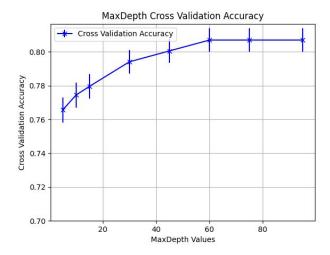
Tune maxDepth on the adult data set with and without numeric features. Create a short (~3-4 figures and 200 words) writeup of what you learned from the tuning runs. Include all the usual elements (bounds, parameter sweeps, ROC curves). Is it better to use the numeric features or not? (Be precise)

## **Categorical Features Cross Validation**

Best MaxDepth: 60

Best MaxDepth Cross Validation Set Accuracy (95% confidence interval): 0.81432 +/- 0.01357

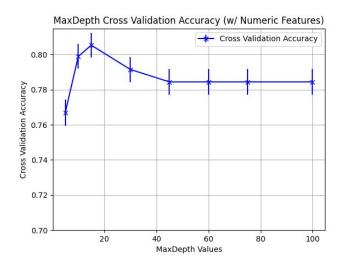


From this tuning run, it seems with categorical features, there is a limit (~60 deep) where the results no longer improves and all you are adding is run time. Also with very shallow trees, it seems it basically regresses to a most common model.

## Categorical + Numerical Features Cross Validation

Best MaxDepth: 10

Best MaxDepth Validation Set Accuracy (95% confidence interval): 0.7890 +/- 0.01424



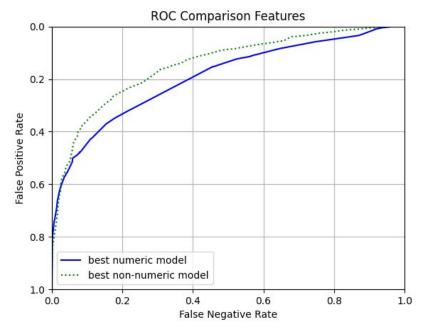
From this tune run, it seems that right around 15 deep is a tree that reports the best cross-validation results, although it is insufficiently better than a 10 deep tree with sufficient certainty (95%) to trade-off against the runtime. It seems adding numerical features actually made the model worse. The numeric features added were:

- 1. Age
- 2. Education number (no idea what this is)
- 3. Fnlwgt (No idea what this is)
- 4. Capital Loss
- 5. Capital Gain
- 6. Hours worked

It is surprising that this did not provide better segmentation of the results. While 4,5,6 would be understandable given that the data was rather poor in these columns (basically just 0's or 40's), the fact age didn't have an effect is interesting.

And similarly to categorical-only, there seems to be a plateau once the tree is 40 some deep.

## Categorical vs. Categorical + Numerical ROC curve @ best MaxDepth



From the ROC graph, it seems that on average, the categorical-only model is uniformly better on false-positives and false-negatives across all thresholds. Further from our two previous tuning runs, it seems that the categorical-only model is better, although not at a 95% confidence interval, but probably at around 75% confidence interval (I mistakenly ran a 50% confidence interval and it was better, so rather than running these models again, I thought I would just take the average as best guess).