

CS 464 Introduction to Machine Learning

Homework 3

Mustafa Gökten Güdükbay

21801740

Question 1

Question 1.1

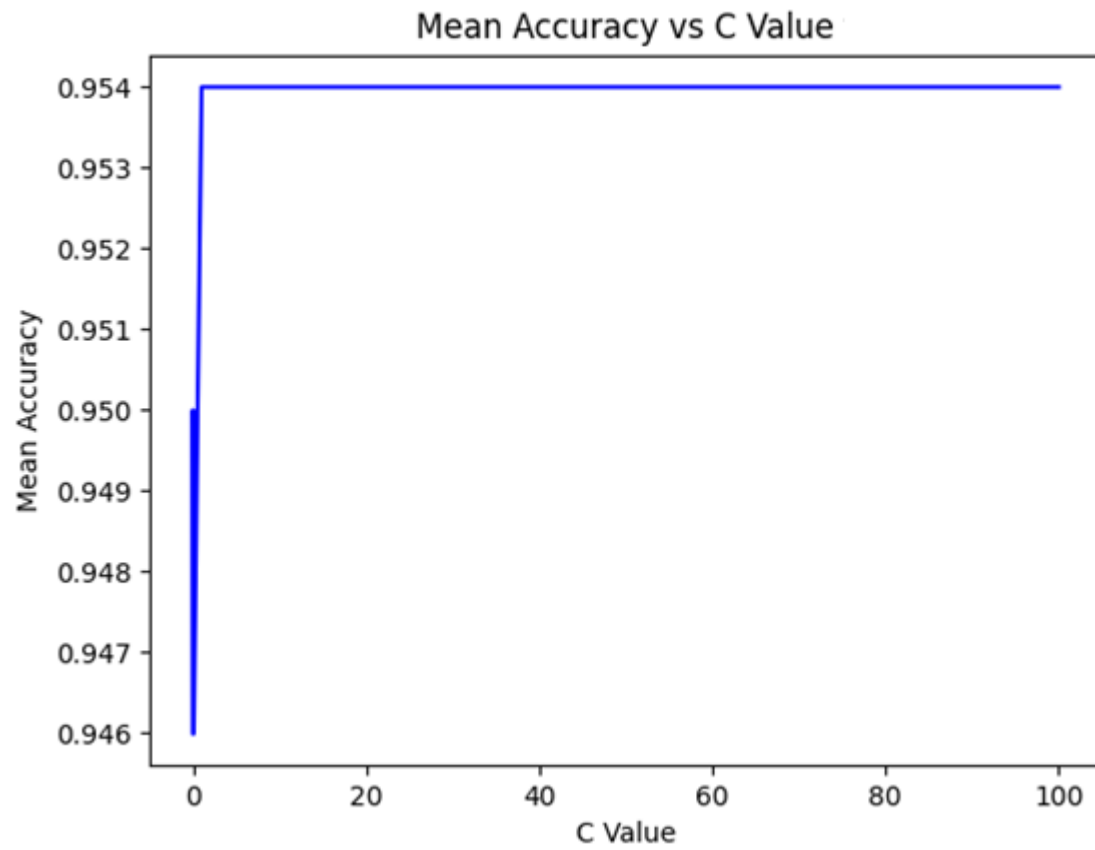


Figure 1. Mean Accuracy vs. C Value

Output

Part 1

Best C Value: 1.0

Confusion Matrix

Classifier	Actual	
	Malignant Tumor	Benign Tumor
Malignant Tumor	44	2
Benign Tumor	0	153

Accuracy: 0.9899497487437185

Precision: 0.9565217391304348

Recall: 1.0

Negative Predictive Values: 1.0

False Positive Rate: 0.012903225806451613

False Discovery Rate: 0.043478260869565216

F1 Score: 0.9777777777777777

F2 Score: 0.990990990990991

Discussion

The c values are 1, 10, and 100 have the best mean accuracy on the validation set. However, I choose 1 for the c value for the model that was tested on the test set. The test set was not balanced, so it may be deceiving to look at accuracy only. Although the recall is 100%, the precision is not 100%. Similarly, the negative predictive value is 100%. If the model predicts Benign Tumor, it is 100% correct. False-positive and false discovery rates are very low; however, the data was imbalanced; so they can be misleading. The F1 and F2 scores were good, which means that the data performed well.

