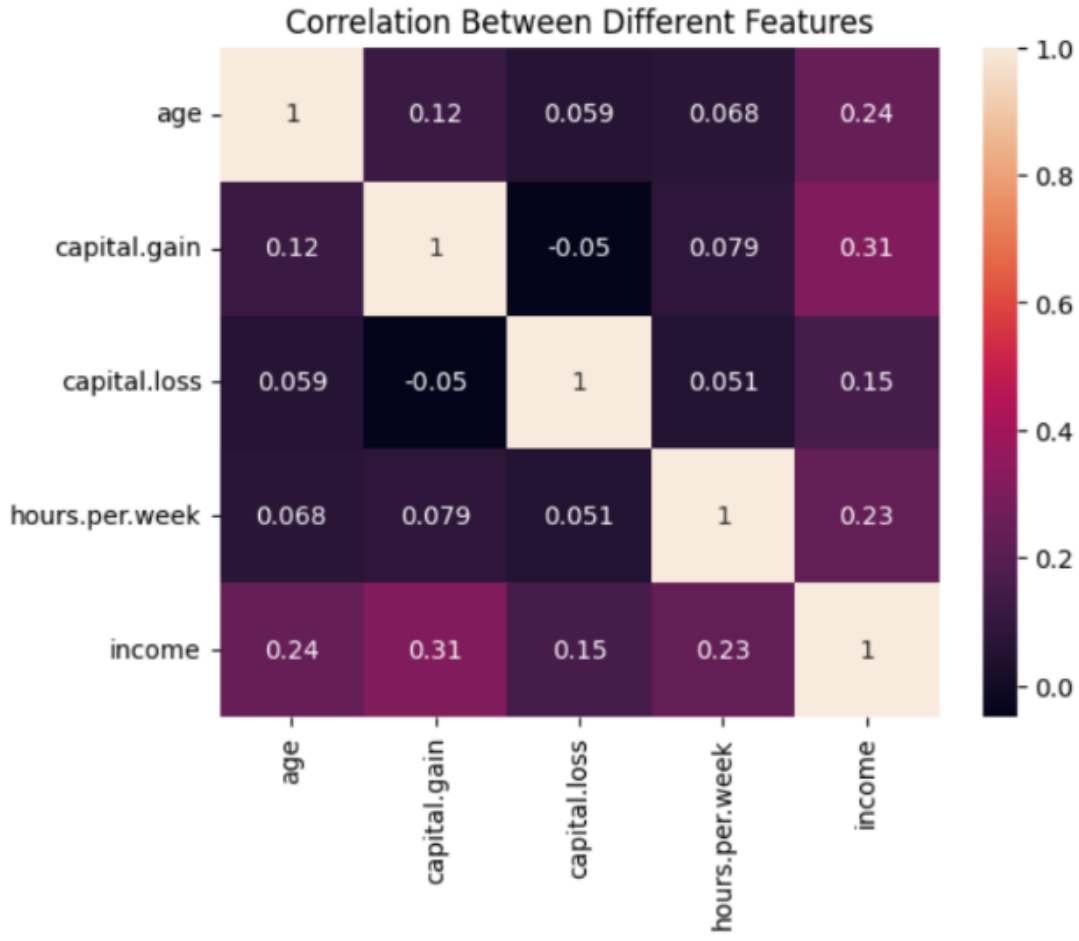
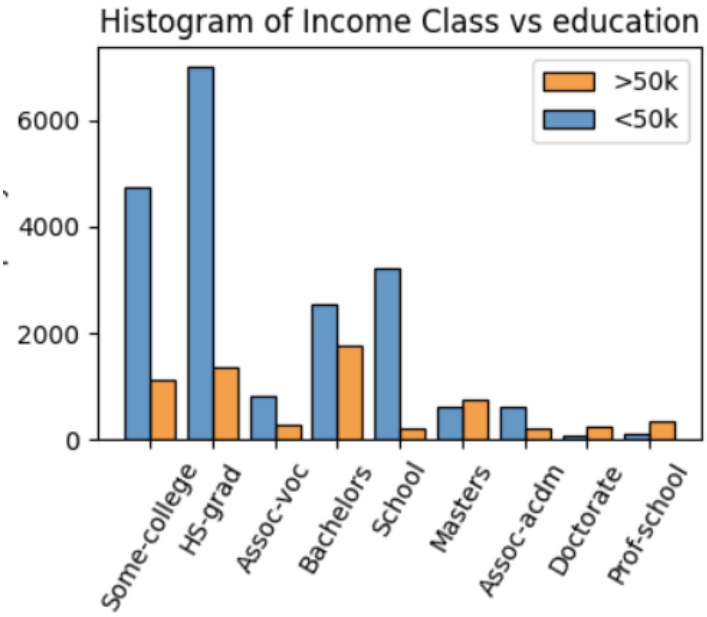
Andy Ho

