

Activation Functions in Neural Networks: A Mathematical Perspective

Nikolai Tristan Pazon
Bachelor of Science in Computer Science
University of San Carlos

October 11, 2024

Abstract

1 Introduction

This paper describes the process of my machine learning model that I created in Java. I chose java because of it is a lower level language compared to python- therefore it requires less processing power. I chose to create a machine learning model to perform what is called a churn analysis.

It is a simple problem that involves binary classification.

1.1 Churn Analysis

A Churn Analysis is a process in business where it refers to understanding why customers stop using their products or services. It also predicts which customers are likely to leave based on the data that a business has collected. In the customer relationship management, or CRM, it is especially important to retain customers because it is often more cost effective than finding new customers and clients. [2]

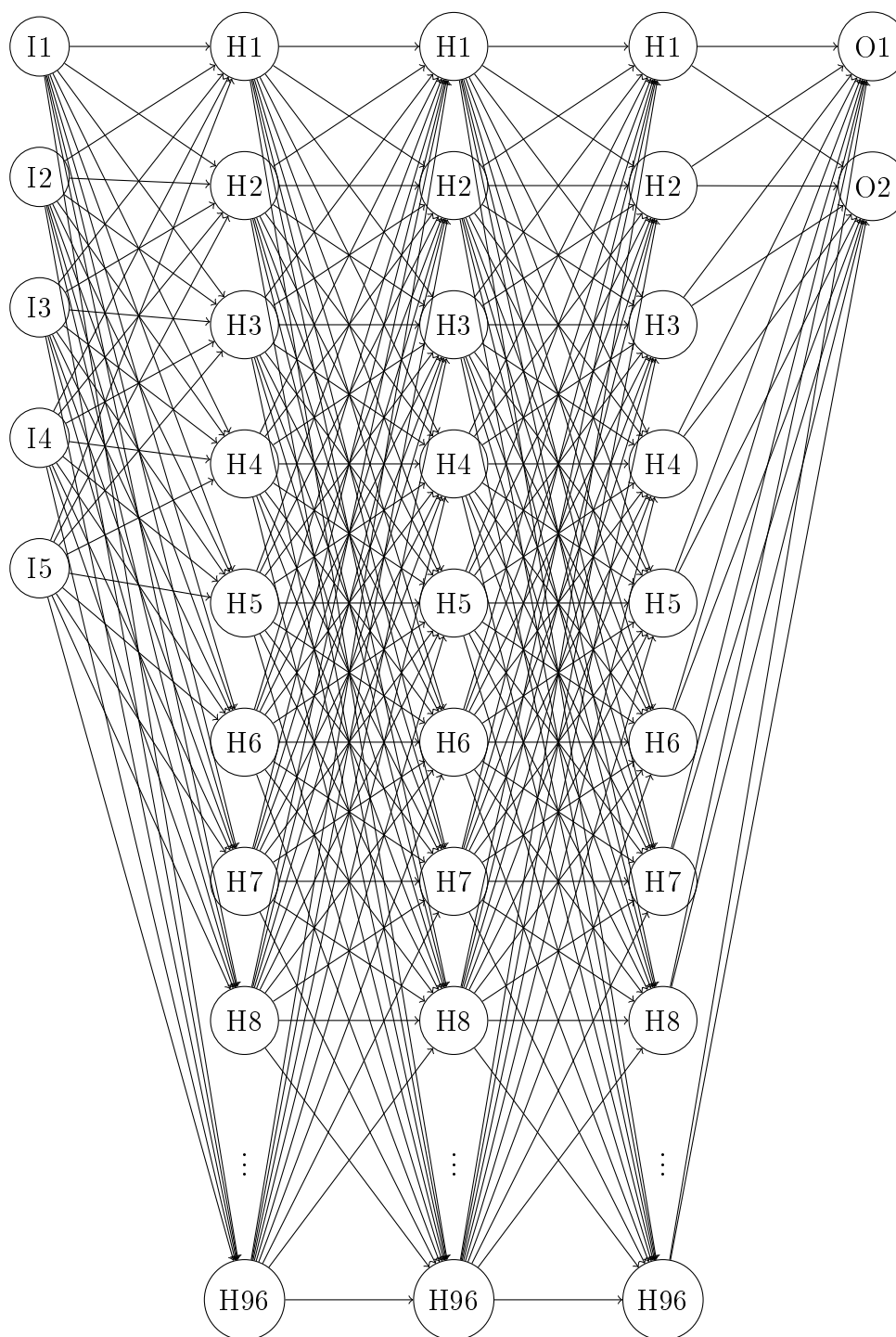
The process typically involves in identifying patters in the behavior of a customer, looking for indicators of dissatisfaction and indicators that a customer will leave. These indicators could be a decline in the engagement between the customer and the business, or it could be poor customer interaction or simply bad product/service quality. [3, 1, 6, 7]

