Capitalized Comparison of Three Machine Learning Models: Linear Model, Decision Tree, Neural Network

Xiantong Zhao *

Ira A. Fulton Schools of Engineering at Arizona State University Tempe, USA

* Corresponding Author Email: Corresponding author: xzhao141@asu.edu

Abstract. As a matter of fact, with the rapid development of computation ability, machine learning has boosted rapidly with much faster training speed in recent years. In reality, machine learning is transforming industries with its ability to derive insights and patterns within massive datasets. Among numerous algorithms available, certain foundational models stand out due to their efficiency and capability. With this in mind, this study compares three typical models among various machine learning scenarios, i.e., linear models, decision trees, and neural network models. According to the analysis, the basic principle, concepts as well as parameters will be demonstrated. While all three has its unique pros and cons, this study aims to guide readers in choosing most fitting model with their tasks, by clarify the difference and future outlooks of these models. At the same time, the future development trends for the machine learning models will be proposed based on the analysis. Overall, these results shed light on guiding further exploration of machine learning.

Keywords: Linear Model, Decision Tree, Neural Network, machine learning.

1. Introduction

Within the last decade, machine learning has gone through revolutionary advance and became the hottest topic in the technology industry, come up with the widely used applications like ChatGPT and autonomous driving. Back to year 1950, Alan Turing posted the question “can machines think?” [1]. He also introduced a test to access machine intelligence, which is now famously referred to as the Turing Test [1], which is still be used to test if a machine might be called intelligent. Six years later, scientist John McCarthy introduced the term “artificial intelligence” at Dartmouth Workshop [1], but this concept was not known by many non-scientists group until the year 1997, a computer chess-playing system defeated the current world chess champion [1], this event represents the potential for artificial intelligence to reach and surpass human intelligence. However, the development of machine learning or AI is not always smooth, since it is a data-driven technology, “most fields have only limited data or poor-quality data” [2]. But in industries with adequate data, machine learning has the potential to process and enable the transformation of big data into actionable intelligence [3], these “actionable intelligence” could be very productive.

Without noticeable feeling, a lot of applications of machine learning are already involved in life. For example, utilize machine-learning techniques to offer customers personalized services such as product recommendation and sales assistance [2]. For daily use phone-based applications like amazon and Instagram, they use these techniques to send the personalized products advertisement or posts. Machine learning also changed the way mankind lived, one of these changes is the autonomous car. Car manufacturers like Tesla, BMW, Ford etc. are in the process to build autonomous cars that be able to handle tricky situation with machine learning technology [4]. With autonomous cars, people could save a lot of energy from driving and eventually could change the way of transportation. Another significant application is ChatGPT, which is a chatbot based on the GPT model developed by OpenAI. One could use ChatGPT as a search engine that give direct answers instead of the way other search engine do: referring to sources where one must find the answer [5]. As a few updated versions, the new model GPT4 is more powerful and could improve efficiency of some working processes like reading documents or brainstorm. Beside those, machine learning also plays a significant role in sounds recognition, this technology is used in voice assistant in almost every mobile device.

Since OpenAI released ChatGPT in December 2022, the popularity of ChatGPT makes people realize the how machine learning could be a revolutionary thing, the attention to machine learning raised to an unprecedented height, a lot of technology companies working on develop their own machine learning model, as for uncountable applications that utilize the machine learning. But the main mechanisms behind the machine learning is accountable, this arose the interests to study different mechanisms or technologies that supports machine learning models, since with this knowledge, one could fundamentally learn the difference and pros and cons of each model. From simple to complex, the technologies behind machine learning include linear model, decision tree and neural network, for those three technologies, there will be individual section to talking about the details and comparisons. Juan and Russell have their explanation: Machine learning (ML) techniques consist of algorithms that attempt to extract patterns from data and to establish associations between these patterns and distinct classes of data samples [6]. Machine learning includes wide-ranging models and algorithms, each designed to address certain problems or process some specific data formats. Machine learning can be categorized into three primary categories: supervised learning, unsupervised learning and reinforcement learning, each of those categories include many Machine Learning models.

