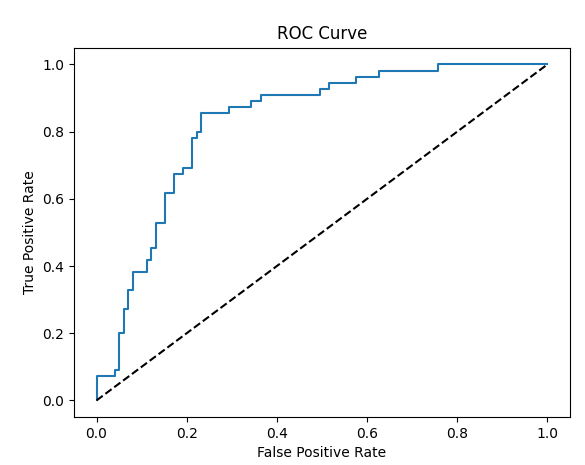
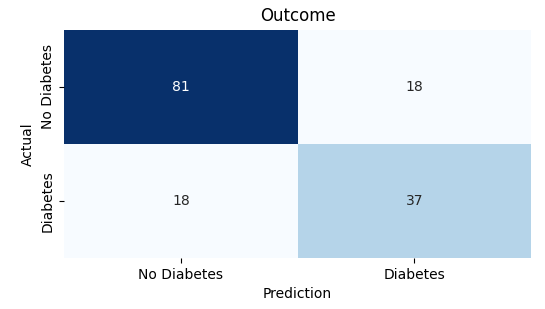
3a -- On Altering Hidden Layers

I found that the architecture of 8, 32, 16, 1 most often gave the highest results in the validation and training accuracies. However, I did find that the architecture of 8, 30, 40, 1 regularly gave a better ROC plot and better results in the confusion matrix, as such

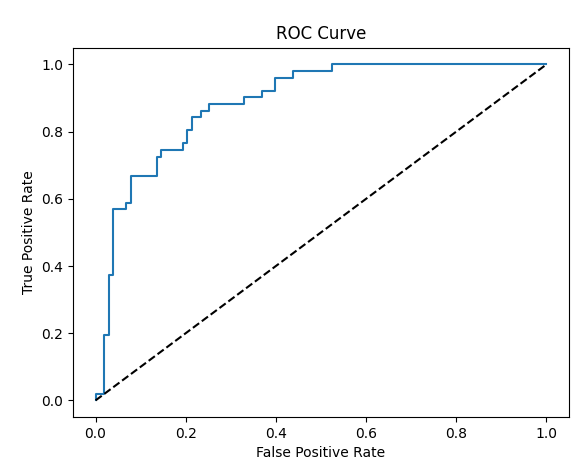
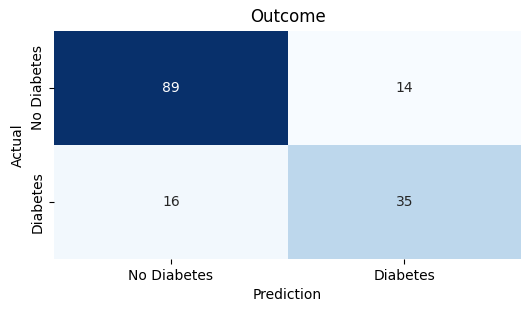


but would also tend to perform worse in its accuracy overall, averaging 90-95% training accuracy and 65-75% validation & testing accuracies. Every other architecture I tried was worse in every category.

3b – On Altering Compile() Parameters (including the optimizer)

This step was somewhat confusing, as the optimizer is part of the compile() function, but the URl links to the fit() function, and the next step says to modify the optimizer, which would theoretically be part of this step. Here is the data that resulted in modifying the compile() function, including changing the optimizer.

Upon changing the parameters in the compile() function, I found that the optimizer SGD combined with 10 steps\_per\_execution performed better in the confusion matrix / ROC Curve than the Adam optimizer at only 1 step\_per\_execution, resulting in the following data:

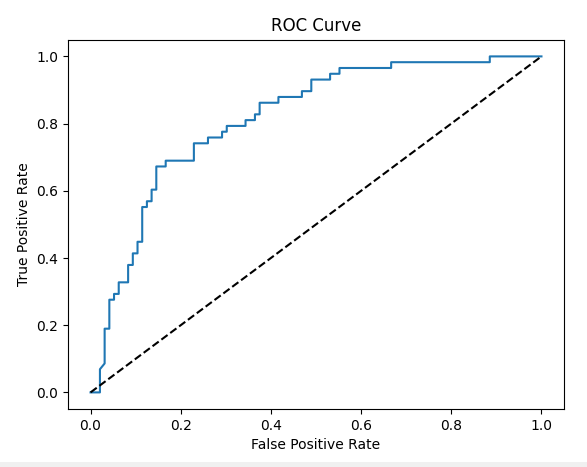


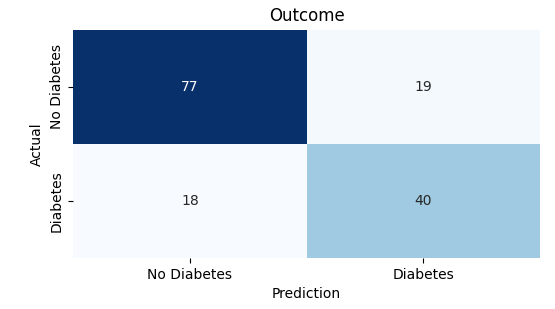
Comparing this graph with the previous step, it seems as though this change made the MLP perform better than altering the number of hidden nodes and could potentially perform even better if combined with earlier alterations. Other changes made to the parameters in the combine() function only seemed to worsen the matrix / ROC Curve and the overall accuracy of the MLP.

3c – On Altering the Fit() Function

Here I wanted to include alterations to the fit() function, as I think that was supposed to be the previous step in this process despite the exact wording.

I found that decreasing the batch\_size bettered the ROC Curve and Confusion Matrix while decreasing the overall accuracy of the MLP, resulting in the following graphs:





This is probably the best confusion matrix encountered in altering any of the previous values, yet the worst accuracy.

In all the alterations made throughout this assignment, there seems to be an inverse relationship between good ROC Curves & Confusion Matrices and MLP accuracies, which seems counterintuitive and leads me to believe there is a strange correlation I’m not able to pick up on based on this simple intro to MLPs. Or, possibly less likely, there was a long string of bad luck that resulted in low accuracy values and high ROC Curve areas / correct confusion matrices. Either way, from the data I collected, there is no clear correct path depending on which values should be min-maxed, or what constitutes the “best” results.