Department of Mathematical and Computational Sciences 2024

Optimizing Neural Network Architecture in Artificial Intelligence

Grace Lindgren – Advised by Dr. Colby Long

THE COLLEGE OF WOOSTER

Abstract

The world is becoming extremely technologically advanced as a result of artificial intelligence. Most may know artificial intelligence as the intelligence behind our everyday technology, including cell phones, computers, and more. However, artificial intelligence is much more. One of the most important concepts within artificial intelligence is neural networking. Neural networks send information from an input, through a network of nodes, to an output. This thesis will follow neural networks and artificial intelligence through a description of their history and purpose. Additionally, the mathematics behind neural networks will be covered as well as optimization methods. An example of optimization will be incorporated through a neural network architecture problem resulting in upwards of 80% accuracy – with little complexity. Finally, the future of artificial intelligence will be touched on.

Artificial Intelligence

Background

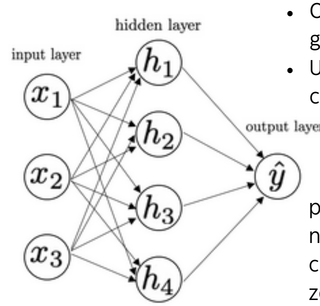
- At first, computer scientists skeptical of how artificial intelligence could work
- 1943 – research began to take off
- 1960s – large number of programs developed
- 1990s – more development; IBM's Deep Blue v. Garry Kasparov chess matches
- 2000s – AI starting to flourish

Neural Networks in Artificial Intelligence: Future Applications

- Chat GPT
  - Neural network that responds to text prompts from users
  - Accuracy learned through training
  - Non-supervised training: conversational responses
- Law
  - Decrease in the amount of lawyers
  - A chatbot can do the work of up to twenty lawyers quicker
- Education
  - Personalized learning for students and strengths/weaknesses grading
- Healthcare
  - More accurate and quicker diagnosis for patients

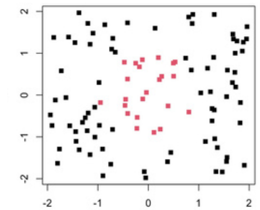
Neural Networks & Optimization

- Input layer, hidden layer(s), output layer
- Modeled off of the human brain
- Simple neural networks – linear decision boundaries
- Complex neural networks – complex decision boundaries
- Can handle thousands of variables
- Activation functions
- "Background" of artificial intelligence



- Optimization using gradient descent method
- Use partial derivatives to compute the gradient of the neural network
  - Minimize the error function (get the prediction from the network – true classification as close to zero as possible)

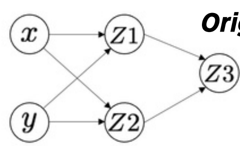
The Unit Circle Example



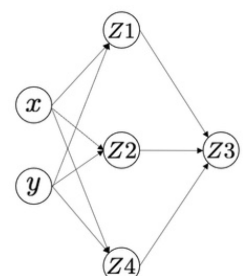
- Goal of example: classify as many points correct as possible using the gradient descent method of optimization
- Neural network learns on a training data set (100 random points; following $x^2 + y^2 \leq 1$ if in red)
- Z3 gives a prediction between zero and one: result is compared to the true classification from $x, y \in [-2, 2]$
- Error is calculated through MSE function and stored in *EE* to keep track of decreasing error
- Use a vector, *v1*, to store current weights
- Essentially, calculating the change in error in accordance with the change in each weight
- Gradient of the error function tells the optimal path of travel to get to the lowest error (travel in the direction of the negative gradient with a rate of 10^{-2})
- Wanted to gain higher classification after trails, so added a neuron to the hidden layer of architecture (Z4)

Architectures

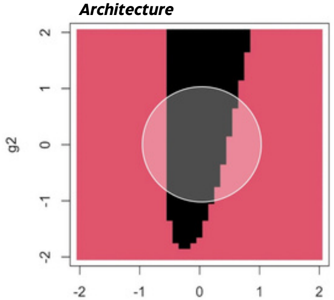
Original



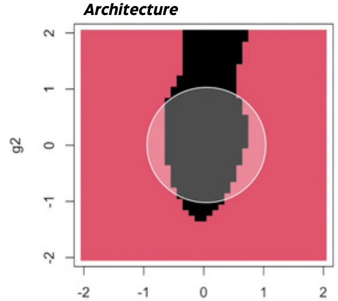
Updated
to Gain Higher Correct Classification



Original Architecture



Updated Architecture



Classification Comparison

Original Architecture

- 92% classification with 5000 iterations on training set

With 10000 iterations:

- 94% classification on training set
- 80.79% classification on testing set; 80.513% average on 10 different sets

Updated Architecture

With 10000 iterations:

- 97% classification on training set
- 87.37% classification on testing set; 87.13% average on 10 different sets

The addition of Z4 adds little complexity, and on average, 6.617% greater classification.

References

[1] A.C. Faul. A Concise Introduction to Machine Learning. Vol. 1. Machine Learning & Pattern Recognition. Taylor & Francis Group, 2020 (pages 13, 19–23, 53–54).
[2] Fernando Iafrate. Artificial Intelligence and Big Data. Vol. 8. Information Systems, Web and Pervasive Computing. ISTE Ltd, 2018 (pages 1, 6–7, 13–14, 18).
[3] Robert Kozma. Artificial Intelligence in the Age of Neural Networks and Brain Computing. Ed. by Francesco Carlo Marabito Cesare Rlippi Yoonsuck Choe. Mara Conner, 2019 (pages 5–6, 15, 17).