In supervised learning, the algorithm is trained on a labeled dataset, meaning that there will be a correct output for each input, the name “supervised” explains itself. Supervised learning generally deals with two tasks: classification and regression, the response of regression is a continuous variable, the response of classification is a categorical variable [7]. Linear regression, decision trees and neural networks are some common models used in supervised learning. Unsupervised learning, on the other hand, working with datasets that do not have a clearly defined dependent variable [8]. Instead of testing hypothesis by statistical methods or to create prediction or classification models based on a set of conditions and a specified response, unsupervised learning focuses on exploring the structure of data and forming hypotheses [8]. One common task in unsupervised learning is clustering (grouping similar data points together). K-Means lustering and autoencoders are two examples of unsupervised learning algorithms. Semi-supervised learning is a middle-ground approach where the algorithm is trained on a mix of both labeled and unlabeled data, the portion of labeled data and unlabeled data is not fixed. The goal of semi-supervised learning is to use the larger amount of unlabeled data to refine the learning process. Reinforcement learning (RL) is an approach that enhances a strategy based on given objective by interacting with an environment, with an agent observes the environment’s state [9]. RL is decision making framework and can be applied in various scenarios whenever an artificial agent needs to decide the choices of actions, its goal is simply to choose actions to select actions that maximize future rewards [10]. Reinforcement learning includes algorithms like Q-Learning and Deep Q-Network. A notable example of reinforcement learning is AlphaGo, which initially used supervised learning based on human playing and then reinforcement learning with self-play [10]. AlphaGo is one of the examples of robotics that been training by reinforcement learning to perform complex tasks.

2. linear model

Linear models are a common type of supervised learning algorithm. The goal for linear models is to learn mapping from input features to output labels, this goal corresponds to the concept of supervised learning. Linear regression, logistic regression and support vector machine (SVM) are three commonly used models. Linear regression is a statistical model which establishes a relationship between two variables (One serves as Dependent Variable and the other as Independent Variable) [11], it has the formula like $Y = aX + b$, where Y is dependent variable and X is independent variable. This model is original widely used in mathematics and statistical field, linear regression is related to machine learning because its nature as predictive modeling technique used to establish the relationship between two variables [11], since supervised learning algorithm trained on a labeled dataset, where there is input features (independent variables) and corresponding target values (dependent variable). The linear regression model fulfills this perfectly, and due to the simplicity of learning regression, it performs efficiently. Linear regression models could be employed to predict real-valued outputs, such as the tasks like trend forecasting, weather prediction etc. [11, 12].

Same as linear regression, logistic regression is also a statistical method that is closely related to machine learning, i.e., a supervised learning model [13]. The distinction between linear regression and logistic regression could been observed in their respective formula, logistic regression usually has the form based on the sigmoid function:

$$y = \frac{e^{(b0 + b1X)}}{1 + e^{(b0+b1X)}} \tag{1}$$

Logistic regression serves as fundamental supervised machine learning algorithm for classification in natural language processing. Logistic regression also shares a strong connection with neural networks. Logistic Regression could be classified into three types: binary, multinomial and ordinal. Medical diagnosis, credit scoring, marketing and customer analytics and various other applications of logistic regression are examples of problems where the primary goal is to classify data into distinct classes or categories based on relevant features and criteria.

