Exploring Machine Learning Model Performance across Diverse Datasets: A Comparative Analysis

Yan Mazheika   Polina Petrova

May 8, 2024

Introduction

This report provides an analysis of the application of various machine learning projects on several datasets. The goal is to identify the best-use scenarios of model variants given the nature of input data. In this report, the authors analyze the

- Decision Tree, *provided by Polina Petrova*

- k Nearest Neighbors, *provided by Polina Petrova*

- Neural Network, *provided by Yan Mazheika*

- Random Forrest, *provided by Yan Mazheika*

against the handwritten digits, titanic survival, loan eligibility, and Parkinson's classification datasets.

It is our aim to explain the nature of the data, our models, and the performance of these classifiers. Throughout our report, we justify our algorithm choice for the given dataset and our choice of hyper-parameters for the tuning of the algorithm. These insights are for the reader's benefit; we also aim to provide insight into how to solve novel machine problems, which algorithms may work best, and how to adjust their hyper-parameters for optimal performance.

Digits Dataset

	Neural Network
Learning Rate α	0.0500
Regularization λ	0.0100
Architecture	[64, 32, 32, 10]
Mean Accuracy	0.9962
Mean F1-score	0.9809
Mean Test Cost	59.3100

Analysis

Titanic Dataset

Analysis

Loan Eligibility Dataset

Analysis

Oxford Parkinson's Disease Dataset

Analysis