Programming Assignment #1 Report

Name: Carter Young

Due: 31 JAN 2024

1. *Correlation Heatmap*

Commentary: We see that all features have a positive correlation with each other, except for capital gain and capital loss (self-explanatory). We also note that the correlation between capital gain and income is markably higher than the other correlations, suggesting that capital gain has a material impact on income. Conversely,
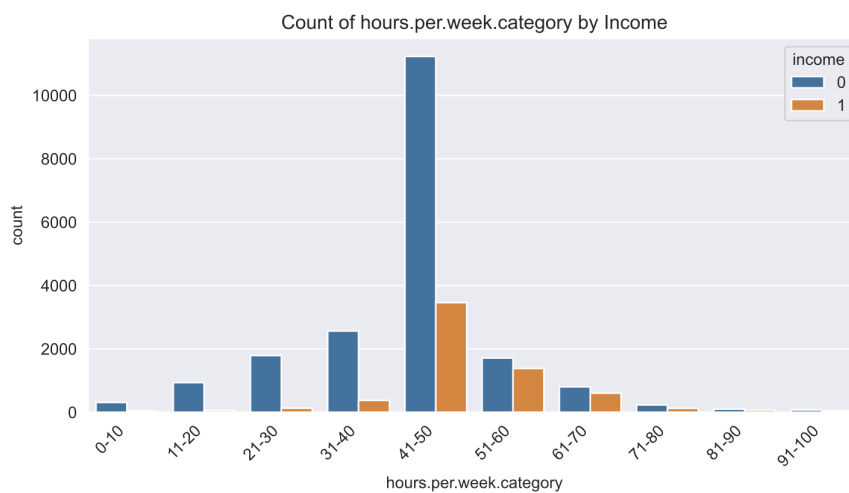
capital loss does not seem to impact income as much as might be expected. This suggests that even if a person takes losses on their investments or holdings, the mere presence of those holdings indicates sufficient or even considerable income.

2. Histograms

Education:
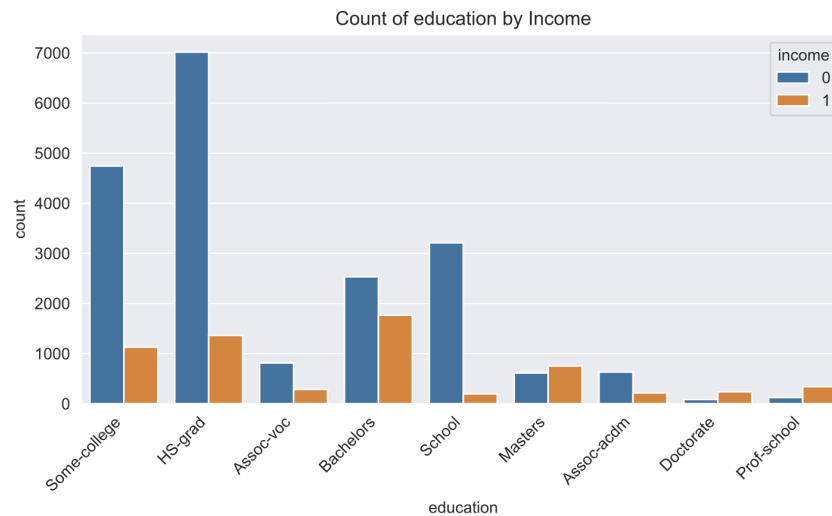
Hours per Week:

Commentary: The only education categories which include more wealthy persons on average are 'Masters', 'Doctorate', and 'Prof-school', which logically makes sense. Conversely, the least wealthy education group is that which did not finish high school. Interestingly, no group in "hours per week" makes more than $50K a majority of the time. However, those who work more than 50 hours per week do experience a marked increase in income, though this is by no means a rule.

3. Log Likelihood Performance
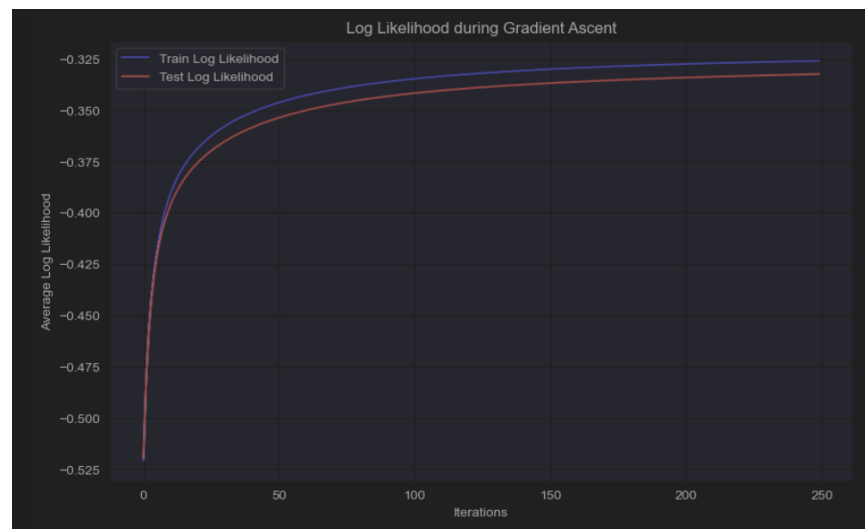
Commentary: Both the train and test log likelihoods increase sharply and then plateau, indicating the model is converging to a solution. The curves are also very close together, which indicates the model does not overfit to the data and generalize well. We note that there are significant diminishing returns past a certain # of iterations, and we may want to decrease the n_iters for future executions.

4. Plot_ROC

Receiver Operating Characteristic
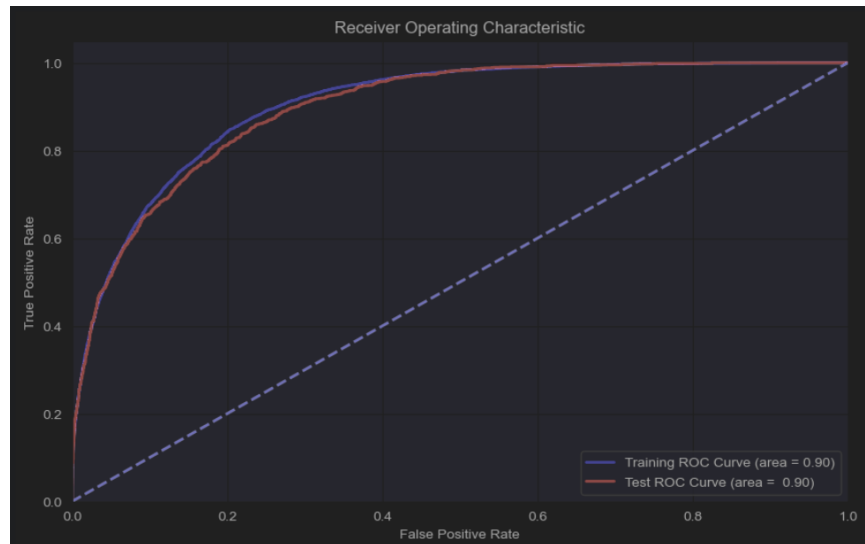
Commentary: In this graph, we see an AUC of 0.90 which suggests the model has a good measure of separability and is able to easily distinguish between classes. As in the prior graph, we see curves which increase sharply and plateau, and are also very tightly fitted. Lastly, both ROC curves are far above the diagonal line, signifying that the model performs significantly better than random chance.