The analysis are incredibly data-driven so methods like machine learning are a perfect use case. Machine learning is a prime tool for this as we need to analyze numerical factors that contribute to a customer's decision to stay or leave. Using this tool we can leverage the analysis to aid a business in designing customer retention strategies. [5, 4]

1.2 Model Architecture

Layer	Type	Number of Nodes	Activation Function
0	Input Layer	10	N/A
1	Hidden Layer	96	ReLU
2	Hidden Layer	96	Sigmoid
3	Hidden Layer	96	Sigmoid
4	Output Layer	2	Softmax

Input Layer Hidden Layer 1 Hidden Layer 2 Hidden Layer 3 Output Layer



- **Input Layer:**
 - **Description:** This is the first layer of the neural network. It inputs the row data or features of the dataset into the model.
 - **Structure:** This consists of n Neurons where n is the number of features / columns in the dataset.
 - **Function:** Data entry without performing computations.
- **First Hidden Layer:**
 - **Activation Function:** Rectified Linear Unit (ReLU)
 - **Definition:** $f(x) = \max(0, x)$
 - **Purpose:** Introduces non-linearity, mitigates the vanishing gradient problem, and allows the model to learn complex patterns.
- **Second and Third Hidden Layers:**
 - **Activation Function:** Sigmoid
 - **Definition:** $f(x) = \frac{1}{1+e^{-x}}$
 - **Purpose:** Maps input values to a range between 0 and 1, useful for learning intermediate representations and refining features.
- **Final Hidden Layer:**
 - **Activation Function:** Softmax
 - **Definition:** $f(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}}$ for each class i
 - **Purpose:** Converts raw output scores into probabilities, providing a probability distribution over possible classes, crucial for predicting customer churn.
- **Output Layer:**
 - **Structure:** Consists of two neurons, each representing one class: churn and non-churn.
 - **Function:** Uses probabilities from the Softmax function to determine the class with the highest likelihood.
 - **Loss Function:** Cross-Entropy
 - * **Definition:** The cross-entropy loss for a single instance is defined as $L = -\sum_i y_i \log(p_i)$, where y_i is the true label (1 for the correct class, 0 otherwise) and p_i is the predicted probability for class i .

- * **Purpose:** Measures the performance of the model by comparing predicted probabilities to actual class labels. It penalizes the model more when the predicted probability for the true class is low, encouraging the model to assign higher probabilities to the correct class.
- * **Advantages:** Cross-entropy is particularly effective for classification tasks as it provides a smooth gradient, which is beneficial for optimization algorithms like gradient descent.

1.3 Training

- **Loss Function:** Cross-Entropy
 - **Purpose:** Suitable for classification tasks, it measures model performance by comparing predicted probabilities to actual class labels, penalizing divergence from actual classes.

2 Activation Functions

3 Mathematical Formulation

4 Comparison of Activation Functions

5 References

References

- [1] Adnan Amin, F. Al-Obeidat, B. Shah, A. Adnan, Jonathan Loo, and S. Anwar. Customer churn prediction in telecommunication industry using data certainty. *Journal of Business Research*, 2019.
- [2] Jonathan Burez and D. V. Poel. Crm at a pay-tv company: Using analytical models to reduce customer attrition by targeted marketing for subscription services. *Expert Systems with Applications*, 2007.
- [3] P. Lalwani, M. Mishra, Jasroop Singh Chadha, and Pratyush Sethi. Customer churn prediction system: a machine learning approach. *Computing*, 2021.

- [4] N. Lu, Hua Lin, Jie Lu, and Guangquan Zhang. A customer churn prediction model in telecom industry using boosting. *IEEE Transactions on Industrial Informatics*, 2014.
- [5] Jitendra Maan and Harsh Maan. Customer churn prediction model using explainable machine learning. *arXiv preprint arXiv:2303.00960*, 2023.
- [6] M. Makhtar, S. Nafis, M. A. Mohamed, M. K. Awang, Mohd Nordin Abdul Rahman, and M. M. Deris. Churn classification model for local telecommunication company based on rough set theory. *Semantic Scholar*, 2018.
- [7] V. Umayaparvathi and K. Iyakutti. A survey on customer churn prediction in telecom industry: Datasets, methods and metrics. *Semantic Scholar*, 2016.