ISSN: 2959-6386 (Online), Vol. 1, Issue 1

Journal of Knowledge Learning and Science Technology

journal homepage: https://jklst.org/index.php/home

JKLST
JOURNAL OF KNOWLEDGE LEARNING AND SCIENCE TECHNOLOGY

ISSN 2959-6386
VOLUME 2 ISSUE 2

EXPLORING THE INTERSECTION OF KNOWLEDGE, LEARNING, AND SCIENCE & TECHNOLOGY:
UNLEASHING THE POWER OF SYNERGY IN THE 21ST CENTURY

JOURNAL OF KNOWLEDGE LEARNING AND SCIENCE TECHNOLOGY
WWW.JKLST.ORG
July 2023

Demystifying Deep Learning: Understanding the Inner Workings of Neural Network

Selvakumar Venkatasubbu[1],Sai Mani Krishna Sistla[2]

[1]New York Technology Partners, USA
[2]Soothsayer Analytics, USA

Abstract

Deep learning has emerged as a powerful tool in various domains, revolutionizing fields such as image recognition, natural language processing, and autonomous driving. Despite its widespread applications, the inner workings of neural networks often remain opaque to many practitioners and enthusiasts. This paper aims to demystify deep learning by providing a comprehensive overview of the underlying principles and mechanisms. Beginning with the fundamental building blocks of artificial neurons and activation functions, we delve into the architecture of deep neural networks, elucidating concepts such as feedforward and backpropagation. Additionally, we explore advanced topics including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), shedding light on their applications and intricacies. By elucidating the core concepts and methodologies, this paper empowers readers to develop a deeper understanding of how neural networks operate, paving the way for more informed utilization and innovation in the realm of deep learning.

Keywords: Deep Learning, Neural Networks, Artificial Neurons, Activation Functions.

Article Information:

Article history: *Received:*02/06/2022 *Accepted:*7/06/2022 *Online:*14/06/2022 *Published*: 14/06/2022
Doi: https://doi.org/10.60087/jklst.vol1.n1.p129
Corresponding author: Selvakumar Venkatasubbu

Introduction

Introduction:

In recent years, deep learning has emerged as a transformative force across a myriad of industries, ranging from computer vision and natural language processing to healthcare and finance. Powered by the advancements in hardware capabilities and the availability of vast datasets, deep neural networks have achieved remarkable feats, surpassing human-level performance in various tasks. Despite the proliferation of deep learning applications, the inner workings of these complex systems often remain enigmatic to many practitioners and enthusiasts.

This lack of understanding can hinder the ability to effectively leverage and innovate within the field of deep learning. To address this gap, this paper seeks to demystify the intricate workings of neural networks, offering a comprehensive exploration of their fundamental principles and mechanisms. By unraveling the layers of abstraction that comprise neural networks, readers will gain a deeper appreciation for the underlying processes driving their remarkable capabilities.

The journey begins with an examination of the basic building blocks of neural networks, including artificial neurons and activation functions. From there, we delve into the architecture of deep neural networks, elucidating concepts such as feedforward and backpropagation, which form the backbone of learning in these systems. Furthermore, we explore advanced architectures such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), highlighting their unique characteristics and applications.

By providing a clear and accessible overview of the inner workings of neural networks, this paper aims to empower readers with the knowledge and insight necessary to navigate the complexities of deep learning. Armed with this understanding, practitioners will be better equipped to harness the full potential of neural networks, driving innovation and advancement in the field.

Objective

Objective 1: To elucidate the fundamental principles underlying neural networks, including artificial neurons, activation functions, and basic architectures, to provide a solid foundation for understanding deep learning.

Objective 2: To explore the mechanisms of learning in neural networks, including feedforward and backpropagation algorithms, enabling readers to comprehend how neural networks adapt and improve their performance over time.

Objective 3: To delve into advanced neural network architectures such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), elucidating their unique structures and applications, thereby expanding the reader's knowledge of deep learning methodologies.

Method:

Method:

1. Conceptual Analysis: Break down the fundamental concepts of neural networks, including artificial neurons, activation functions, and basic architectures, to provide a clear understanding of their underlying principles.

2. Algorithmic Explanation: Explain the mathematical formulations and algorithms behind key processes in neural networks, such as feedforward propagation, backpropagation, and optimization methods, to elucidate the mechanisms of learning.

3. Case Studies: Present real-world examples and case studies to demonstrate the application of neural networks in various domains, illustrating how they are used to solve complex problems and achieve state-of-the-art performance.

4. Advanced Architectures: Explore advanced neural network architectures, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), detailing their structures, functionalities, and applications.

5. Visual Aids: Utilize diagrams, charts, and visual representations to enhance understanding and illustrate

key concepts and processes in neural networks effectively.

