1. Ag3743, hs2923, js4637, rap2194
2. As we know, neural networks are considered to be a black box model by many people. What we are looking to solve is to be able to understand what our neural network is doing at each layer. Thus, we want to interpret what feature map our neural network uses and how each layer reacts to the input.
3. We will be looking at two main datasets. First, we will look at a relatively simpler dataset of MNIST and look at how our neural network reacts and then we will look at a slightly more complex dataset of CIFAR 10 and see how feature visualization works in that case.
4. Our starting point is something that we worked on in class. The feature visualization for classification of E, F & L ; we would look at the weights of each layer and look to visualize them. This is feasible for smaller neural networks but not possible for larger neural networks. Then, our next step would be too look at the feature visualization as an optimization problem and see which image activates the network the most. There are multiple methods of looking at feature visualization: Using a deconvoluted neural network; regularized optimization in image space.
5. The most difficult aspect of the project will be to look at the optimization of the images and the complexity of the model which will make running times for the code to be quite high. We are looking at multiple papers which give us methodologies to look at this optimization and we plan on running the code in a GPU to ensure that it runs at a good speed.