student1 number: 1000698328
student1 utorID: gnanaran
student2 number: 1001716869
student2 utorID: huyi10

1)

Yes, the points from the different classes seem to be hard to differentiate when only examining points, but when colored they are reasonably separable and identifiable.

2)

"for dimension 3"
"Classification accuracy on training set: 65.4088"
"Classification accuracy on testing set: 64.2086"

"for dimension 5"
"Classification accuracy on training set: 85.5346"
"Classification accuracy on testing set: 84.4424"

"for dimension 6"
"Classification accuracy on training set: 86.5229"
"Classification accuracy on testing set: 85.7914"

"for dimension 7"
"Classification accuracy on training set: 87.062"
"Classification accuracy on testing set: 85.8813"

"for dimension 9"
"Classification accuracy on training set: 89.1285"
"Classification accuracy on testing set: 87.1403"

"for dimension 10"
"Classification accuracy on training set: 88.9488"
"Classification accuracy on testing set: 88.759"

"for dimension 15"
"Classification accuracy on training set: 92.4528"
"Classification accuracy on testing set: 92.0863"

The optimal dimensionality for the GCC classifier is 15 for our testing cases because the accuracy of 15 dimension on train and testing is 92% which higher than the rest. If the dimensionality were to increase from 15, the accuracy will also relatively increase.

3)

"for full-dimensionality data"
"Classification accuracy on training set: 100"
"Classification accuracy on testing set: 95.3237"

For this classification problem, NB is a better classifier due to the accuracy being 100% on the training set, and 95% on the testing set. This is because NB can use full dimensionality and so works well with large data set. The weaker classifier, GCC, is less accurate due to it not being able to work well with full dimensionality. As our data is very large and after applying PCA we are not using our entire data set, only the PCs with the highest variability, this means the accuracy of our training data would be reduced which is not what happed with NB.