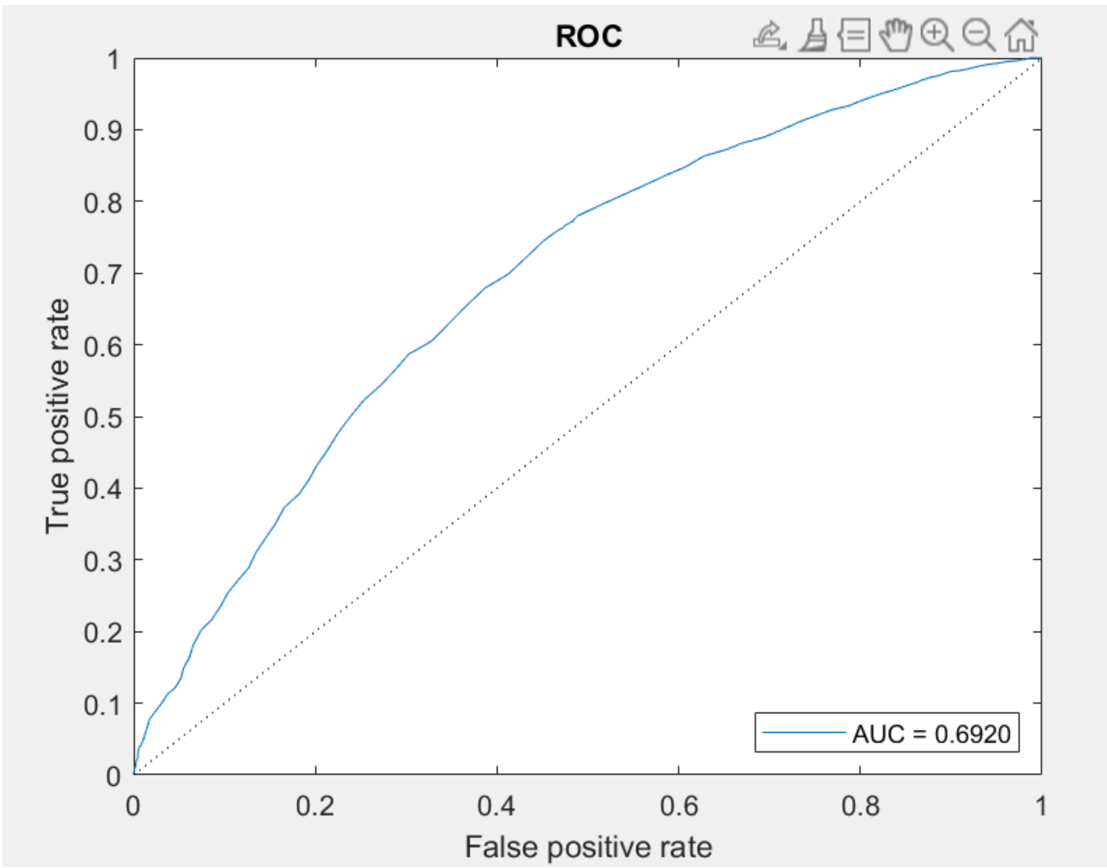
Steganography

Tientso Ning

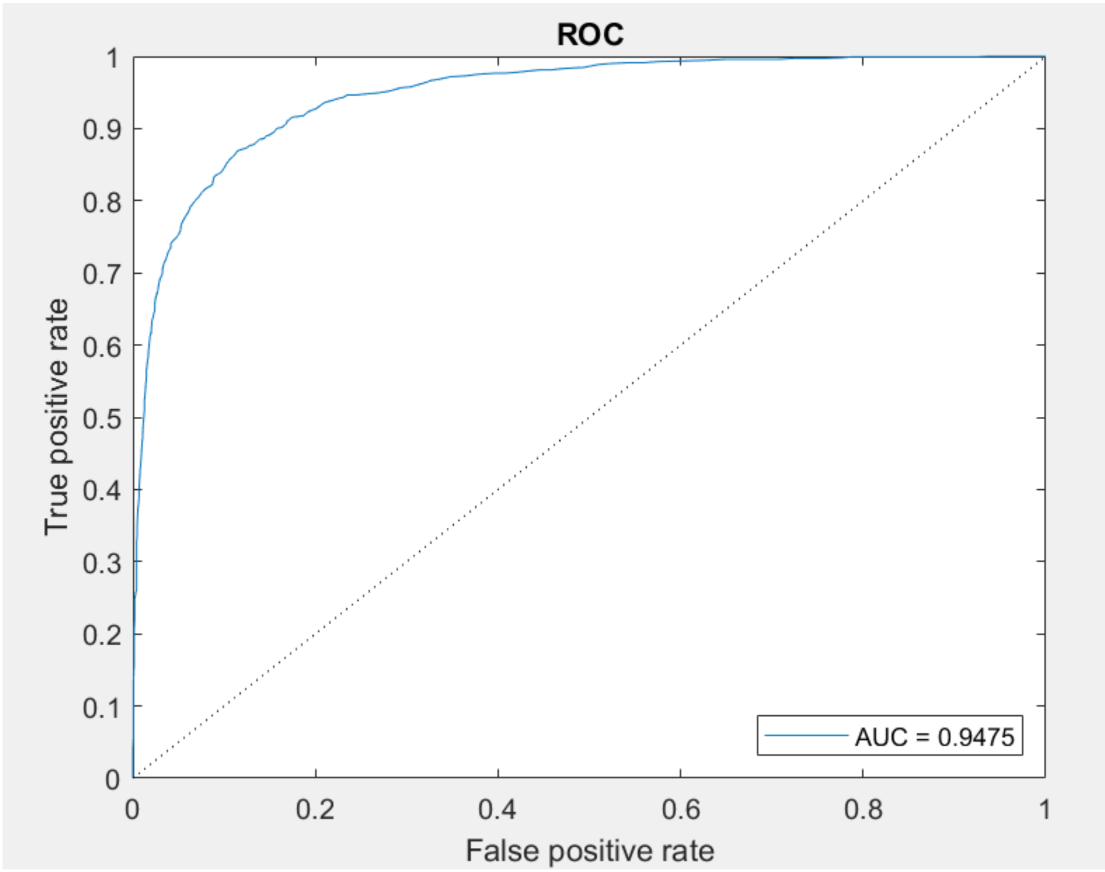
Exercise 2

Investigate the impact of