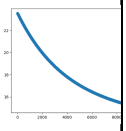
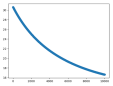
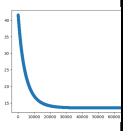
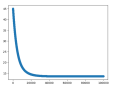
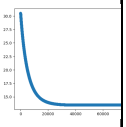
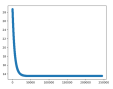
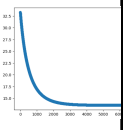
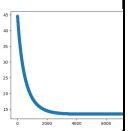
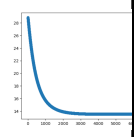
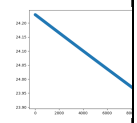
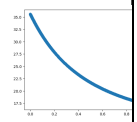
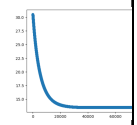
Miner- deeppp15

Assignment 2

Observations:

Learning Rate	No of Iterations	Classification Error	Training Time	Cross Entropy Error	Converged	Cost Variance across Iterations	Self Gradient Descent Accuracy	SciKit Learn Accuracy	Tolerance
0.001	10000	0.17105263157894735	0.8831398487091064	16.78728630750741	NA		0.7960526315789473	0.8157894736842105	10^{-3}.
0.001	10000	0.1842105263157895	89.41564488410 95	17.20657415363371	NA		0.8092105263157895	0.8157894736842105	10^{-6}.
0.001	100000	0.17105263157894735	6.782255649566 65	13.530576949678357	77105		0.8289473684210527	0.8157894736842105	10^{-3}.
0.001	100000	0.17105263157894735	9.005282640457153	13.536374264 73552	NA		0.8289473684210527	0.8157894736842105	10^{-6}.
0.001	1000000	0.17105263157894735	6.624827384948 7305	13.538984077203132	75878		0.8092105263157895	0.91	10^{-3}.
0.001	1000000	0.17105263157894735	21.77542567253 1128	13.541037707765641	244724		0.8289473684210527	0.91	10^{-6}.
0.01	1000000	0.17105263157894735	0.7015998363494873	13.534510215085046	7622		0.8289473684210527	0.91	10^{-3}.
0.01	100000	0.17105263157894735	0.7452270984649658	13.531995521419306	8112		0.8289473684210527	0.8157894736842105	10^{-3}.

Learning Rate	No of Iterations	Classification Error	Training Time	Cross Entropy Error	Converged	Cost Variance across Iterations	Self Gradient Descent Accuracy	SciKit Learn Accuracy	Tolerance
0.01	10000	0.17105263157894735	0.6727602481842041	13.52674732575593	7492		0.8289473684210527	0.8157894736842105	10^{-3}.
0.00001	10000	0.2828947368421053	0.9259915351867676	33.42486070638004	NA		0.7171052631578947	0.8157894736842105	10^{-3}.
0.00001	100000	0.3092105263157895	9.272836923599243	31.185127256497697	NA		0.7631578947368421	0.8157894736842105	10^{-3}.
0.00001	1000000	0.1907894736842105	87.38178491592407	16.524903545937647	NA		0.7894736842105263	0.91	10^{-3}.

In this above table as we can see we have tested on different parameters of the logistic regression model where we try different tolerance rates and learning rates.

Generalization Properties

In our training data the model sometimes overfits with a higher number of iterations with a higher accuracy, but our test_classification_error is high. Also to note that the accuracy in the Sci-kit logistic regression model is high compared to the self implemented gradient descent and sigmoid function.

Tolerance Rates

Tolerance in algorithms is used to determine the level of accuracy we need to achieve.
A less tolerance level can lead to faster convergence whereas a higher tolerance can lead to slower convergence.

In our example of 1000000 iterations and learning rate, our gradient converges quickly at 75878, whereas with a tolerance of 10^{-6} our gradient converges at 244724. Which is much higher than our previous convergence.

Learning Rates

Learning rate enables the algorithm to take smaller steps towards the minimum and prevent overshooting, it can result in slower convergence but improved accuracy. On the other hand, a higher learning rate may cause the algorithm to overshoot, oscillate around the minimum, or even diverge, which would result in a faster convergence but lesser accuracy.
In our 0.001 learning rate and 1000000 we get convergence at 75878 but accuracy at 0.7631578947368421 with low cross entropy error , whereas in our 0.00001 learning rate we do not get a convergence, and an accuracy of 0.80 with a comparatively high cross entropy error. Hence, larger rates can cause slower convergence or a high divergence which is our case.