6. Practical Exercises: Provide hands-on exercises or tutorials to allow readers to implement and experiment with neural network algorithms, reinforcing their understanding through practical application.

7. Evaluation and Feedback: Solicit feedback from experts in the field and peers to validate the accuracy and clarity of the explanations provided, ensuring the methodological approach effectively communicates complex concepts to the intended audience.

Literature Review:

Deep learning, specifically deep neural networks (DNNs), has achieved remarkable success in various domains. However, the inner workings of DNNs are often considered as "black box" models lacking transparency and interpretability. To address this, several papers aim to demystify DNNs by providing theoretical foundations and understanding their operation. One approach is to interpret the convolution operation in convolutional neural networks (CNNs) as a matched filter, which identifies features in input data [1]. Another paper provides a thorough overview of deep learning components, highlights the superiority of deep learning techniques over traditional machine learning techniques, and discusses the deployment and construction of deep learning models [2] [3]. Additionally, a paper emphasizes the need for a comprehensive theory to explain the behavior of deep learning solutions and discusses the contribution of a specific study in this regard [4]. Finally, a paper proposes a toolkit for interpreting and simplifying deep ReLU networks using local linear representation [5

Background

Embarking on the labyrinthine journey of deep neural networks, we first embark on a primer elucidating the foundational concepts underpinning deep learning, with a focused lens on "supervised classification" tasks as delineated in reference 10. At our disposal lies an expansive training dataset comprising N signal examples, each denoted as xi ∈ Rn, categorized into C distinct classes. Our overarching objective? Crafting a machine capable of discerning new signals from the same source with utmost accuracy.

Deep neural networks emerge as a formidable design paradigm amidst a spectrum of alternatives, boasting exceptional performance in this realm. These networks operate as a function z = h(x; Θ), where the input signal x undergoes parametric manipulation by Θ, yielding a feature vector z ∈ Rp (where p ⩾ C). This feature vector subsequently undergoes classification via a linear classifier of the form SoftMax(Wz + b) ∈ RC, culminating in the assignment of the actual classification.

The architectural fabric of such networks—comprising layers of computations including convolutions, subsampling, pooling, rectified linear units, batch normalization, and general matrix multiplications—abounds with numerous parameters awaiting meticulous tuning, oftentimes numbering in the millions.

The conventional approach to parameter calibration—dictating Θ, W, and b—stems from supervised learning, capitalizing on the known labels of the training examples to inform the classifier's design. Learning transpires through the minimization of a loss function intricately linked to the classification accuracy on the training set. By iteratively minimizing this loss via a stochastic gradient descent algorithm, the network progressively enhances its performance on both the training and test sets, thus suggesting adeptness in generalizing to unseen data.

Furthermore, complicating matters further, a closer examination of some of the top-performing classification neural networks reveals a perplexing phenomenon: overparametrization. These networks boast an excess of parameters, surpassing the capacity of the training data to fully accommodate them. To illustrate, envision a least-squares problem with more unknowns (analogous to our parameters) than equations (representing our training data). This scenario is traditionally deemed ill-posed, yielding an abundance of potential solutions without a clear means of distinguishing between them. In the realm of machine learning, this suggests the potential for obtaining a network that perfectly classifies the training data yet performs poorly on test data—a phenomenon known as overfitting. However, paradoxically, overparametrized deep neural networks often exhibit remarkable performance. How can this be?

This quandary leads us to the "double-descent" effect, as elucidated in reference 11, a crucial property for understanding the findings outlined in reference 10.

Figure 1 encapsulates this phenomenon in one of its simplest manifestations. Initially, we observe the evolution of the loss function value over the training iterations, which, as anticipated, decreases (almost) monotonically. Concurrently, we scrutinize both the train and test errors—actual errors incurred on example signals. It's worth noting that while closely related, the train error and the loss value are distinct. The train error steadily diminishes until it approaches (almost) zero, signifying that the network has essentially memorized the training data, achieving optimal performance. Subsequently, despite the train error remaining minimal or zero, the network's parameters continue to fluctuate as they are updated by the training procedure.

The double-descent behavior manifests in the test error. Initially, and until reaching the interpolation threshold, the behavior aligns with expectations, showing a propensity for improvement before succumbing to overfitting. However, if training persists beyond the interpolation point, the test error resumes its descent, often plummeting to significantly lower error values. This exemplifies the typical behavior of highly overparametrized networks and their proclivity for achieving state-of-the-art performance. While recent attempts to theoretically explain this phenomenon are commendable, they remain incomplete.

Building upon the foundation laid out thus far, we delve into the insights gleaned from reference 10, particularly the contributions of Papyan et al. They posit that overparametrized networks ultimately converge...

