

First Model Evaluation Report: SGDClassifier

Paula Lozano Gonzalo

April 4, 2025

1 Model Summary

The SGDClassifier was trained on the diabetes dataset with the following best parameters found through grid search:

Best Model Parameters:

```
{
    'model__alpha': 0.0001,
    'model__average': False,
    'model__class_weight': None,
    'model__early_stopping': True,
    'model__epsilon': 0.1,
    'model__eta0': 0.1,
    'model__fit_intercept': True,
    'model__l1_ratio': 0.5,
    'model__learning_rate': 'adaptive',
    'model__loss': 'hinge',
    'model__max_iter': 1000,
    'model__n_iter_no_change': 5,
    'model__n_jobs': None,
    'model__penalty': 'elasticnet',
    'model__power_t': 0.5,
    'model__random_state': None,
    'model__shuffle': True,
    'model__tol': 0.001,
    'model__validation_fraction': 0.1,
    'model__verbose': 0,
    'model__warm_start': False
}
```

2 Performance Metrics

3 Analysis

The model shows significant overfitting:

- High training accuracy (0.961) vs moderate F1 score on validation data (0.545)

Metric	Training	Validation	Test
Accuracy	0.961	0.959	-
True Negative	64,608	6,875	-
False Positive	1,307	413	-
False Negative	2,232	217	-
True Positive	3,853	495	-
Precision	0.747	0.545	-
Recall	0.633	0.695	-
F1 Score	0.685	0.611	-

Table 1: Model performance metrics with confusion matrix results

- Precision-Recall metrics indicate room for improvement in classification quality

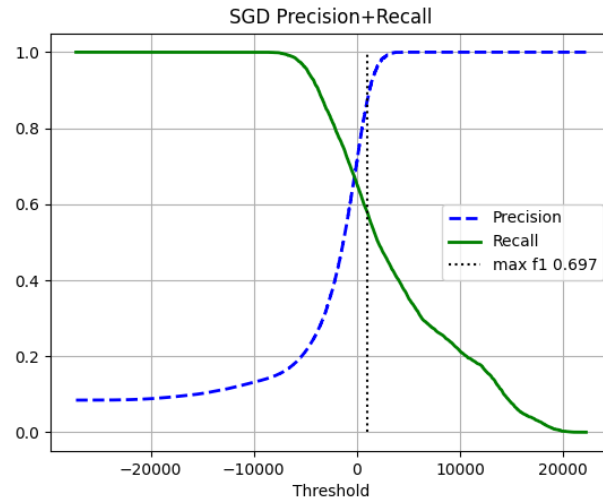


Figure 1: Precision-Recall Plot

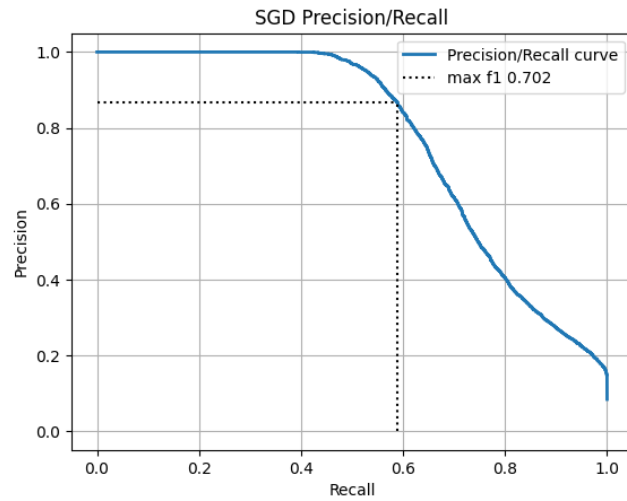


Figure 2: Precision-Recall Curve

4 Conclusion

The SGDClassifier achieved high accuracy scores but shows signs of overfitting. The discrepancy between validation accuracy and F1 score suggests potential issues with overfitting. This was just my starter model just like I always do. I will probably move to a forest classifier.