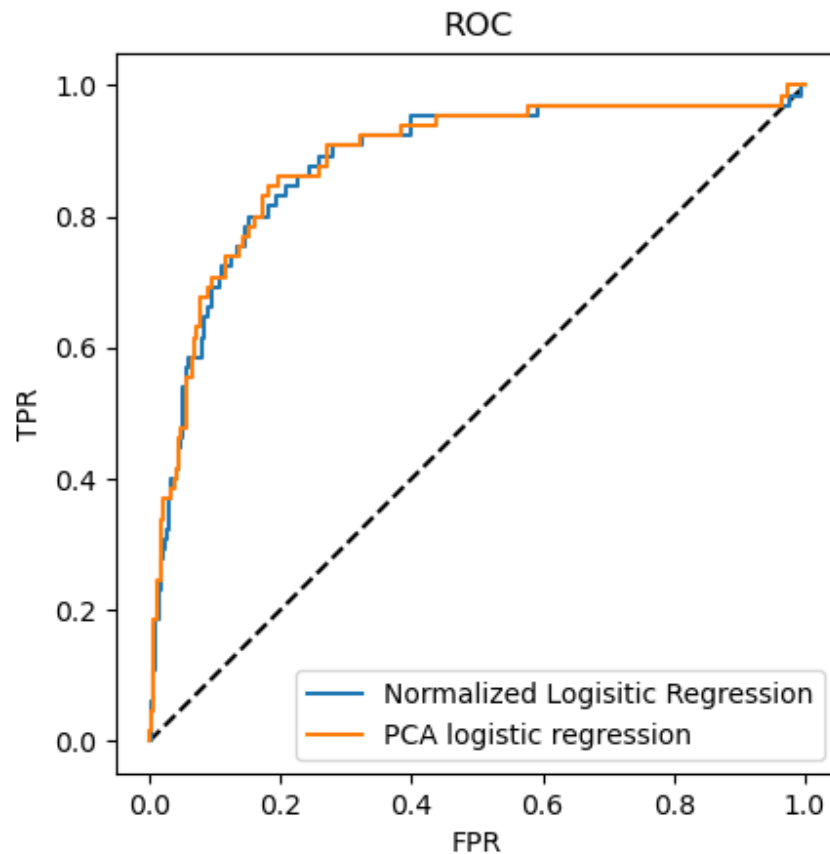


1d)

AUC for unregularized Logistic regression on normalized data: 0.879925857275255

AUC for unregularized Logistic regression on unnormalized data: 0.8834105653382762



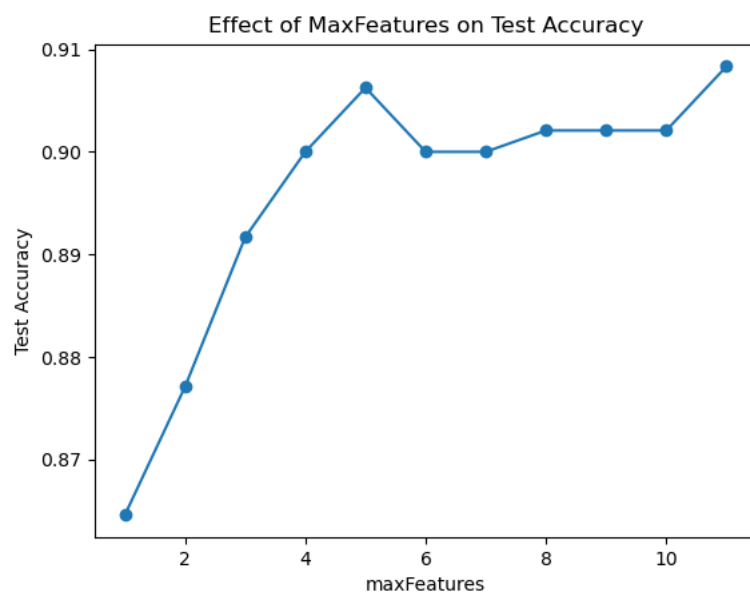
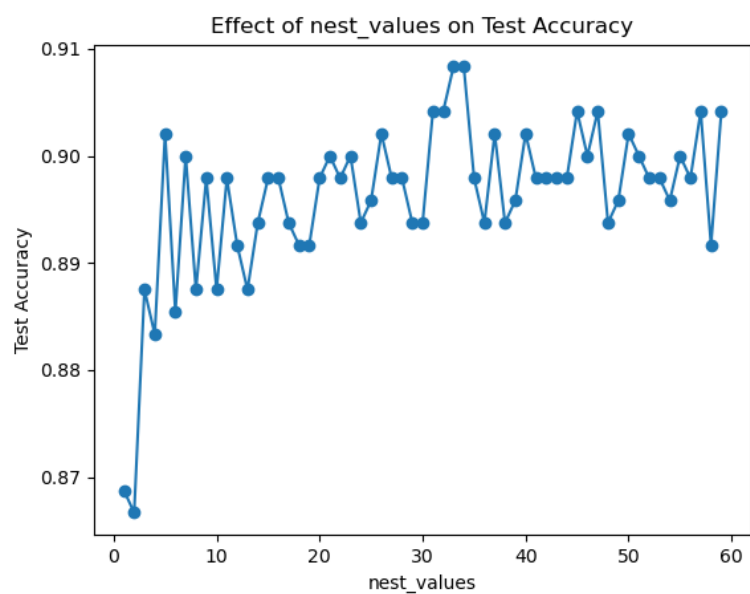
My plot seems to imply the two shapes are very similar. Neither curve dominates the other, and neither curve has a particular advantage over the other for any meaningful amount of space.