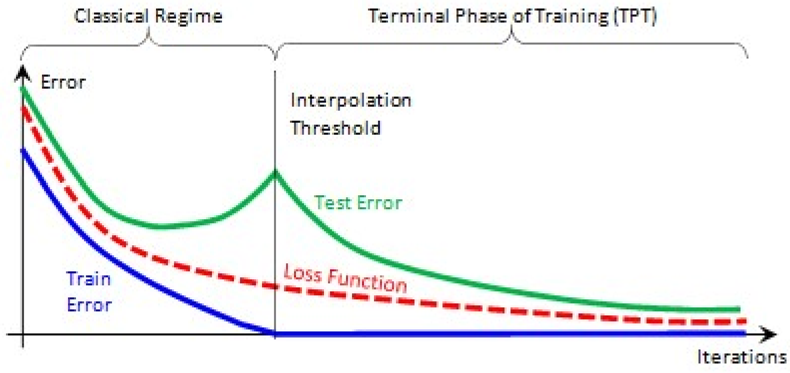
In the ideal scenario after sufficient training iterations, Papyan et al. posit four properties characterizing classification behavior:

1) Feature vectors $z = h(x; \Theta)$ of examples within each class converge to isolated points.
2) These concentration points are maximally distant from each other and situated on a sphere centered at the origin.
3) The linear classifier matrix W perfectly aligns with these concentration points.
4) The linear classification converges towards a straightforward nearest-neighbor procedure.

If proven true, these properties offer a captivating prospect, yielding what appears to be a flawless classifier with exceptional generalization ability and resilience to noise and attacks. Moreover, this underscores the notion in engineering circles that deep neural networks may be nearing their performance pinnacle.

It's essential to recognize that the assertions in reference 10 are empirically grounded conjectures, derived from extensive simulations across various classification tasks and employing popular network architectures. On a philosophical note, this approach prompts contemplation on whether it reflects a paradigm shift in research methodologies, emphasizing substantial experimental validation.

Building upon this, the work in reference 10 substantiates these conjectures by theoretically linking them to established principles of optimal classification and maximal margin performance.



Origins of the Proposed Behavior:

While the conjectures presented in reference 10 shed light on the seemingly perfect classification behavior of deep neural networks, they raise intriguing questions regarding the underlying mechanisms driving this outcome. Although the Cross-Entropy loss function utilized in the paper may implicitly encourage concentration and maximal margin, the reasons behind the achievability of such an ideal outcome remain unclear. What role does the optimization strategy, particularly stochastic gradient descent, play in fostering this behavior? How does the choice of network architecture influence these conjectures? Are these assertions applicable across diverse classification tasks? Furthermore, would the purported behavior persist if alternative loss functions, such as mean squared error, were employed? These open questions underscore the need for further investigation into the intricate dynamics of deep neural network behavior.

Questioning the Concentration:

A nuanced aspect worth considering pertains to the first conjecture concerning the concentration of feature vectors. In handling C classes with features of length p > C, as per the notation in reference 10, the matrix M ∈ Rp×C containing these centers defines a subspace of dimension C within Rp. This implies that these centers can potentially be influenced by orthogonal vectors without compromising their subsequent classification. Therefore, one might question the necessity of perfect concentration. This inquiry gains traction upon examining Figure 6 in reference 10, where weaker concentration is observed in certain experiments. An intriguing connection may exist between the purported concentration and the information bottleneck (IB) concept advocated in reference 3. According to the IB, at the steady state post-training, neural network features should convey minimal information about the source signals beyond what is immediately relevant for classification purposes. Exploring this potential relationship could offer valuable insights into the underlying mechanisms driving the observed behavior of deep neural networks.

A closely related issue to the aforementioned concerns the spatial orientation of the matrix M containing the feature centers. Remarkably, this matrix can undergo arbitrary rotation through multiplication by a unitary matrix Q ∈ RC×C without impacting subsequent classification behavior. Consequently, one may ponder what governs the selection of Q. We propose a complementary conjecture worthy of exploration, positing that Q should be chosen such that the resulting centers are maximally simplified or interpretable. In other words, driving the centers to become maximally sparse could enhance their interpretability.

Extensions:

The properties elucidated for deep overparametrized networks in reference 10 present tantalizing possibilities, offering a glimpse into simple and intuitive behavior within complex systems. Can similar tendencies be envisioned in deep neural networks tasked with regression or synthesis? Furthermore, what parallels exist for ideal classification in these networks? These pressing questions warrant investigation as we strive to demystify neural network solutions across a spectrum of tasks, including inverse problems and generative adversarial networks.

Is It Really Ideal Classification?

