



Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

Machine Learning Emerging & Trends

Name : Haoyi Sun

Student Number : 1904101

Module: Emerging and future trends

20/05/2020

Contents

1. Background	4
2. Machine Learning	5
2.1 The scope of machine learning.....	5
3. Machine Learning Approaches	7
3.1 Regression algorithm.....	7
3.2 Neural networks.....	8
3.3 SVM (Support Vector Machine).....	9
3.4 Clustering algorithm.....	10
3.5 Downside algorithm.....	10
3.6 Recommendation algorithm.....	11
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 General Neural Networks & Weight Agnostic Neural Networks

Figure 6 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.

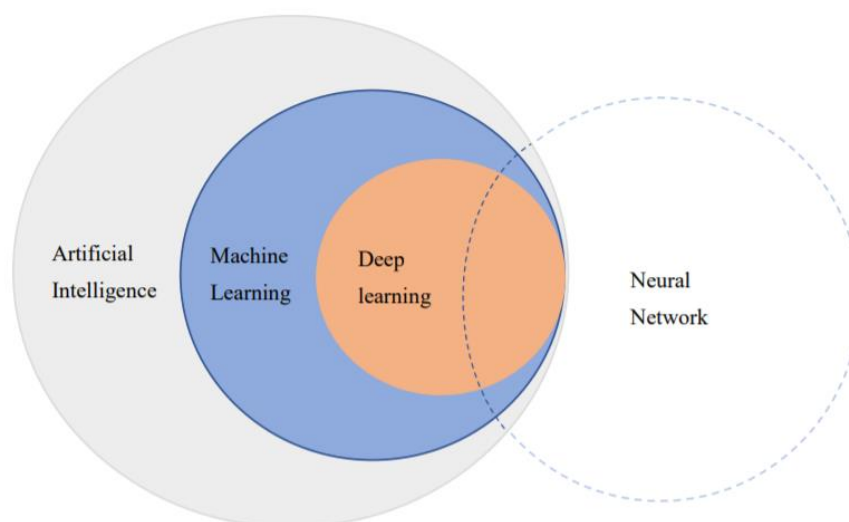


Figure 1 Venn diagram

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.