the size of training set on the quality of steganalysis.

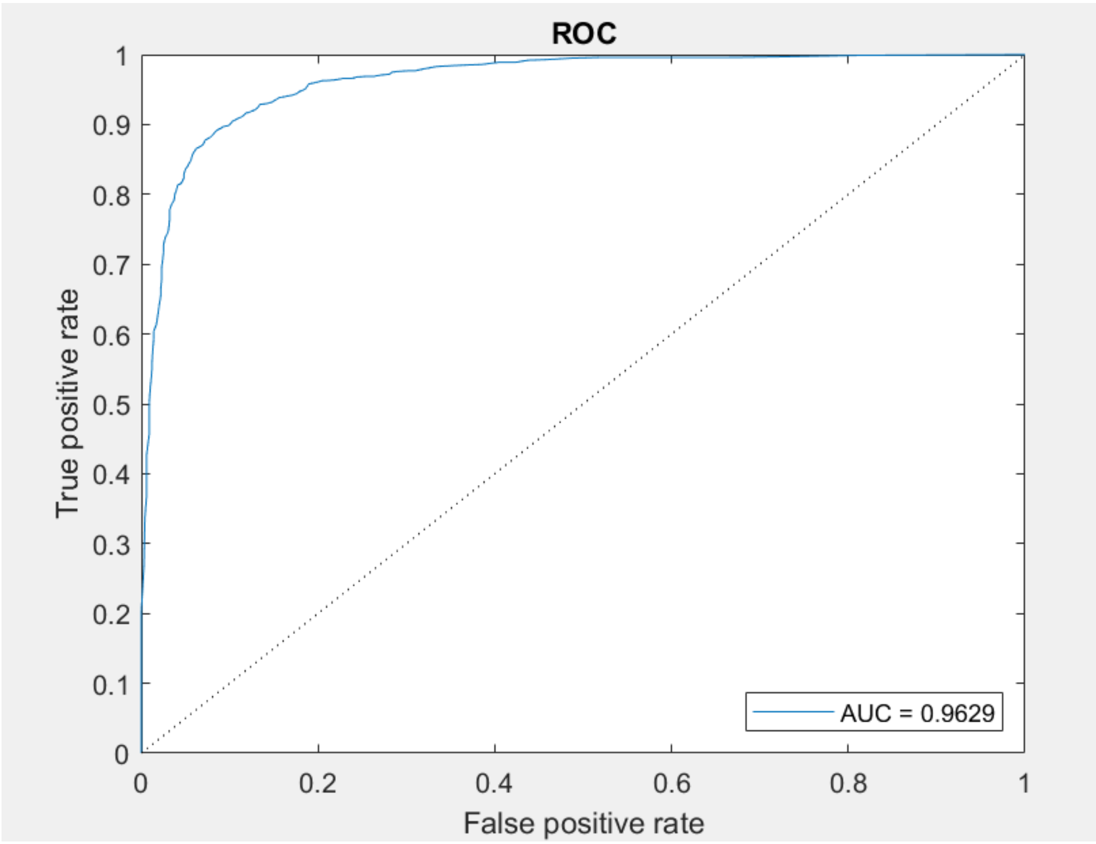
Training Size = 0.01 (1% set aside for training)