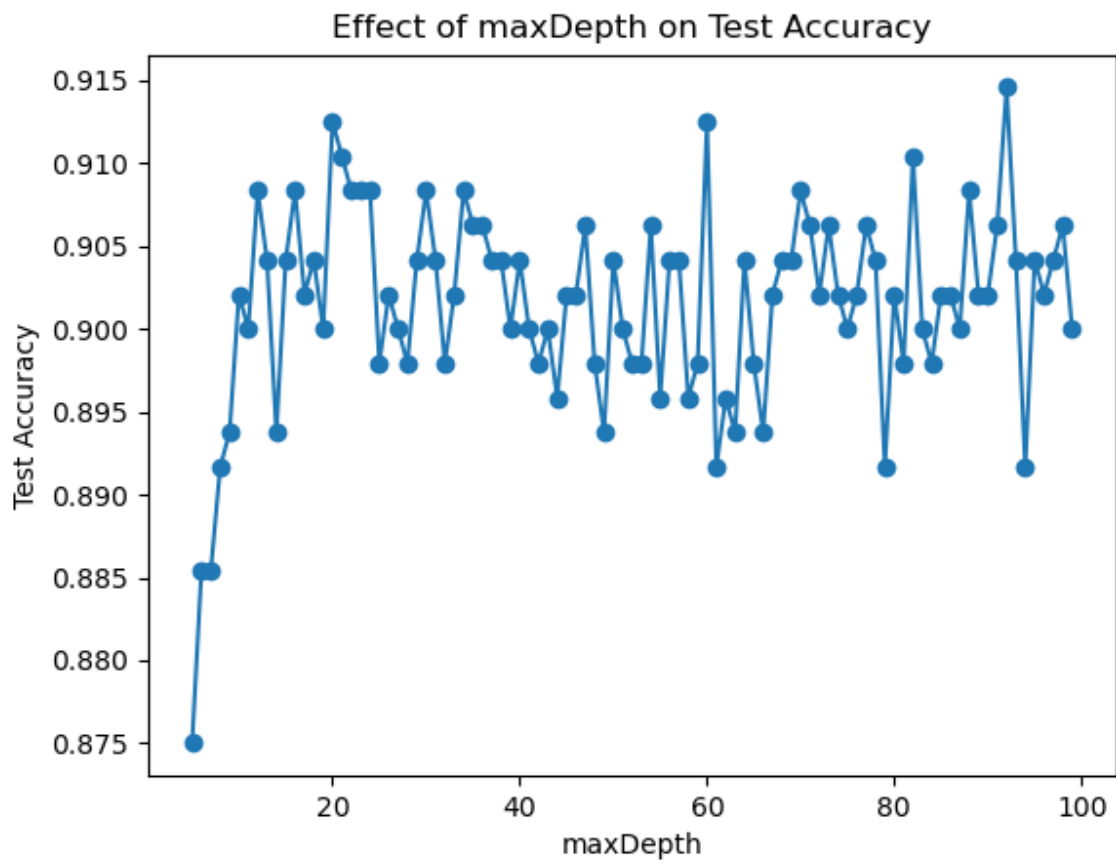
2b)

