

Hyperparameter Tuning on the Supervised Algorithm

Amrit Bag

Dec 5, 2024

Abstract

This report contains the results of Hyperparameter Tuning on Logistic Regression model applied to Instacert Dataset. A **Grid search** approach is used. A grid search approach was used to evaluate combinations of regularization strength (C), solver, and convergence tolerance (tol). The dataset was preprocessed to handle categorical features and split into training and test sets. The best-performing model achieved a test accuracy of **85.94%** with the following parameters: L1 penalty, $C=0.01$, solver liblinear, and $tol=0.0001$. The findings indicate that stronger regularization and precise convergence criteria improve model performance.

Context

1. Introduction.....	1
1.1 Tuning of Supervised Learning Algorithm.....	1
1.2 Tuning of Unsupervised Learning Algorithm.....	2
2. Methodology.....	2
2.1 Supervised learning.....	2
2.2.1 Hyperparameter Grid.....	2
2.2.2 Evolution Metric.....	2
3. Observation and Results.....	2
3.1 Supervised Learning.....	2
3.1.1 Result.....	2
3.1.2 Observation.....	3
5. Conclusion.....	3
5.1 Best Model for.....	3

1.Introduction

1.1 Tuning of Supervised Learning Algorithm

This report contains the process and result of **Hyperparameter Tuning** on a **Logistic Regression** Model using **Grid Search**. The objective is to find the optimal combination of the Hyperparameters that maximises the **accuracy** for classification of Instacart dataset

2. Methodology

2.1 Hyperparameter Grid

The following Hyperparameters were tuned using **Grid Search** technique:

- **C**: [0.01, 0.1, 1, 10, 100] (Inverse of regularization strength)
- **Solver**: ['liblinear'] (selected for compatibility with L1 penalty)
- **Tolerance(tol)**: [1e-4, 1e-3, 1e-2] (Stopping criteria)

2.2 Evolution Metric

- **Metric used**: Test Accuracy
- **Scoring**: Models were trained on the training set and evaluated on the test set using `accuracy_score`

3. Observation and Results

3.1 Result

Rank	Penalty	C	Solver	Tolerance	Test Accuracy
1	L1	0.01	liblinear	0.0001	0.8594
2	L1	1	liblinear	0.0001	0.8594
3	L1	10	liblinear	0.0001	0.8593
4	L1	100	liblinear	0.0001	0.8593
5	L1	0.1	liblinear	0.0001	0.8593

3.2 Observation

- Regularization Strength (C):**
 - Smaller values of C (e.g., 0.01) yielded the highest accuracy. This indicates that stronger regularization is beneficial for the dataset.
- Solver:**
 - The `liblinear` solver consistently performed well with the L1 penalty.
- Tolerance (tol):**
 - The lowest tolerance value (0.0001) provided the best results, indicating the importance of precise convergence criteria.
- Test Accuracy:**
 - The highest test accuracy achieved was **0.8594**, observed with two different combinations of C and tol.

5. Conclusion

5.1 Best Model

The optimal set of hyperparameters for the Logistic Regression model:

- **Penalty:** L1
 - **C:** 0.01
 - **Solver:** liblinear
 - **Tolerance:** 0.0001
 - **Test Accuracy:** **0.8594**
-