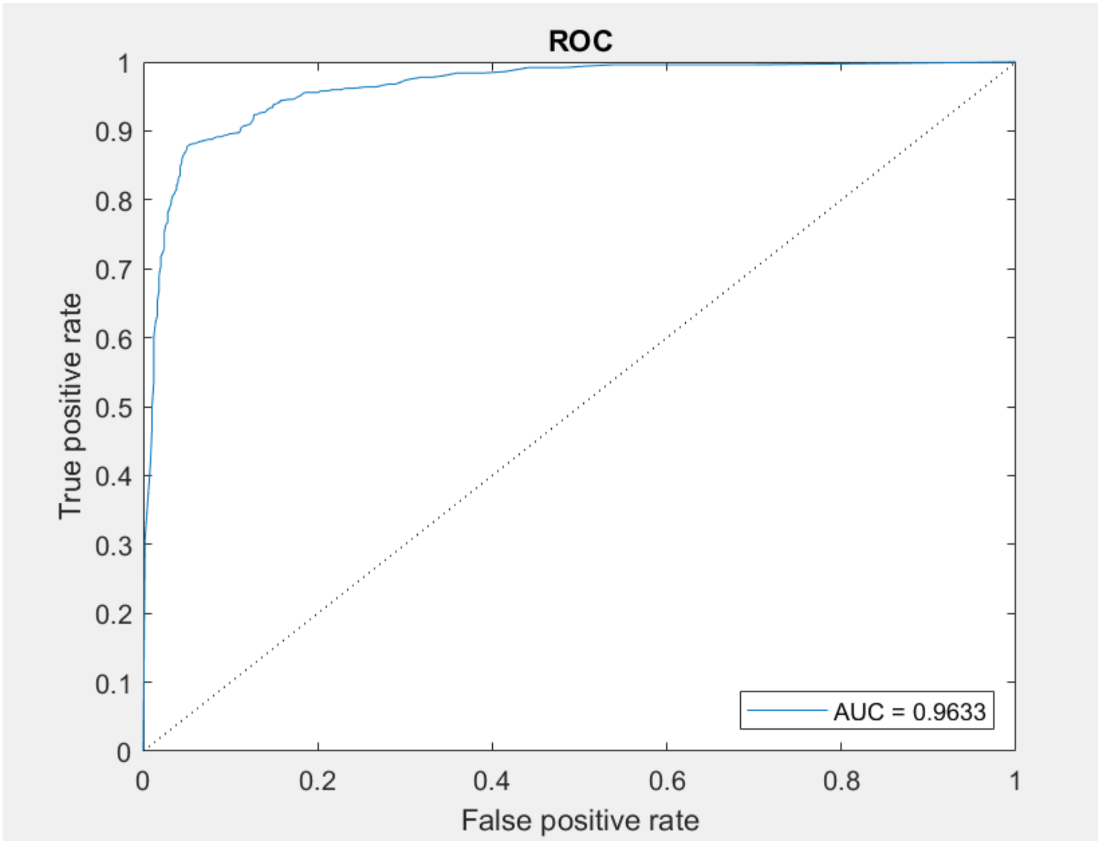
Training Size = 0.25 (25% set aside for training)