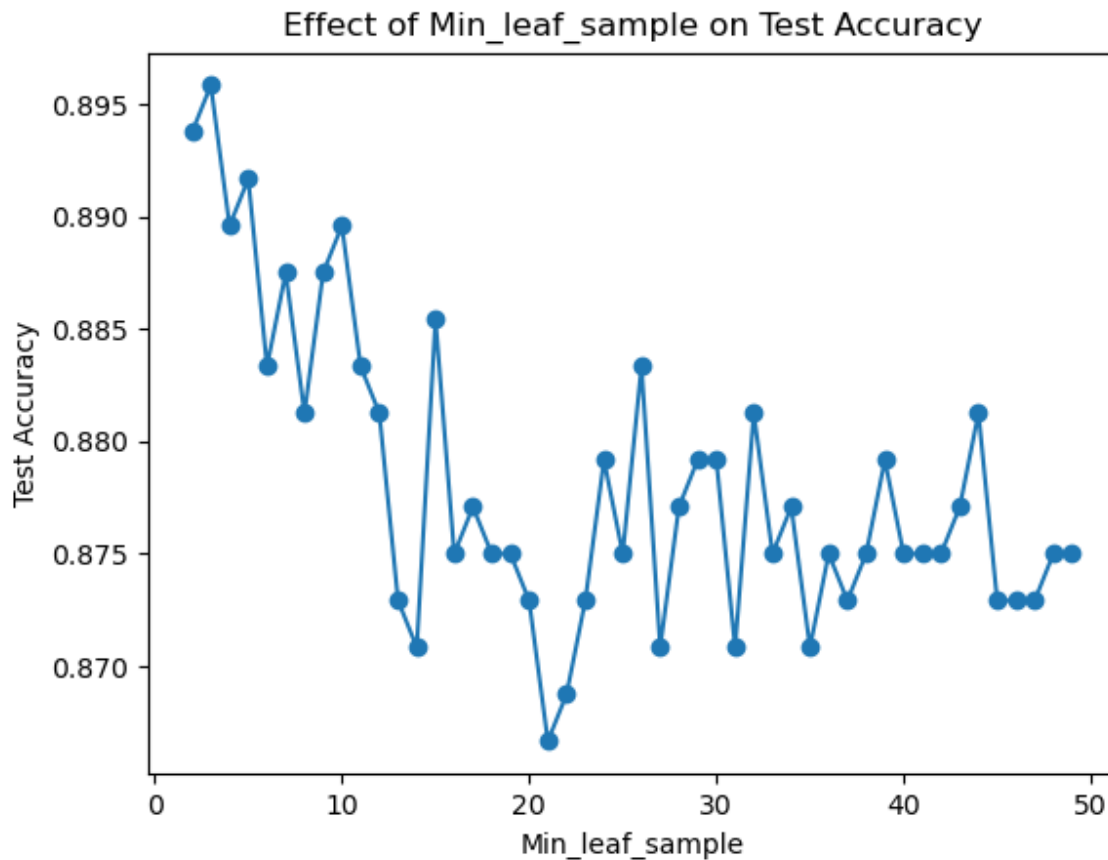
I used these defaults when running individual tests on each of the parameters:

```
model = RandomForest(50, 5, "gini", 10, 2)
```

Here are some graphs when only changing such variables:







1c)

I then picked the generally best value from each. I decided on 35 trees, 5 features, gini, a depth of 25, and a min leaf sample of 2.

My test accuracy was 0.9125.

My out-of-bag error on the final tree was 0.02980625931445602.

I think that this seems to be a relatively successful model. With a high test accuracy and a low oob error, my model seems to have a high generalizability.