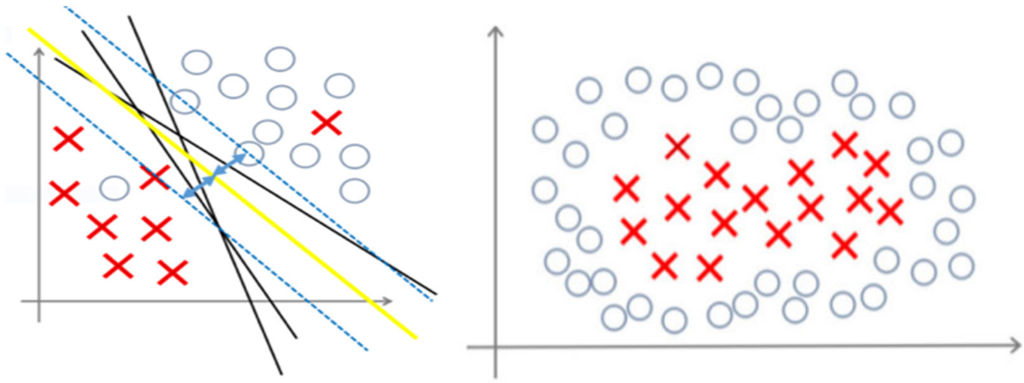

Figure 1. A sketch of SMV and separating results [15].

As supervised machine learning model, Support Vector Machines (SVMs) are powerful tools that can be used for both classification and regression tasks. SVMs are used to find a hyperplane that best separates data into distinct classes while maximizing the space between these classes [14]. Hard margin, soft margin and kernel functions are three different variations of SVM. In hard margin, hyperplanes that could separate every training sample. However, most data cannot be separated linearly, this would be the case of soft margin and kernel function. Soft margin allows SVM to accommodate certain incorrect samples, as shown in Fig. 1. In the case of kernel function, in order to separate these samples correctly, one needs to map the sample to a higher dimensional space so one can identify suitable hyperplanes [15].

3. Decision Tree

Decision tree is a machine learning model that mimics a decision-making process, it's used for both classification and regression tasks. Decision tree has the structure like a tree, shown as Fig. 2, where each internal node represents a decision, each branch represents an outcome or choice, and each leaf node represents a final decision or prediction. There are some reasons that decision trees could be a foundation of machine learning. Two of these reasons include their predictive accuracy (when deal with tabular data [16]) and efficiency (can be trained with relatively compact datasets [16]).

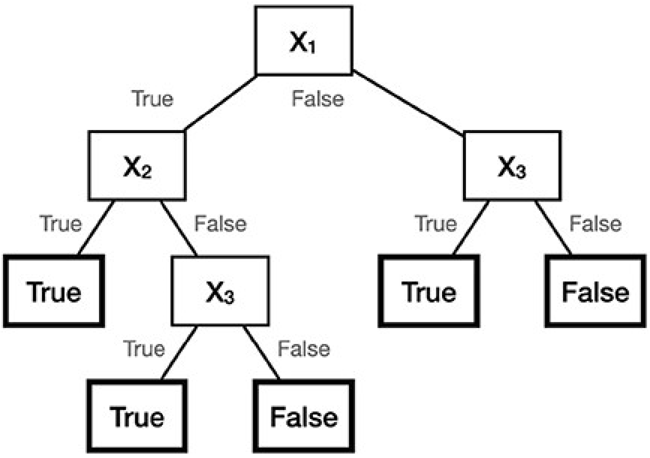

Figure 2. A sketch of decision tree [15].

Random Forest, LightGBM and XGBoost are powerful ensemble methods that utilize decision trees as their base model. Formally, a random forest is a predictive model consisting of a group of randomly generated individual regression regression trees [17]. For each tree, random forest selects a subset of features randomly at each split point, this feature randomization further adds diversity and help mitigate the risk of overfitting. Another advantage of random forests is high predictive accuracy. Both LightGBM and XGBoost are gradient boosting algorithms, and they belong to the category of Gradient Boosted Decision Trees (GBDT) [18], since there are a lot of comparisons between them. LightGBM includes two innovative techniques: Gradient-based One-Side Sampling and Exclusive Feature Bundling to handle a large amount of data samples [18]. XGBoost contains techniques such as weighted quantile sketch and sparsity-aware split finding [19]. Like most other boosting algorithms split the tress depth wise or level wise, XGBoost has level-wise algorithm when tree growing, but leaf wise algorithm is used in LightGBM (seen from Fig. 3), when expanding on the same leaf in LightGBM, the leaf-wise algorithm has the capability to minimize loss to a greater extent compared to level-wise algorithm. This leads to significantly improved accuracy which is rarely achievable with other existing boosting algorithm [20], in terms of computational speed and memory consumption, LightGBM can outperform XGBoost as well [18] but that does not mean XGBoost is an inferior algorithm in every term compared to LightGBM, when deal with smaller datasets or when interpretability is crucial [20], XGBoost may perform better than LightGBM.