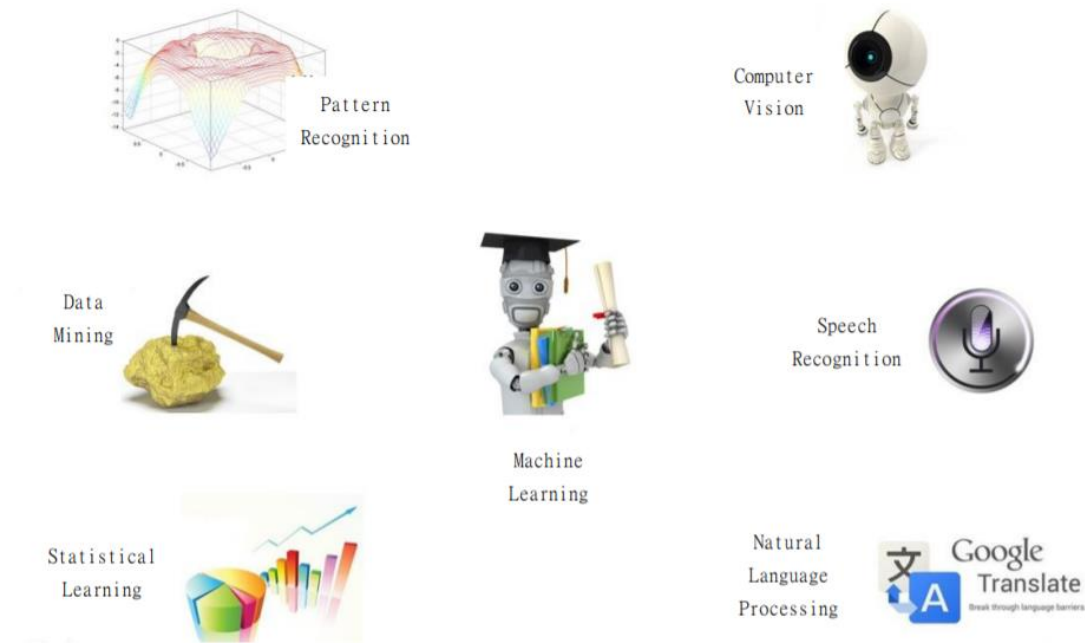


Figure 2 The scope of machine learning

① Pattern Recognition

Pattern recognition = machine learning. The main difference between the two is that the former is a concept developed from industry, while the latter is largely derived from the discipline of computing. In the famous book *Pattern Recognition And Machine Learning*, Christopher M. Bishop begins by saying that "Pattern recognition comes from industry, and machine learning comes from the discipline of computers. However, the activities in them can be seen as two aspects of the same field, while they have both come a long way in the last 10 years."

② Data Mining

Data mining = machine learning + database. The concept of data mining is all too familiar these days. Almost tantamount to hype. But anyone who talks about data

mining will brag about how data mining is, such as digging gold out of data, and turning discarded data into value, and so on. But as much as I may dig for gold, I may dig for "stone". The implication of this statement is that data mining is just a way of thinking that tells us that we should try to mine knowledge from data, but not every piece of data is going to mine gold, so don't myth it. A system will never become omnipotent just because it's on a data mining module (which IBM loves to brag about), quite the contrary, a person with a data mining mindset is the key, and he must also have a deep understanding of the data so that it's possible to derive patterns from it to guide business improvement. Most of the algorithms in data mining are machine learning algorithms optimized in the database.

③ Statistical Learning

Statistical learning approximation equals machine learning. Statistical learning is a discipline that highly overlaps with machine learning. Because most of the methods in machine learning come from statistics, it can even be argued that the development of statistics has contributed to the prosperity of machine learning. The famous support vector machine algorithm, for example, is derived from the discipline of statistics. However, there is a difference between the two to some extent. The difference is that a statistical learner focuses on the development and optimization of statistical models, which is more mathematical, while a machine learner focuses on the ability to solve problems, which is more practical, so a machine learning researcher will focus on the efficiency and accuracy of learning algorithms implemented on the computer.

④ Computer Vision

Computer vision = image processing + machine learning. Image processing techniques are used to process images into inputs that fit into a machine learning model, and machine learning is responsible for identifying the relevant patterns from the images. There are many computer vision related applications, such as Baidu mapping, handwritten character recognition, license plate recognition and other applications. This is an area with a very hot application outlook and also a hot research direction. With the development of the new field of deep learning in machine learning, which has greatly facilitated the effectiveness of computer image recognition, the future prospects of the computer vision community are inestimable.

⑤Speech Recognition

Speech recognition = speech processing + machine learning. Speech recognition is a combination of audio processing technology and machine learning. Speech recognition technologies are not generally used on their own, but are generally combined with technologies related to natural language processing. Current related apps are Apple's voice assistant SIRI, among others.

⑥Natural Language Processing

Natural language processing = text processing + machine learning. Natural language processing technology is primarily a field that allows machines to understand human language. In natural language processing techniques, there is a great deal of use of techniques related to the principles of compilation, such as lexical analysis, syntactic analysis, etc. In addition to this level of understanding, semantic understanding, machine learning and other techniques are used. As the only symbol created by humans themselves, natural language processing has been a constant research direction in the machine learning community. According to Baidu machine learning expert Yu Kai, "listening and seeing, to put it bluntly, both cats and dogs know, and only language is unique to humans. How to use machine learning techniques for deeper understanding of natural language has been the focus of attention in industry and academia.

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 tumor 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 final 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".