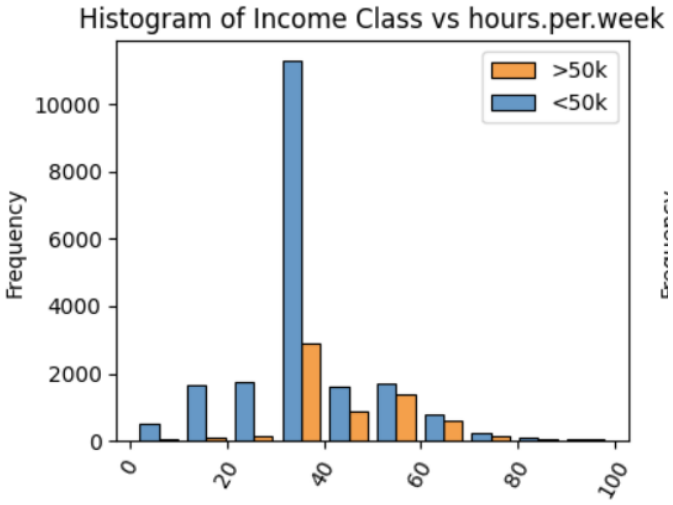
Project 1 Report:

Heatmap Analysis: For this heatmap, it's clear to see that there really aren't linear relationships between the features. The greatest correlation between the features is between age and income, but even then, it's quite small at roughly .12. However, once comparing them to the income category, we can see that the relationship between income and [age, capital gain, capital loss, and hours per week is apparent with relationships ranging from .15 (capital loss to income) to .31 (capital.gain to income).

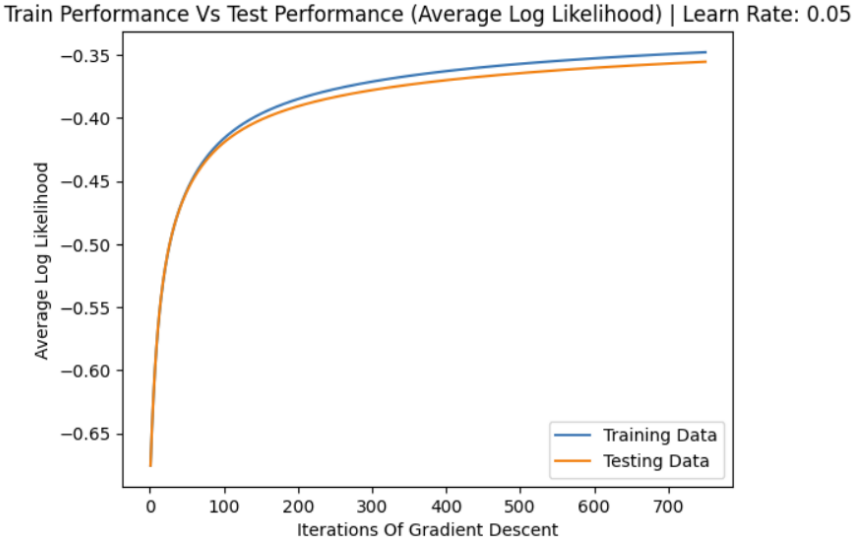


Histogram Analysis: In these histograms of income vs education, one of the most standout obvious things is that if you're working less than 40 hours a week, you're almost certainly not making more than 50 thousand dollars. However, the inverse isn't quite true, as if you work more than 40 hours you're not guaranteed to make more than 50 thousand dollars, but your chance of making >50k certainly increases. This is seen in how the proportion of people making >50k only increases as hours increase. As far as education goes, it seems pretty clear that the more education one receives, the more likely they'll find themselves making >50k.