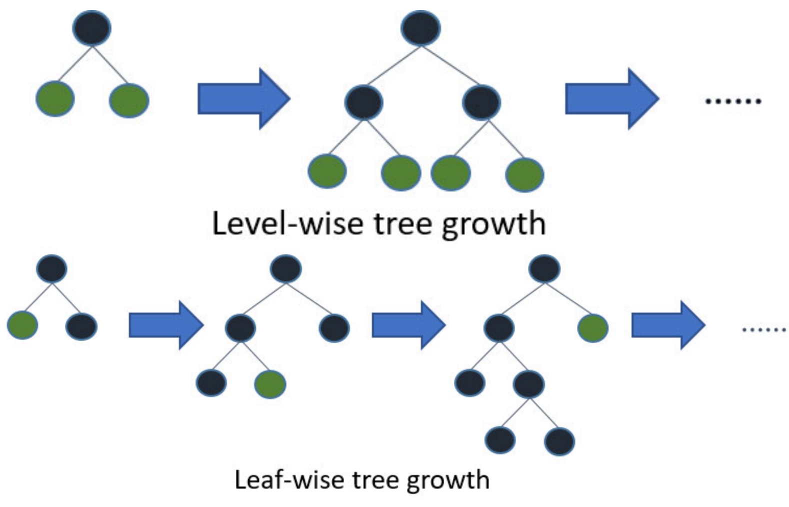

Figure 3. A sketch of LightGBM [20].

4. Neural Network

Neural Network (NN) is a fundamental concept of machine learning. It is a computational model inspired by the operation and constructure of human brain [21], designed to solve a wide range of tasks, from understanding image and speech to processing natural language. The fundamental layout of an Artificial Neural Network (ANN) can be depicted as shown in Fig. 4. One significant drawback of conventional ANN approaches is their difficulty in handling the computational demands with processing image data [21]. CNN, RNN, GAN and transformers are some of the important Neural Network models.

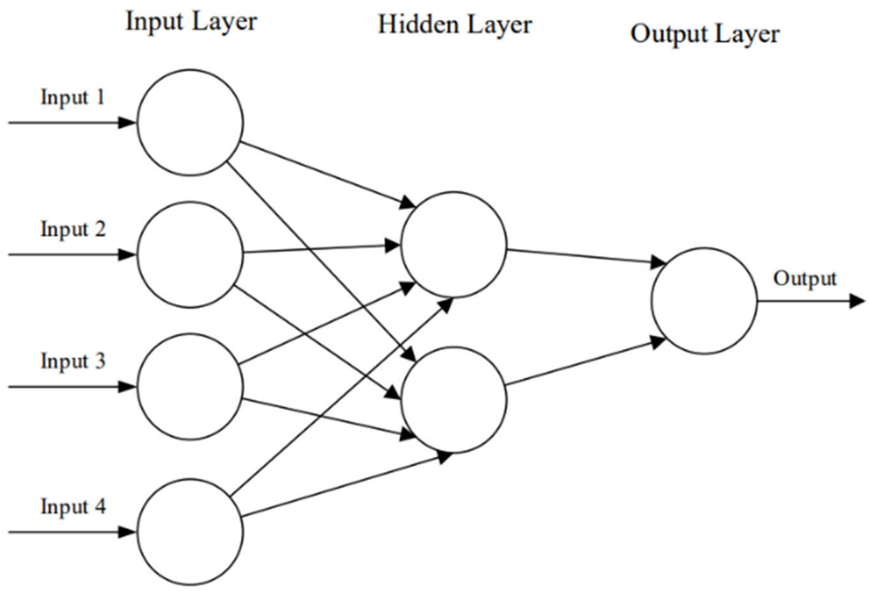

Figure 4. A sketch of neural network [21].

