homework 4

Training the NN

1. What did you try to get the optimal performance?

Following is a detailed table for the parameters that i have tried and their corresponding test accuracy:

	Batch Size	HiddeSizes	Eta	Beta1	Beta2	Epsilon	Num epochs	Test accuracy
1	128	[16, 64, 16]	0.001	0.9	0.999	1*10^-8	10	0.96
2	128	[16, 64, 16]	0.001	0.9	0.999	1*10^-8	10	0.95
3	128	[16, 64, 16]	0.01	0.9	0.999	1*10^-8	10	0.9525
4	128	[16, 64, 16]	0.005	0.5	0.999	1*10^-8	10	0.957
5	128	[16, 64, 16]	0.001	0.5	0.999	1*10^-8	20	0.958
6	64	[16, 64, 16]	0.001	0.5	0.999	1*10^-8	10	0.957
7	128	[16, 16, 16]	0.001	0.5	0.999	1*10^-8	10	0.95
8	128	[16, 16, 16]	0.001	0.5	0.999	1*10^-8	10	0.95
9	128	[16, 16, 16]	0.001	0.5	0.999	1*10^-8	10	0.974
10	128	[64, 64, 64, 16]	0.001	0.5	0.999	1*10^-8	20	0.975
11	128	[64, 64, 64, 16]	0.001	0.5	0.999	1*10^-8	10	0.973
12	128	[64, 64, 64, 16]	0.001	0.5	0.999	1*10^-8	10	0.974
13	128	[128, 64, 64, 16]	0.001	0.5	0.999	1*10^-8	10	0.9775
14	128	[128, 64, 64, 16]	0.001	0.5	0.999	1*10^-8	10	0.9788

2. What are the best hyper parameters you found?

Batch Size	HiddeSizes	Eta	Beta1	Beta2	Epsilon	Num epochs	Test accuracy
128	[128, 64, 64, 16]	0.001	0.5	0.999	1*10^-8	10	0.9788

2. Plot the best parameters?