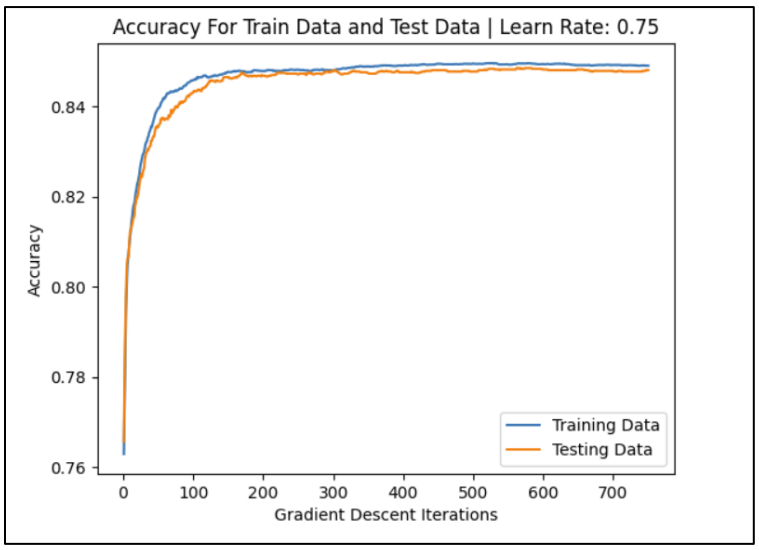


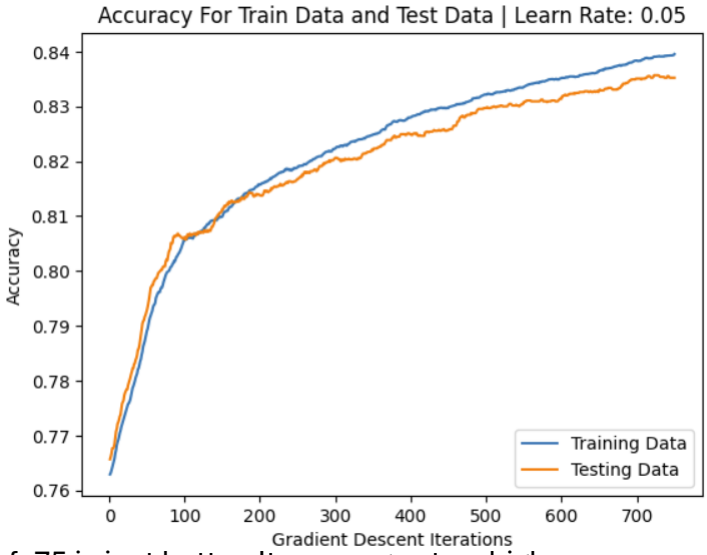
Average Log Likelihood Analysis and Accuracy: See below:



Average Log Likelihood: So in these two plots, it seems as if the learn rate of .75 converges much faster onto the optimal parameters for the weights than the learn rate of .75. This can be seen in how quickly the graph flattens out over time and how sharp the curve's "turn" is, wherein the learn rate of .75 makes a turn closer to 90 degrees. Something important to note is that the learn rate of .75 not only reaches optimal parameters faster, but it also reaches better parameters in general, with it being able to maximize the train set to around -.325, wherein after 700 iterations the learn rate of .05 maximizes the parameters to -.35.

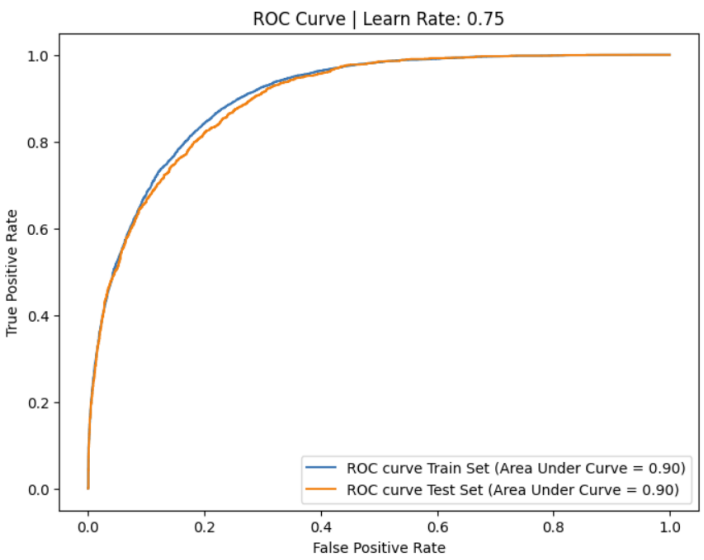
Accuracy:

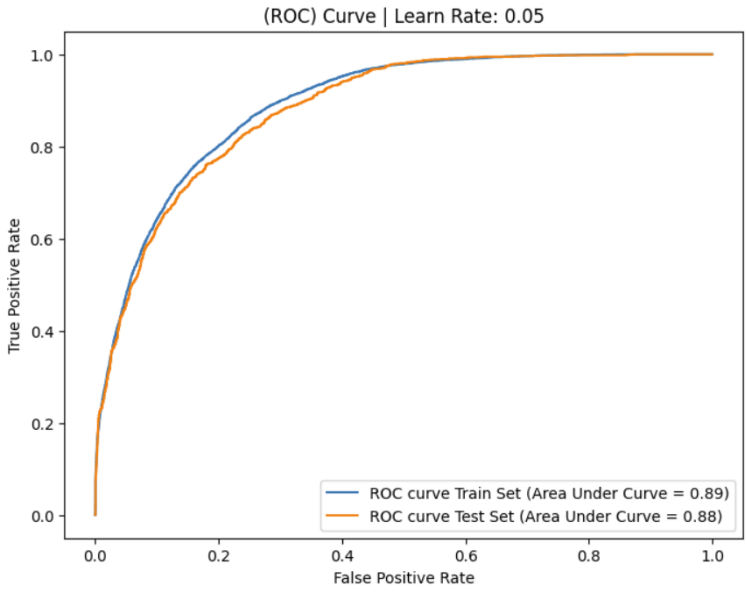




Similarly to the log Likelihood, the learn rate of .75 is just better. It converges to a higher accuracy, doing it much faster reaching relatively optimal parameters at around 400 iterations in comparison to .05's 700 iterations. Both of these learn rates have similar accuracy after full training, where in they both hover around 84%-85% accuracy.

ROC Plots:





Unlike the other plots the ROC curves don't really seem to be that different between each of the learning rates. Not only are the shapes of the graph roughly the same, but the area underneath each curve is also roughly the same between the two plots. I guess the ROC curve with the learn rate of .75 is ever so slightly better though because the AUC of test set is greater .9 > .88.