CNNs are similar to traditional ANNs. In both, each neuron receives inputs and perform an operation, from the raw image vectors as inputs to the final class score output [21]. CNNs are powerful machine learning algorithms, however they can be demanding significant computational resources. What sets Convolutional Neural Networks apart from other types of ANN is their ability to utilize what they know about the inputs they are given, rather than trying to handle everything [21]. As a result, a simpler network design becomes possible. There is a problem that traditional neural networks have: not be able to retain past or historical information. RNNs have a solution for that bring the idea of memory into neural networks by including the dependency between data points [22]. Based on this feature, RNNs can be trained to remember concepts based on context. But this feature also results in a short-term memory and makes RNNs can handle sequential data. What RNNs do better than other NNs is it be able to handle inputs of varying length. RNNs suffer from the vanishing gradient descent, which could prevent the network from learning new weights, this results in the RNN forgetting what is seen in longer sequences [22].

GAN, short for Generative Adversarial Network, made its debut in a 2014 paper titled "Generative Adversarial Nets" authored by Ian Goodfellow and his colleagues. GANs are designed to generate new, synthetic data that is similar to a given dataset. GAN is defined as a innovative framework for estimating generative models using an adversarial process, involving the training of two models: a generative model and a discriminative model [23]. It is like a creative contest between two computer programs. One program tries to make fake data that looks real, while other program tries to tell if something is real or fake, they both learn and get better at their tasks by competing against each other. Examples of what GANs can be used for including enhancing image quality, generating art and translating images between different styles. Transformer is a newer concept compared to the models mentioned above. The transformer's key innovation is its attention mechanism, which enables the modeling of dependencies within input or output sets, regardless of the distance between these sets [24]. There are many advantages of transformers: highly efficient (parallelism [24]), capturing long-range dependencies (attention mechanism [24]), state-of -the-art performance (multi-head attention [24]) etc. Some of disadvantages of transformer include lack of interpretability, data requirements and attention overheads.

Linear model, decision tree and neural network, these three models all have their advantages and disadvantages, one could apply these models accordingly. Linear models emphasize establishing relationships between variables, they shine in their simplicity and direct interpretability but may fall short when data isn't linearly separable or has complex patterns. Linear models offer reliable methods for predictions and classification. Decision trees with the hierarchical structure are easily to understand and can dissect data into decisions at each node, which offers insights into the underlying structure of the data. However, decision trees sometimes could be overfitting, which ensemble methods like Random Forest and LightGBM help to counteract, these ensemble methods also have high predictive accuracy. Small changes in the data could also lead to different tree structures, makes decision trees unstable. Decision trees also tend to favor dominant classes in imbalanced datasets which could cause bias. On the other hand, neural networks are the most complex of the three models. They can approximate any function with models like CNNs, RNNs and transformers. While these models offer a broad spectrum of capabilities, these models also require a significant number of resources (data, memory, computational power etc.) and sometimes hard to interpret due to their complexity.

The future is promising for these three machine learning models. As professionals continue to innovate in the field. Linear models are simple and could handle a more complex data, due to its regularization techniques. Decision trees, usually have problems of overfitting and instability, may see improvements through enhanced pruning strategies and ensemble techniques, making them more robust and reliable. Neural networks, while powerful, will likely become more interpretable through ongoing research in explainable AI, and their training will become more efficient with the aid of hardware advancements and transfer learning methods. Additionally, hybrid models [25] and collaborative approaches that combine the strengths of these three models are anticipated to lead the way in addressing complex real-world problems, providing solutions that are both accurate and interpretable, while also optimizing resource utilization.

