

My Colab link is:

<https://colab.research.google.com/drive/1FdAikPbr4ZA6wmMKmV8uUDPE-YBga27I>

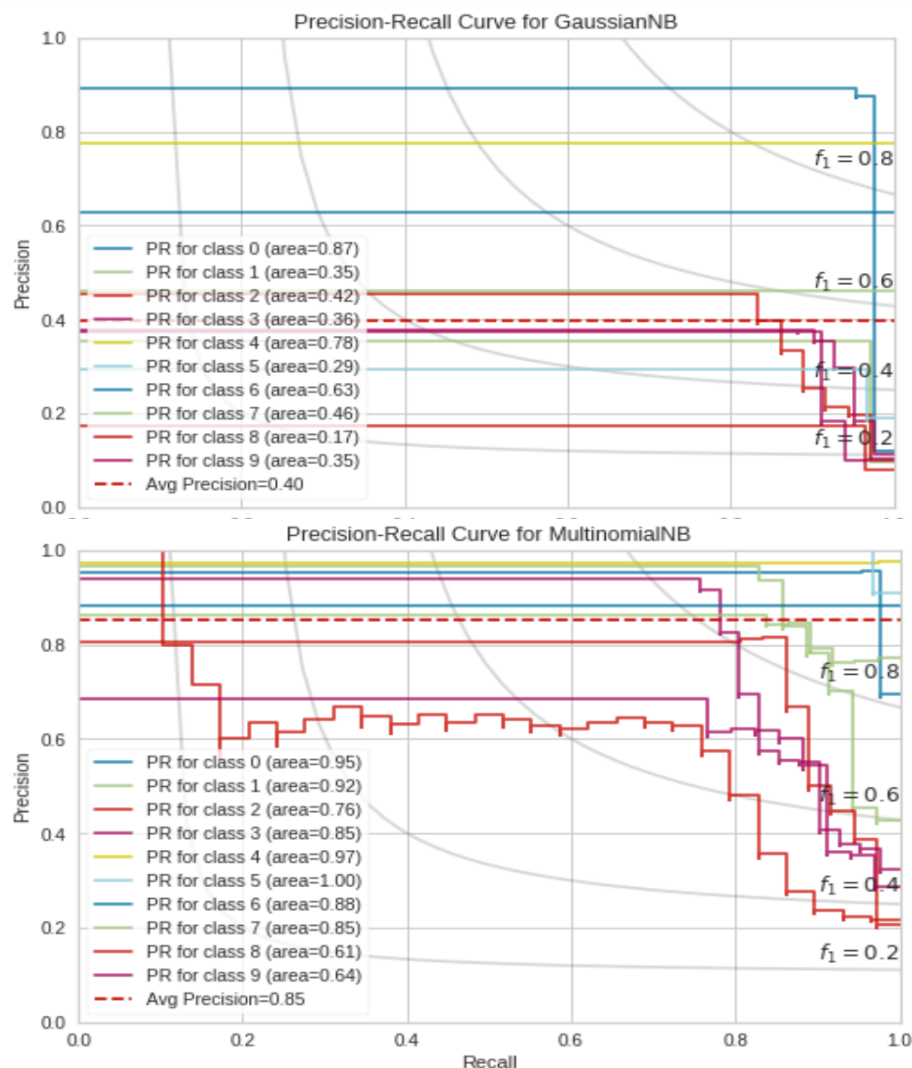
Overview

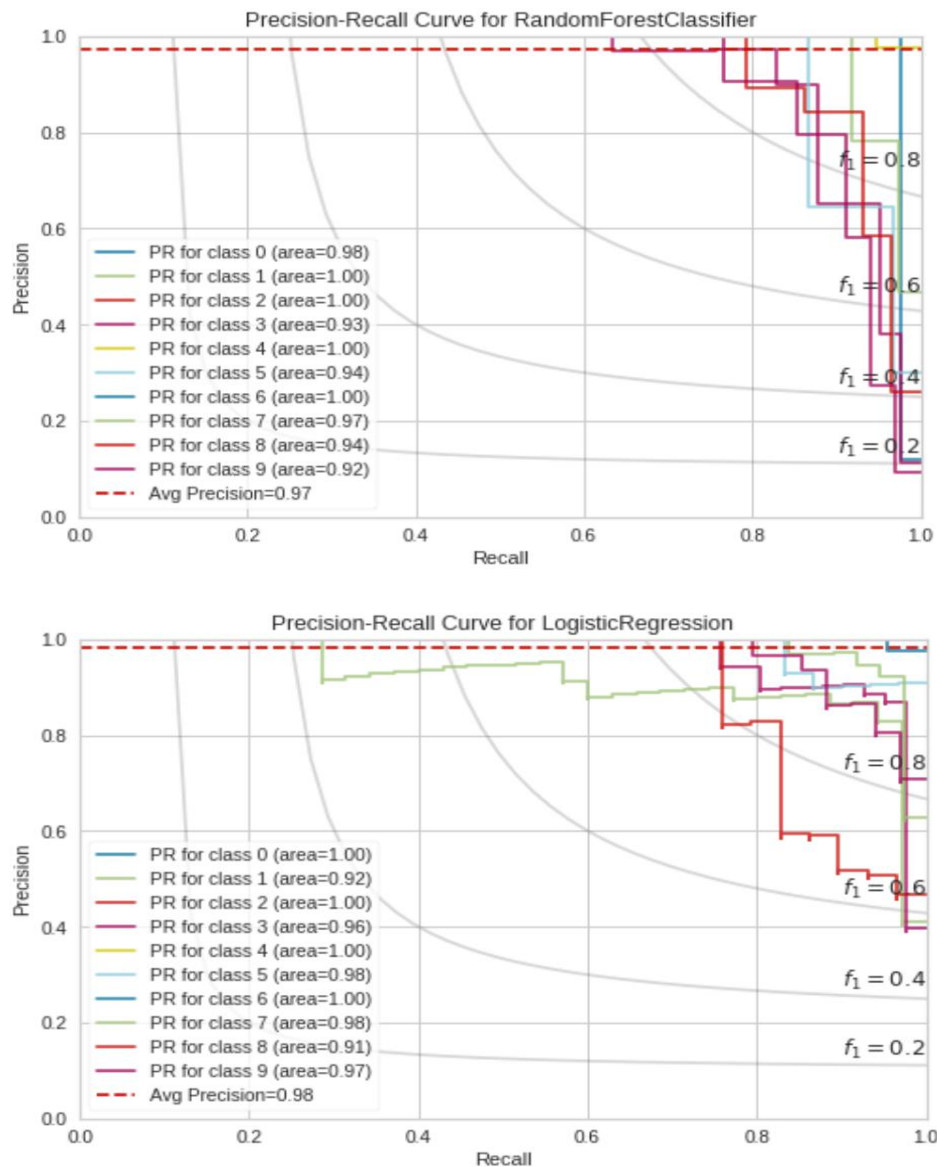
In this assignment, I compare four different models according to their accuracy scores and precision-recall curves. The data in this work is digit data.

Technique

First, I load the data and split the data randomly into train and test with `train_test_split` commend. Then, I use GaussianNB as my first model and the accuracy score is 0.3988. The next model I try is MultinomialNB, whose accuracy score is 0.8525. I also use Random Forest and Logistic Regression to fit the data and get the accuracy scores, 0.9715 and 0.9814.

To visualize the performances of these four models, I draw their precision-recall curves.





Conclusion

1. Logistic Regression has the highest accuracy score among the four models and the accuracy scores are consistent of the precision-recall curves. For the four models, the accuracy score increases and the curves in the plot also move to the right up corner, which means the model is becoming better.
2. The precision-recall curve shows the tradeoff between precision, a measure of result relevancy, and recall, a measure of how many relevant results are returned.
3. A large area under the curve represents both high recall and precision, the best-case scenario for a classifier, showing a model that returns accurate results for the majority of classes it selects.