https://github.com/Gauravmadhukarkhandekar/activity-3.ipynb.git

Part 1: Binary Classification on MNIST

In this activity, the MNIST dataset was used to perform a binary classification task. Instead of using the commonly shown example of digits 3 and 5, a different digit was selected in order to practice classification on digits other than 3 or 5. In this case, the digit 2 was treated as the positive class, and all remaining digits were treated as the negative class.

An SGDClassifier was trained using the training portion of the dataset. Since MNIST contains a large number of samples, stochastic gradient descent is a good choice because it is efficient and works well with large datasets. Cross-validation was applied to evaluate the model rather than relying on a single train-test split.

Although the classifier achieved a high accuracy score, this value alone does not fully represent the model's performance. The dataset is imbalanced because most images are not of the digit 2. As a result, a model can obtain high accuracy even if it performs poorly on identifying the target digit.

To better understand the classifier's behavior, additional evaluation metrics were used. The confusion matrix showed the number of correct and incorrect predictions for both classes. Precision indicated how many of the images predicted as the digit 2 were actually correct, while recall measured how many actual instances of the digit 2 were successfully identified. The F1-score, which combines precision and recall, provided a more balanced evaluation of the model's performance.

To compare SGD with another machine learning approach, a Random Forest Classifier was also evaluated. ROC curves and ROC AUC scores were used for comparison. The Random Forest model achieved a higher ROC AUC score and showed better overall performance. This suggests that ensemble methods are better at capturing complex patterns in pixel-based image data compared to a linear classifier like SGD.

Part 2: Summary of Findings

From this experiment, it is clear that accuracy alone is not a reliable metric when working with imbalanced datasets. Precision, recall, and F1-score provide a more meaningful understanding of how well a classifier performs on the target class.

While the SGD classifier performed reasonably well and was computationally efficient, the Random Forest classifier produced better results across most evaluation metrics, particularly in terms of ROC AUC. This activity highlighted the importance of choosing appropriate evaluation metrics and showed how different models behave when applied to the same binary classification problem.