Report on Project 2

The results obtained show that after performing PCA for only 7 components, the accuracy is 95% which means there is 95% chance of survival from a mine with the data provided.

The confusion matrix tells that True positive(mine) are 30 and True negatives are 21 and 4 false positives and 8 false negatives at the maximum accuracy.

The plot for components versus accuracy is fairly uneven but it has a sharp increase as we increase the components to 4 components. PCA for 7 components give 95% accuracy which says that 7 principal components are enough to determine the output accurately. Even if we provide more components, it is only redundant data which does not contribute to the variance in result. In fact, The accuracy might decrease as we increase the components as the additional of a new principal component might result in making the already fitted sample an outlier and at the same time doesn't fit the already outlier samples which leads to reduction in accuracy . Accuracy increases sometimes a little if addition of a component fit a few outlier samples.

For the layers less than 200, iterations<2000, any other activation function apart from logistic, any other solver apart from "adam", model is *underfit* for the data: that is, it does not have enough model flexibility to suitably account for all the features in the data.

For Layers more than 200, iterations >200, the model fit has enough flexibility to nearly perfectly account for the fine features in the data, but even though it very accurately describes the training data, its precise form seems to include noise rather than the intrinsic properties of data.

So, I selected these parameters for MLP classifier.

-SAMPRITI NEOG

(1217397714)