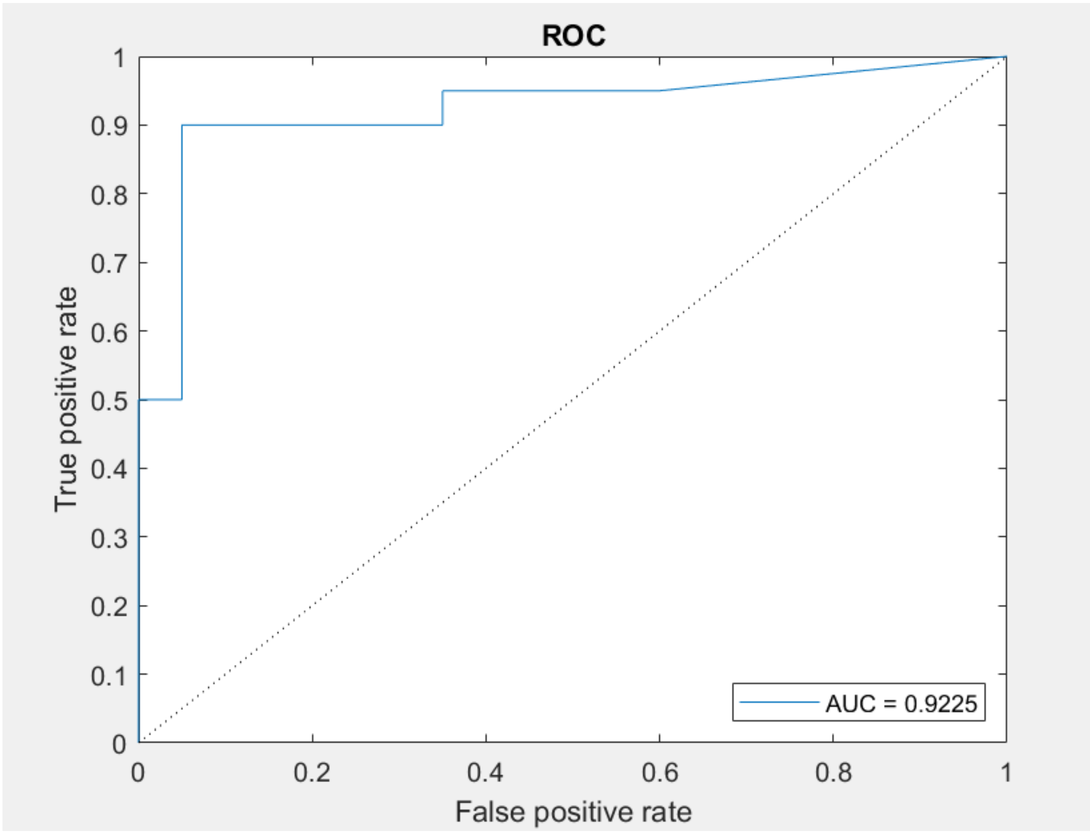
Training Size = 0.5 (50% set aside, also the default setting)



Training Size = 0.75 (75% set aside for training)



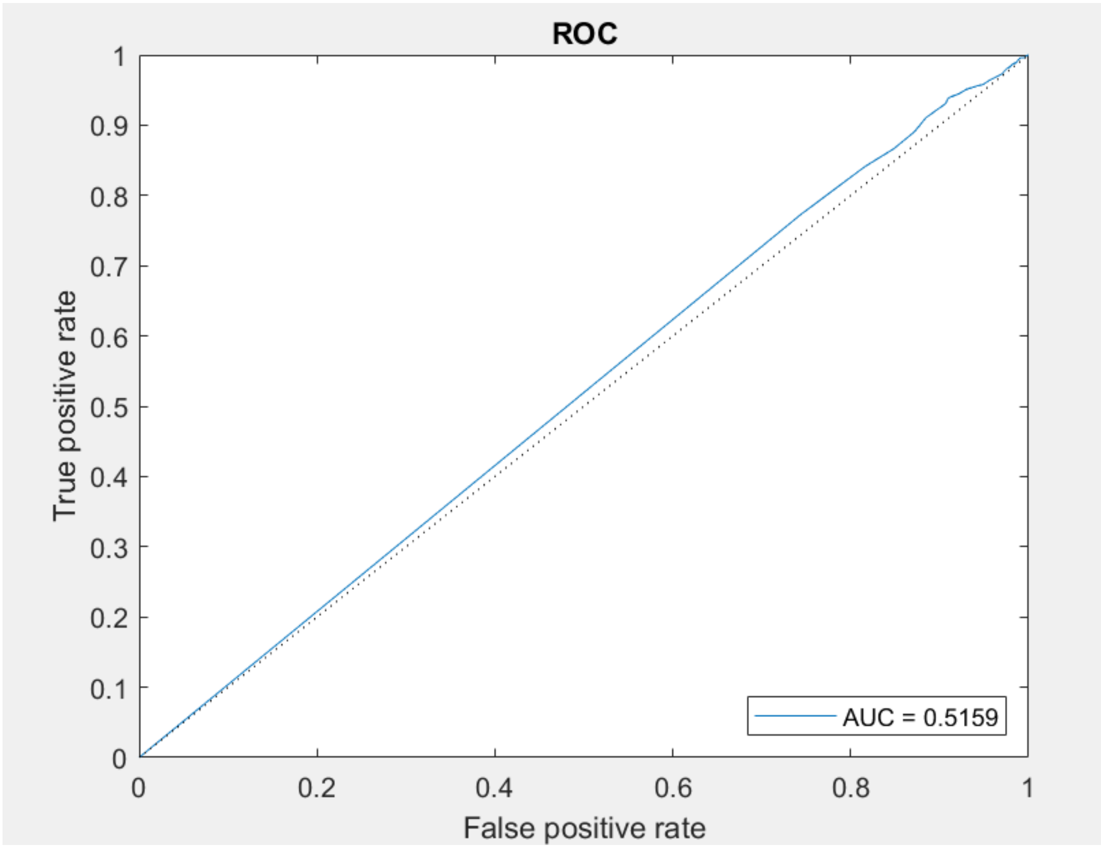
Training Size = 0.99 (99% set aside for training)



