Writeup for homework.
Satwik Singh (satwiks2)

Current method reaches an accuracy of 91.89%. I implemented a stochastic gradient descent algorithm, where I shuffle the dataset after every batch of 3000 iterations and do this 50 times, updating the theta in every iteration using the method provided in the lecture notes and class notes that I took. I took the hyperparameter alpha to be 0.003, I was initially going to implement a stepwise alpha but 0.003 got me the results so I didn’t play around. I initialized theta using np.random.randomn() that gives a nd array of normally distributed values. After I get the final theta I use it to get the argmax() and predict labels, using the sample code that was provided to us. I call my function matrix_mult() which multiplies the theta and x_vec to generate a vector of probabilities out of which the argmax is predicted.