5. Conclusion

To sum up, this study explores three foundational machine learning models: linear models, decision trees, and neural networks, clears up their basic information, strengths, and applications. Linear models are good at understanding relationships between variables due to their simplicity and interpretability. Decision trees, which break down data hierarchically, offer insights into data structure but can be prone to overfitting, however, ensemble methods like Random Forest and LightGBM can counteract the overfitting. Neural networks, being the most complex, can approximate any function, with models like CNNs, RNNs, and transformers standing out, but require substantial resources and lack straightforward interpretation. With ongoing advancements, each model type will see enhancements in their capabilities and effectiveness, promising exciting developments in the field of machine learning. This review makes it easier to understand the basics of machine learning models. By highlight the strengths and limitation of each model, also with detailing each model's principles, functions, and application example. Through the comparison, the article aims to guide readers in choosing the appropriate model based on specific tasks and data complexities.

References

[1] Royal Society & British Academy, *Machine Learning: The Power and Promise of Computers That* Learn *by Example* (RS&BC, London, 2017), pp.1 - 331.

[2] M. Yang, Y. Liu, T. Chen, and Y. Tong, ACM Transactions on Intelligent Systems and Technology 10 (2), 1 – 19 (2019).

[3] C. Zhou, S. Pan, J. Wang and A. V. Vasilakos, Neurocomputing (Amsterdam) 237, 350 – 361 (2017).

[4] V. Muthalagu, A. S. Bolimera, D. Duseja, and S. Fernandes, Transport and Telecommunication 22 (4), 383 – 391 (2021).

[5] P. Gordijn and H. Have, Medicine, Health Care, and Philosophy 26(1), 1 – 2 (2023).
[6] J. Jovel and R. Greiner, Frontiers in Medicine 8, 771607 – 771609 (2021).
[7] D. Valkenborg, American Journal of Orthodontics and Dentofacial Orthopedics 164 (1), 146 – 149 (2023).
[8] D. Valkenborg, American Journal of Orthodontics and Dentofacial Orthopedics 163 (6), 877 – 882 (2023).
[9] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction* (MIT press, Cambridge, 2018).
[10] Y. Matsuo, Neural Networks 152, 267 – 275 (2022).
[11] K. S. Priya, International Journal of Research in Applied Science and Engineering Technology 15 (2021).
[12] D. Jurafsky and J. H. Martin, *Speech and Language Processing (Chapter 5)* (Stanford University, Stanford 2023).
[13] J. Hair, *Análise multivariada de dados* (Porto Alegre, Bookman Editora, 2009).
[14] D. Boswell, *Introduction to Support Vector Machines.* (Caltech press, Pasadena, 2022).
[15] L. Jun, Journal of Physics. Conference Series 1748 (5), 52006 (2021).
[16] D. Blockeel, L. Devos, B. Frénay, G. Nanfack and S. Nijssen, Frontiers in Artificial Intelligence 6, 1124553 (2023).
[17] G. Biau, Journal of Machine Learning Research 13, 1063 - 1095 (2012).
[18] G. Ke, Q. Meng, T. Finley et al., NIPS 17, 3149 - 3157 (2017).
[19] T. Chen and C. Guestrin, Advances in Pure Mathematics 6, 9 (2016).
[20] E. Khandelwal. Which Algorithm Takes the Crown: Light GBM vs XGBOOST? Analytics Vidhya. Retrieved from https://www.analyticsvidhya.com/blog/2017/06/which-algorithm-takes-the-crown-light-gbm-vs-xgboost/.
[21] K. O'Shea, and R. Nash, *An Introduction to Convolutional Neural Networks* (Springer, Berlin, 2015).
[22] J. Dancker, *A Brief Introduction to Recurrent Neural Networks*, (Towards Data Science, London, 2022).
[23] A. Creswell, T. White, V. Dumoulin, et al., IEEE signal processing magazine 35 (1), 53 - 65 (2018).
[24] S. Vaswani, N. Shazeer, N. Parmar, et al., Advances in neural information processing systems 30 (2017).
[25] M. Zhou, H. Fan, Y. Liu, H. Zhang and R. Ji, Applied Sciences, 13 (10), 5870 (2023).