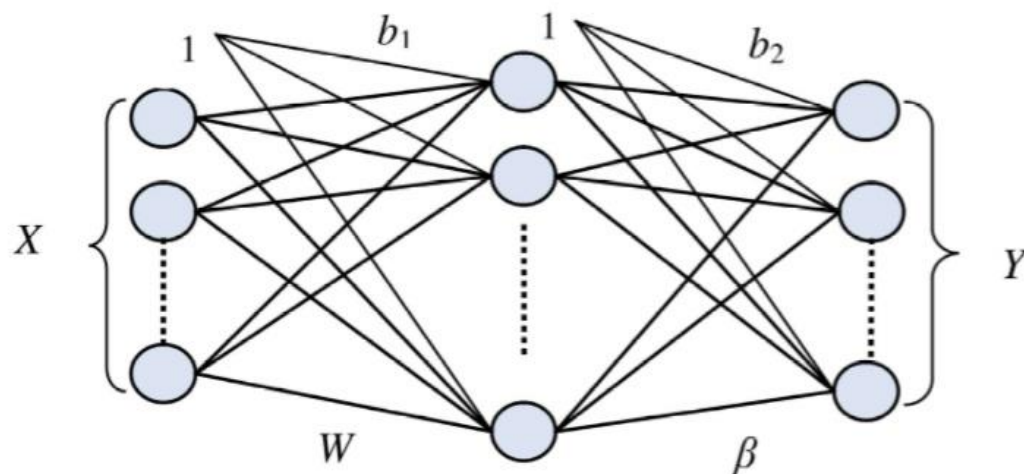


Figure 3 Single-hidden Layer Feedforward Neural Networks (SLFNs) Structure

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.

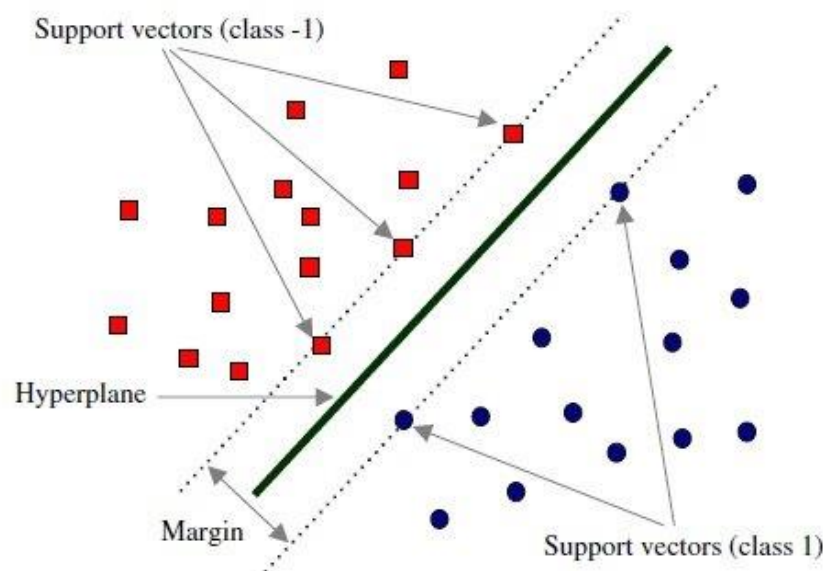


Figure 4 Support Vector Machine Structure

However, by combining with Gaussian "nuclei", the support vector machine can express very complex classification boundaries and thus achieve very good classification results. The "nucleus" is in fact a special function, the most typical feature of which is the ability to map low-dimensional space to high-dimensional space.

How do we delineate a circular classification boundary in a two-dimensional plane? It can be difficult in a two-dimensional plane, but by "nuclei" you can map two-dimensional space to three-dimensional space, and then use a linear plane to achieve a similar effect. That is, the nonlinear classification boundary delineated by the two-

dimensional plane can be equated to the linear classification boundary of the three-dimensional plane. Thus, we can achieve the effect of non-linear division in the two-dimensional plane by simple linear division in the three-dimensional space.