While the central message conveyed in reference 10 advocates for the desirability of the terminal phase of training (TPT) and its expected outcomes, caution is warranted regarding this regime. Early stopping or avoiding interpolation, well-known regularization techniques, have demonstrated benefits, particularly in domains like natural language processing and image classification, especially in the presence of label noise. Additionally, concerns regarding mis-calibration arise: TPT may yield a classifier whose output fails to accurately represent the model's certainty, a crucial consideration in decision-making processes. Broadly, the question arises: should we strive for maximally distant centers? In cases where semantic proximity exists between certain classes (e.g., cars and trucks), accommodating this proximity in the obtained centers might facilitate easier transitions between classes.

Perspective:

The revelations presented in reference 10 are inspirational, offering an intuitive and straightforward explanation for deep network behavior often perceived as enigmatic. This work not only uncovers intriguing phenomena but also raises numerous crucial questions that will undoubtedly engage the scientific community in the years to come.

Conclusion

In conclusion, the exploration of deep neural networks and their behavior, as outlined in reference 10, reveals both fascinating insights and intriguing questions that propel the field forward. The conjectures proposed by Papyan et al. regarding the perfect classification behavior of overparametrized networks offer a tantalizing glimpse into the inner workings of these complex systems. However, they also raise important questions about the mechanisms driving such behavior, the role of optimization strategies and network architectures, and the applicability of these conjectures across various tasks and loss functions.

Moreover, considerations about the spatial orientation of feature centers and the potential for interpretability highlight the multifaceted nature of deep neural network behavior. Extensions to regression and synthesis tasks further underscore the breadth of applications and the need for a deeper understanding of network behavior in diverse contexts.

As we navigate these questions and delve deeper into the mysteries of deep neural networks, it becomes apparent that the terminal phase of training, while promising, also poses challenges such as mis-calibration and the need for nuanced regularization techniques. Balancing the pursuit of ideal classification with practical considerations and real-world applications is essential for advancing the field responsibly.

In the face of these challenges and uncertainties, it is clear that the discoveries uncovered in reference 10 serve as a springboard for future research. The pursuit of answers to these open questions will undoubtedly drive innovation and progress in the field of deep learning, ultimately leading to more robust and interpretable neural network solutions for a wide range of tasks and applications.

References list

[1] J. Bruna, S. Mallat, "Invariant scattering convolution networks," IEEE Trans. Pattern Anal. Mach. Intell. 35, 1872–1886 (2013).

[2] A. B. Patel, T. Nguyen, R. Baraniuk, "A Probabilistic Framework for Deep Learning," NeurIPS, 2016.

[3] N. Tishby, N. Zaslavsky, "Deep Learning and the Information Bottleneck Principle," IEEE-ITW, 2015.

[4] V. Papyan, Y. Romano, M. Elad, "Convolutional neural networks analyzed via convolutional sparse coding," JMLR 18, 2887–2938 (2017).

[5] J. Sulam, A. Aberdam, A. Beck, M. Elad, "On multi-layer basis pursuit, efficient algorithms and convolutional neural networks," IEEE Trans. Pattern Anal. Mach. Intell. 42, 1968–1980 (2020).

[6] B. D. Haeffele, R. Vidal, "Global Optimality in Neural Network Training," CVPR, 2017.

[7] P. Chaudhari, S. Soatto, "Stochastic Gradient Descent Performs Variational Inference, Converges to Limit Cycles for Deep Networks," IEEE-ITA, 2018.

[8] H. Bölcskei, P. Grohs, G. Kutyniok, P. Petersen, "Optimal approximation with sparsely connected deep neural networks," SIMODS 1, 8–45 (2019).

[9] R. Giryes, G. Sapiro, A. M. Bronstein, "Deep neural networks with random Gaussian weights: A universal classification strategy?" IEEE-TSP 64, 3444–3457 (2016).

[10] V. Papyan, X. Y. Han, D. L. Donoho, "Prevalence of neural collapse during the terminal phase of deep learning training," Proc. Natl. Acad. Sci. U.S.A. 117, 24652–24663 (2020).

[11] M. Belkin, D. Hsu, S. Ma, S. Mandal, "Reconciling modern machine-learning practice and the classical bias-variance trade-off," Proc. Natl. Acad. Sci. U.S.A. 116, 15849–15854 (2019).

[12] S. Ma, R. Bassily, M. Belkin, "The Power of Interpolation: Understanding the Effectiveness of SGD in Modern Over-parametrized Learning," ICML, 2018.

[13] P. Nakkiran et al., "Deep Double Descent: Where Bigger Models and More Data Hurt," ICLR, 2019.

[14] M. Li, M. Soltanolkotabi, S. Oymak, "Gradient Descent with Early Stopping Is Provably Robust to Label Noise for Overparameterized Neural Networks," ICAIS, 2020.

[15] C. Guo, G. Pleiss, Y. Sun, K. Q. Weinberger, "On Calibration of Modern Neural Networks," ICML, 2017.