Project: Machine Learning Mine Versus Rock

This project uses MLP classifiers to differentiate the hidden mines from the rocks in the ocean. The given dataset is split into 30% test data and 70% training data. Using PCA the accuracy is measured for different numbers of components ranging from 1 to 60. The different values of the random state of the MLP classifier does not cause any significant change in the accuracy, as long as the same random state is used for all the components. The accuracy which is obtained from using different random state values for different components is not accurate, as each value of the component gets trained in a different model. From running the program multiple times, the accuracy was found to be almost always above 65%. The average maximum accuracy is 94% and the average number of components to achieve the maximum accuracy is 7. This could be because reducing some features with PCA might help reduce overfitting and hence increase the accuracy. Hence the overall graph starts low and reaches a peak then drops a little and stays steady till it drops a little more again towards the end. The hidden layer parameter of the MLP classifier was chosen to be 100, since the increased number of hidden layers results in a higher accuracy. The solver was chosen as ‘adam’ since ‘adam’ is known to perform well with large datasets. Since we are using ‘adam’, we need the total number of iterations(max_iter) to be specified as well, by default the value is 200, but I used it as 2000. Since we need the sigmoid function, the logistic activation is used. The tolerance for optimization(tol) is useful when the loss or score is not improving by at least tol for n consecutive iterations, convergence is considered to be reached and training stops. The confusion matrix tells about the performance about the classifier. Knowing the true positives and false negatives helps us to know about the prediction errors made.