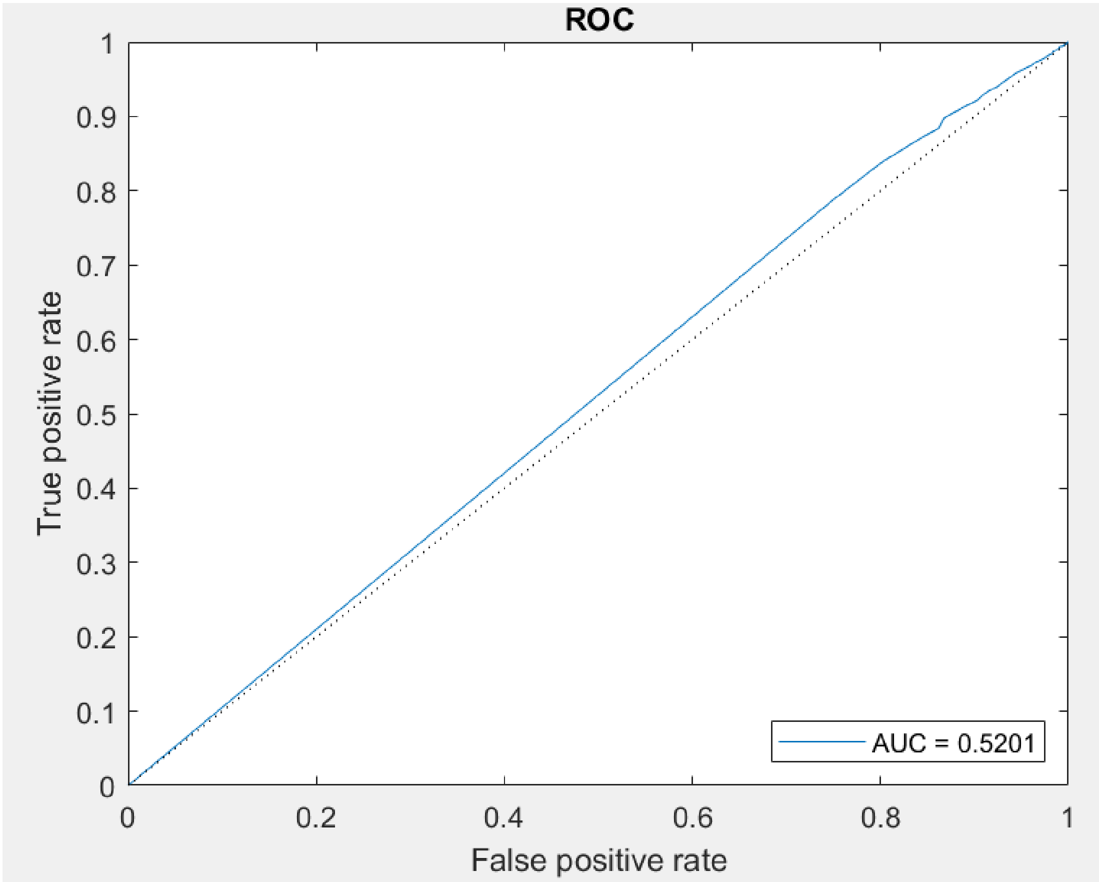
Conclusion: As we increase the training size set, we are also shifting the burden on a decreasing test set. As a result, we can see that if we use too little training data (1%) that the accuracy is not great, middling around 69%. However, we can also see that if we increase the training data size too much (99%), we aren't able to generalize well, and the accuracy also suffers.

