density estimation - estimating the intrinsic distribution of the data, with some one-dimensional points at the top, we fit this density distribution using a Gaussian function.

c) Reinforcement learning is often used in robotics (e.g., robotic dogs) where it receives the current state of the robot and the goal of the algorithm is to train the

machine to perform various specific behaviours. The workflow is mostly: the machine is placed in a specific environment where the machine can continuously train itself, and the environment gives either positive or negative feedback. The machine will gain from previous action experience and ultimately find the best content of knowledge to help it make the most effective behavioural decisions.

5. Evaluation of machine learning models

There are three commonly used evaluation metrics: accuracy, recall and Precision.

a) Accuracy is the proportion of correctly classified samples to the total number of samples.

$$Accuracy = \frac{n_{correct}}{n_{total}}$$

$n_{correct}$ is the number of samples correctly classified, n_{total} is the total number of samples.

Limitations of accuracy: Accuracy is the simplest and most intuitive evaluation indicator of classification problems, but there are obvious drawbacks in that when the sample proportions of different general categories are very uneven, the large category is often the most important factor affecting accuracy. For example, when the negative sample is 99%, the classifier predicts a 99% accuracy rate for all negative samples, in other words, a high overall accuracy rate does not mean a high accuracy rate for a small proportion of categories.

b) Precision is the ratio of the number of correctly classified positive samples to the number of samples that the classifier determines to be positive.

c) Recall rate is the proportion of the number of correctly classified positive samples to the number of true positive samples.

Precision value and Recall value are two contradictory and unified indicators, in order to improve the Precision value, the classifier needs to try to predict the sample as a positive sample when it is 'more confident', but currently, it is often too conservative and misses many 'unconfirmed' positive samples, resulting in a lower Recall value.

In a sorting problem, there is usually no definite threshold that directly determines the result to be a positive or negative sample, but instead uses the Precision and Recall values of the TopN return result to measure the performance of the sorting model, i.e., consider the TopN result returned by the model to be the positive sample determined by the model, and then calculate the Precision at N positions and the Recall at the top N positions.

6. Emerging & Future Trends

6.1 Deeping Learning

In 2006, Geoffrey Hinton published [an](#) article in the scientific journal Science that argued two points.

Commented [GD5]: Ok but show the ref in your references/bibliography section

a) Neural networks with multiple [cryptic](#) layers have excellent feature learning capabilities, and the learned features have a more intrinsic portrayal of the data, thus facilitating visualization or classification.

Commented [GD6]: ?

b) The difficulty of training deep neural networks can be effectively overcome by "layer-by-layer initialization".

With such findings, not only the computational difficulty of neural networks is solved, but also the excellence of deep neural networks in learning is illustrated. Since then, neural networks have re-emerged as the dominant and powerful learning technology in the machine learning community. Meanwhile, neural networks with multiple hidden layers are called deep neural networks, and learning studies based on deep neural networks are called deep learning.

At present, many of the industry's image recognition technology and voice recognition technology advances are derived from the development of deep learning, in addition to the Cortana and other voice assistants mentioned at the beginning of this article, but also include some image recognition applications, the typical representative of which is the following figure Baidu mapping function.

Deep learning belongs to a subcategory of machine learning. The development of depth-based learning has greatly contributed to the elevated status of machine learning and, even further, to the industry's renewed focus on the AI dream of its parent class).

The essence of deep learning is to learn more useful features by building machine learning models with many hidden layers and huge amounts of training data, which ultimately improves the accuracy of classification or prediction. Thus, "depth models" are the means and "feature learning" is the end.

6.2 Weight Agnostic Neural Networks

In 2019, A paper by Google AI David Ha et al titled "Weight Agnostic Neural Networks" sparked the machine learning community. Its "subversive" theory makes one exclaim: "At the end of the day we know nothing about neural networks?"

Commented [GD7]: make ref here to the paper included bibliography

Some researchers on Reddit have argued that the more interesting significance of the Weight Agnostic Neural Networks paper is that it also proclaims the end of the life of the explanation of deep learning hierarchical coding features.

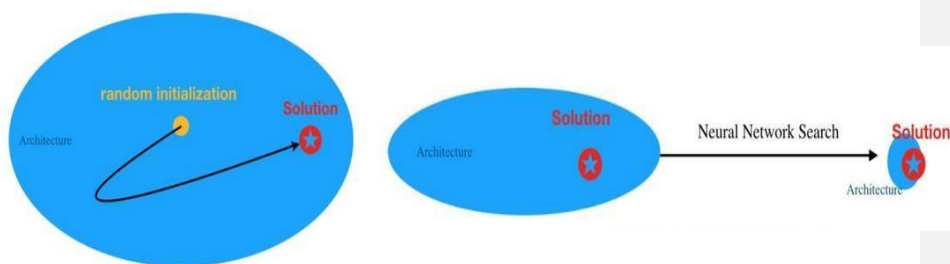


Figure 6 General Neural Networks & Weight Agnostic Neural Networks

As above is our intuitive understanding of the two networks. A general neural network randomly initializes the weights within the architecture and learns the weights to find the optimal solution, so that the model can accomplish the specific task. Generally, as long as the architecture is "big enough" then it is likely to contain the optimal solution and the gradient descent will roughly find it.

