1. (a)

Max Iterations	E_{in}	Binary E_{in}	Binary E_{out}	Time (s)
10^4	0.586	0.309	0.317	0.24
10^5	0.469	0.224	0.207	2.18
10^6	0.436	0.151	0.131	21.48

The logistic regression model generalizes well. The E_{in} and E_{out} are consistently very similar indicating that there is enough data for learning to occur, at least with this simple model. Increasing the maximum number of iterations, at least to up to 10^6, continues to improve the model's performance.

(b)

Learning Rate	E_{in}	Binary E_{in}	Binary E_{out}	Iterations	Time (s)
0.01	0.69	0.171	0.11	22429	0.49
0.1	0.663	0.171	0.11	2239	0.05
1	0.487	0.171	0.11	220	0.01
4	0.522	0.171	0.11	51	0.003
7	0.767	0.171	0.11	44	0.00299
7.5	0.813	0.171	0.11	180	0.00601
7.6	0.822	0.171	0.11	446	0.012
7.7	0.408	0.164	0.117	1000000	22.03

I normalized both the training and testing set to the mean and variance of the training set. This is because we never want to include the testing data in training.

Normalizing the data increases the performance of the model. Using the same parameters without normalized data yields E_{in} and E_{out} (binary) approaching random guessing.

Changing the learning rate usually arrives at the same error but just at different rates. The smallest learning rate takes > 20000 iterations to arrive at the same prediction model as does a learning rate of 7 (which takes just 44). Too big of a learning rate becomes cumbersome very quickly however. Increasing the learning rate to 7.6 makes us take 446 iterations and 7.7 makes us max out the number of iterations at 10^6!

3 (c) plot.