COSC 325 HW1 Writeup Zebulon Mcknight

Task 2 Questions:

a. My primary design philosophy was to keep the code I was writing as simple as possible. My limited experience with machine learning was the primary motivator behind this, and I was willing to trade off performance by excluding more complicated steps of the pipeline if it contributed to my understanding. In effect this means I modeled my code after the code provided in class as closely as possible. One major design deviation I made was an update to the imputer implementation. My model uses two imputers (one numeric and one categorical) to avoid dropping samples and ensure compatibility with the Kaggle competition.
b. SKLearn validation accuracy = 0.7985074626865671
Scratch validation accuracy = 0.8134328358208955

   Both models performed fairly similarly on validation data, but neither model performed exceptionally well. Approximately 80% is good accuracy for a simple beginning model, but better data preparation steps and more in depth training (or a different model) could perform better.
c. My best performing model came from my scratch model with a learning rate of 0.0275 training over 1500 epochs. It resulted in a testing accuracy of 0.75358 which placed me at 8531 on the leaderboard.

Task 3 Questions:

1. Overall the course has been very satisfying up to this point. My one suggestion would be the incorporation of more code into the lectures. While the code provided in the zip files is extremely helpful, I think it may be easier to understand the content as it is taught if the instructor were to show the connection between the theory in the PowerPoints and the implementations in the provided code.
2. The most difficult, but rewarding, part of the homework was designing the scratch model. The limited instructions and theory heavy content made it difficult to understand at first, but also forced me to learn the material as I worked out how to implement the concepts from class.