Lab 2

Task B: Supervised Learning

UNIVERSIDAD
DE MÁLAGA

Alejandro Silva Rodríguez

Marta Cuevas Rodríguez

Aprendizaje Computacional
Universidad de Málaga

October 2024

Contents

1 Introduction

This project explores both fundamental and advanced concepts in supervised learning, placing a particular emphasis on the significant potential of Deep Learning in biomedical fields. Supervised learning, a technique where models learn to classify or predict outcomes based on labeled data, is foundational in machine learning, particularly for applications involving complex classification tasks. However, with the advent of Deep Learning, models now have the capacity to process and analyze high-dimensional data with enhanced accuracy and flexibility, making them highly effective in medical and health-related applications.

In this project, we begin by examining essential concepts, such as partitioning datasets into training and testing sets to evaluate model generalizability, and the use of cross-validation to optimize model performance. We then delve into the structure of neural networks, differentiating traditional feed-forward networks from more advanced architectures used in Deep Learning. These modern architectures—characterized by multiple hidden layers, complex connections, and sophisticated activation functions—enable models to learn hierarchical representations of data, capturing intricate patterns and dependencies.

Moreover, this project addresses critical challenges that arise when applying Deep Learning to biomedical contexts. These include the risk of overfitting, where models may perform well on training data but fail to generalize to new data, and the common issue of data scarcity, given the limited availability of labeled biomedical datasets. We will explore techniques to alleviate these challenges, such as regularization, data augmentation, and synthetic data generation, which expand the training dataset and improve model robustness.

Finally, we will review recent advancements in Deep Learning applied to healthcare, showcasing examples where such models have significantly improved diagnostics, disease prediction, and patient outcomes. This project aims to build a comprehensive understanding of how supervised learning, and Deep Learning in particular, can drive impactful changes in the healthcare sector.

1.1 Objectives

- **Develop a foundational understanding of supervised learning techniques**: Explore how supervised learning methods work, emphasizing their role in solving classification tasks in healthcare.
- **Examine Deep Learning concepts and architectures**: Differentiate traditional feed-forward networks from advanced Deep Learning architectures, analyzing the benefits of additional hidden layers, complex connections, and hierarchical data representation.
- **Understand data partitioning and evaluation**: Learn the importance of splitting datasets into training and test sets to ensure model performance generalizes to new data, and explore cross-validation as a technique to further enhance reliability and robustness.
- **Implement overfitting prevention techniques**: Identify strategies to reduce overfitting, such as regularization, dropout, and early stopping, ensuring that models perform well not only on training data but also on new, unseen data.
- **Address data scarcity through data augmentation and synthetic data generation**: Explore how data scarcity impacts biomedical applications and implement data augmentation and synthetic data generation techniques to expand limited datasets and enhance model accuracy.
- **Analyze real-world applications of Deep Learning in healthcare**: Review case studies where Deep Learning has been successfully applied in the medical field, such as diagnostics and predictive models for disease progression, highlighting its potential to transform patient care.

2 Methodology and Results

- When developing a supervised method, why do we need to split the dataset into training and test sets? - What is cross-validation? - What is artificial neural networks? - What is Deep Learning (DL)? - What is new in DL models with respect to traditional feedforward neural networks? - Google (among others) has produced astonishing results in the application of DL models in different domains. Mention two of these cases describing shortly the problem solved. **Alphafold Geminy** - What is overfitting and how DL models avoid it? - Lack of data is a big limitation regarding the application of DL models to biomedical problems. What techniques can be applied to alleviate this problem.

3 Conclusion

4 Repository Access

All additional information, including the source code and full documentation of this project, is available in the GitHub repository [1].

References

[1] Marta Cuevas. Lab2_computational_learning. `https://github.com/martacuevasr/Lab2_Computational_learning`, 2024. Último acceso: 1 noviembre 2024.