Investigate the impact of cover data distortions on the quality of steganalysis.

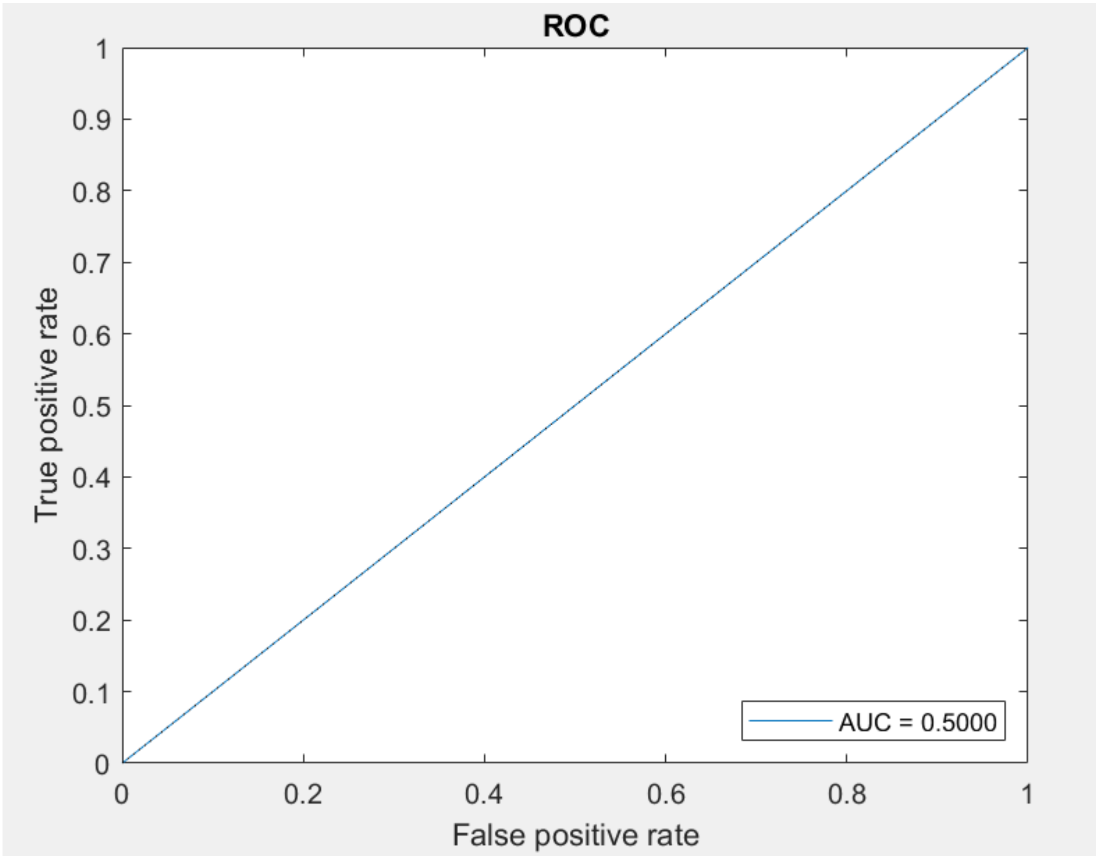
<u>Gaussian noise, mean 1, std 1</u>



<u>Gaussian noise, mean 1, std 10</u>



Gaussian noise, mean 2, std 5



Conclusion: Here we can see that even a basic addition of noise (gaussian noise) makes the accuracy of the model more or less of a complete guess (close to fifty-fifty). We expect this since the distortions of the cover data can hide or ruin the statistical visibility in the features that the classifiers attempt to look for, making each classification essentially a guess.