But for neural networks that do not use learning weights, it is equivalent to constantly specializing the architecture, or reducing the variance of the model. This way, as the architecture gets smaller and contains only the optimal solution, the randomized weights will solve the practical problem. Of course, searching from small to large architectures, as the researchers did, is also possible, as long as the architecture happens to enclose the optimal solution.

Core idea - In order to search for neural network architectures with strong inductive bias, the researchers propose to search for architectures by reducing the importance of weights. The specific steps are 1) to provide a single shared weighting parameter for each network connection and 2) to evaluate the network over a larger range of weighting parameter values.

Instead of optimizing fixed network weights, this study optimizes architectures that perform well across a wide range of weight values. The researchers demonstrated that the method can generate networks that perform different continuous control tasks using random weighting parameters.

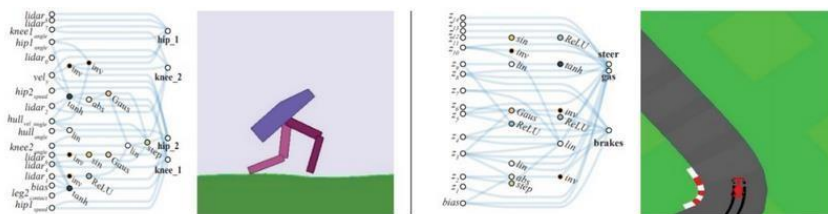


Figure 7 Examples of WANN

As above: examples of weight-independent neural networks: bipedal walking intelligent body (left), racing car (right). Researchers search architecture by reducing the importance of weights. The network uses a single shared weight value. All architectures optimized for performance on a wide range of weighted values are still able to perform different tasks without weighting training.

The goal of **NAS** technology is to generate architectures that go beyond human manual design when training is complete. No one has ever claimed that the solution is inherent to that network architecture.

Commented [GD8]: NAS ? expand the abbreviation

In order to generate an architecture that can be coded and decoded by itself, the importance of weights must be minimized. In evaluating network performance, the researcher did not choose the network that used the optimal weight value, but instead extracted the weight value from a random distribution. Replacing weighted training with weighted sampling ensures that performance is only related to the network topology.

However, due to the high dimensionality, reliable sampling of the weight space is not feasible on all but the simplest of networks. Although the dimensionality problem prevented the researcher from sampling the high dimensional weight space efficiently, the number of weight values was reduced to 1 by implementing weight sharing on ownership weights.

Systematically sampling individual weights is simple and efficient, allowing us to approximate network performance with just a few experiments. This approximation can then be used to search for a better architecture.

7. References

- [1] Mitchell, Tom M. Machine learning[M]. McGraw-Hill, 2003.
- [2] Bishop, Christopher. (2006). Pattern Recognition and Machine Learning. 10.1117/1.2819119.
- [3] Glickman M , Balthrop J , Forrest S . A Machine Learning Evaluation of an Artificial Immune System[J]. Evolutionary computation, 2005, 13(2):p.179-212.
- [4] Hinton G E , Osindero S , Teh Y W . A Fast Learning Algorithm for Deep Belief Nets[J]. Neural Computation, 2006, 18(7):1527-1554.
- [5] C. Saunders, M. O. Stitson, J. Weston. Support vector machine[J]. computer science, 2002, 1(4):1-28.
- [6] Hagan, Martin T, Demuth, Howard B, Beale, Mark H. Neural network design[M]. China Machine Press, 2002.
- [7] Gaier A , Ha D . Weight Agnostic Neural Networks[J]. 2019.

MARKING CRITERIA AND STUDENT FEEDBACK

This section details the assessment criteria. The *extent* to which these are demonstrated by you determines your mark. The marks available for each criterion are shown. Lecturers will use the space provided to comment on the achievement of the task(s), including those areas in which you have performed well and areas that would benefit from development/improvement.

Common Assessment Criteria Applied (Assignment 1)	Marks available	Marks awarded
1. Engagement with Literature Skills Extent of research and/or own reading, selection of credible sources, application of appropriate referencing conventions.		
Your references are of a high quality. Very good supporting choices. But try to include more. You need to apply the school referencing system correctly.	20	14
2. Knowledge and Understanding Skills Extent of knowledge and understanding of concepts and underlying principles associated with the discipline.		
You have explained and discuss the selected topic in a meaningful way. You spent a lot of time on discussing established ideas before moving on to new and emerging ideas. Try to be more disciplined here.	20	16
3. Cognitive and Intellectual Skills Analysis, evaluation and synthesis; logic, argument and judgement; analytical reflection; organisation of ideas and evidence		
A very detailed introduction with plenty of definitions but you covered a lot of material only superficially. I would have preferred it if you focused on one emerging area in this – which you did, eventually. A shame as you have identified and studied relevant and emerging ideas in NNets.	50	34
4. Practical Application Skills Deployment of methods, materials, tools and techniques; application of concepts; formulation of innovative and creative solutions to solve problems.		
5. Transferable Skills for Life and Professional Practice Attributes in professional practice: individual and collaborative working; deployment of appropriate media; presentation and organisation.		
Try to write less about the underlying topics as you leave too little space for the interesting emerging ideas. Some minor grammar/ style issues. Otherwise structured but no real conclusions.	10	5

Assignment Mark (Assessment marks are subject to ratification at the Exam Board. These comments and marks are to give feedback on module work and are for guidance only until they are confirmed.)

Late Submission Penalties (tick if appropriate)
Up to four weeks late
First Attempt?

69%