Question 1.2

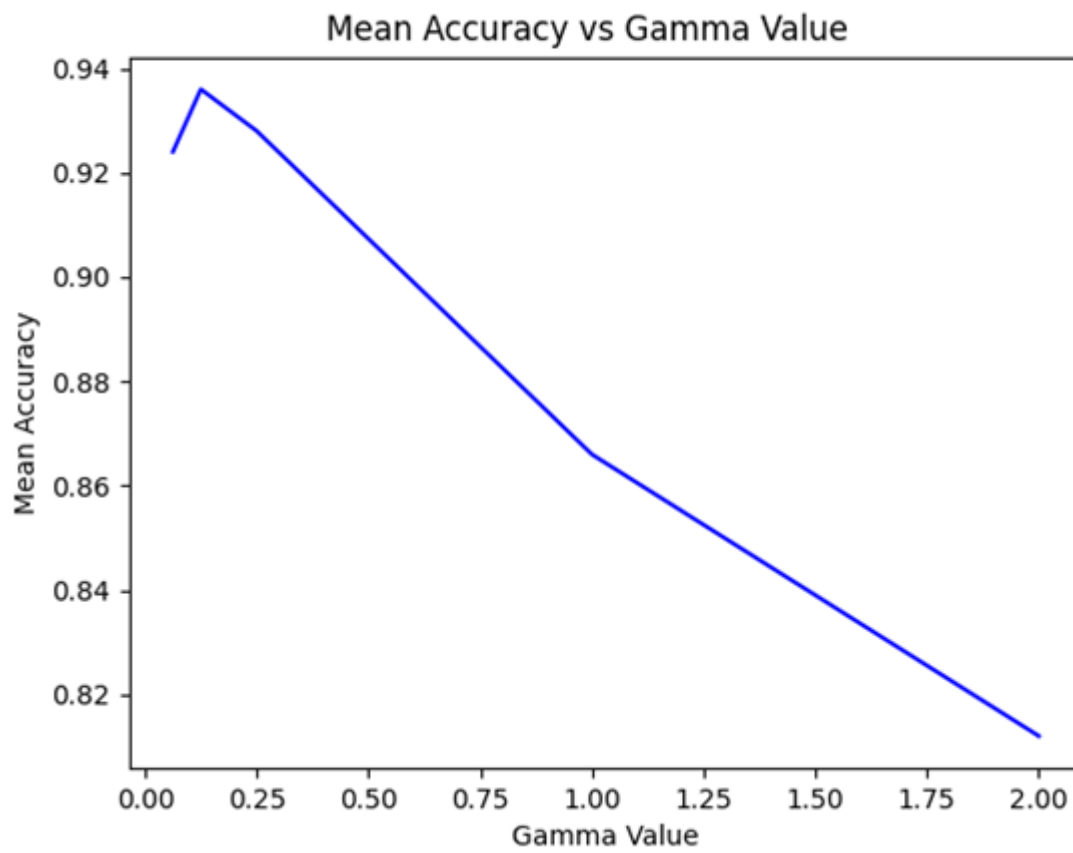


Figure 2. Mean Accuracy versus Gamma Value

Output

Part 2

Best Gamma Value: 0.125

Confusion Matrix

Classifier	Actual	
	Malignant Tumor	Benign Tumor
Malignant Tumor	44	2
Benign Tumor	0	153

Accuracy: 0.9899497487437185

Precision: 0.9565217391304348

Recall: 1.0

Negative Predictive Values: 1.0

False Positive Rate: 0.012903225806451613

False Discovery Rate: 0.043478260869565216

F1 Score: 0.9777777777777777

F2 Score: 0.990990990990991

Discussion

The gamma value 0.125 had the best mean accuracy on the validation set. The confusion matrix was similar to the previous part, so the scores were the same. Both models are suitable for processing this data.