A support vector machine is a machine learning algorithm with a strong mathematical component (neural networks, by contrast, have a biological science component). In the core step of the algorithm, one step proves that mapping data from low to high dimensions does not lead to a final increase in computational complexity. Thus, by supporting the vector machine algorithm, it is possible to maintain computational efficiency and at the same time obtain very good classification results. Therefore, the support vector machine has occupied the most central position in machine learning throughout the late 1990s, essentially replacing neural network algorithms. It is not until now that neural networks have resurfaced with deep learning that the delicate balance between the two has shifted again.

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.

The main role of the degradation algorithm is to compress data and improve the efficiency of other machine learning algorithms. Data with thousands of features can be compressed down to several features through a reduced dimensional algorithm. Another benefit of the downscaling algorithm is the visualization of the data, such as

compressing 5 dimensions to 2 dimensions, which can then be visualized in a 2-dimensional plane. The main representative of the reduced dimensional algorithm is the PCA algorithm (i.e. principal component analysis algorithm).

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 learning is the most common type of machine learning in which the training data is labelled and the training goal is to be able to give the new data (test data) the correct label. For example, to classify emails as spam or not, we start by training some emails and their labels (spam or non-spam) together, learning that the model is constantly capturing the connections between these emails and labels for self-adjustment and refinement, and then we give some new emails without labels and let the model classify the new emails as spam or not.

b) Unsupervised learning is often used for data mining, for finding something in a large amount of unlabelled data. There are three main types of unsupervised: clustering, discrete point detection and downscaling.

Its training data is unlabelled and the training goal is to be able to classify or differentiate the observed values, etc. Unsupervised learning, for example, should be able to distinguish a picture of a "cat" from a large number of different pictures based solely on the characteristics of all "cat" pictures without giving any additional cues.

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.

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.

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).

Ultimately, what is the "depth" of deep learning is the number of layers experienced from the "input layer" to the "output layer", i.e. the number of "hidden layers", the more layers, the deeper the network structure. So the more complex the choice issue, the more layers are needed. Of course, in addition to the number of layers, the number of "neurons" per layer is also greater. For example, AlphaGo's strategy network is 13 layers. The number of neurons in each layer is 192.

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?"

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.

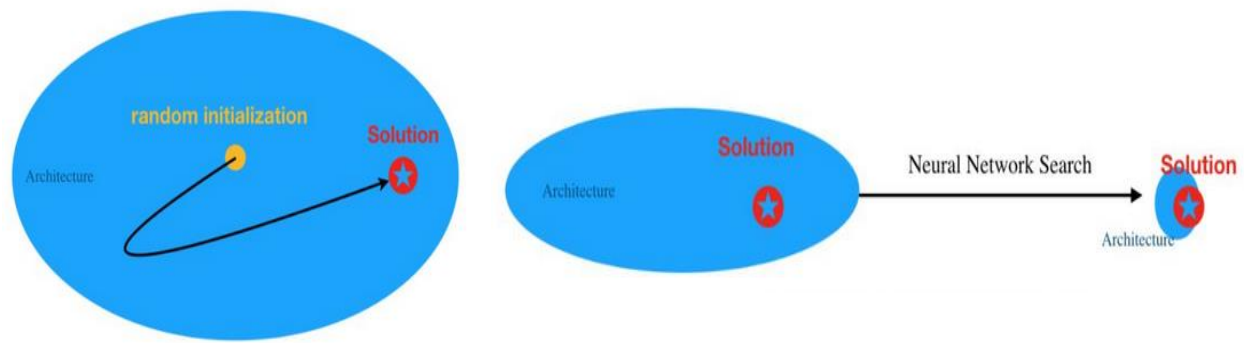


Figure 5 General Neural Networks & Weight Agnostic Neural Networks

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.

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.