

CS 6375

ASSIGNMENT _____1_____

Names of students in your group:
Liangyu Nie

Number of free late days used:
_____0_____

Note: You are allowed a total of 4 free late days for the entire semester. You can use at most 2 for each assignment. After that, there will be a penalty of 10% for each late day.

Please list clearly all the sources/references that you have used in this assignment.

During the assignment 1, I use SGD as modelling, , it has two important parameter: Learning rate and epoch, also for the evaluation, I evaluated MSE ,precision,recall and F-1 score. I set the learning rate at 0.01, 0.1and 0.5. Run the epoch 500,1000,1500.

Different learning rate and different epoch have different result:

eg:

Trial with learning_rate=0.01, epochs=1000, Test MSE=0.2647058823529412

precision recall f1-score support				
0	0.66	0.76	0.70	113
1	0.81	0.72	0.76	159
accuracy			0.74	272
macro avg	0.73	0.74	0.73	272
weighted avg	0.75	0.74	0.74	272

Trial with learning_rate=0.01, epochs=2000, MSE=0.011029411764705883

precision recall f1-score support				
False	0.98	1.00	0.99	138
True	1.00	0.98	0.99	134

accuracy		0.99	272
macro avg	0.99	0.99	0.99 272
weighted avg	0.99	0.99	0.99 272

Also the best test dataset error values for the best set of paper is learning_rate=0.5, epochs=1500, MSE=0.0018399264029438822 the performance is:

	precision	recall	f1-score	support
False	1.00	1.00	1.00	567
True	1.00	1.00	1.00	520

accuracy		1.00	1087
macro avg	1.00	1.00	1.00 1087
weighted avg	1.00	1.00	1.00 1087

I am satisfy with this performance, because with this setting the the predict is 100% correct.

For the part2

the best SGD Classifier parameters: penalty=None, max_iter=2000, tol=1e-05

Training accuracy: 0.7157313707451702

Testing accuracy: 0.7573529411764706

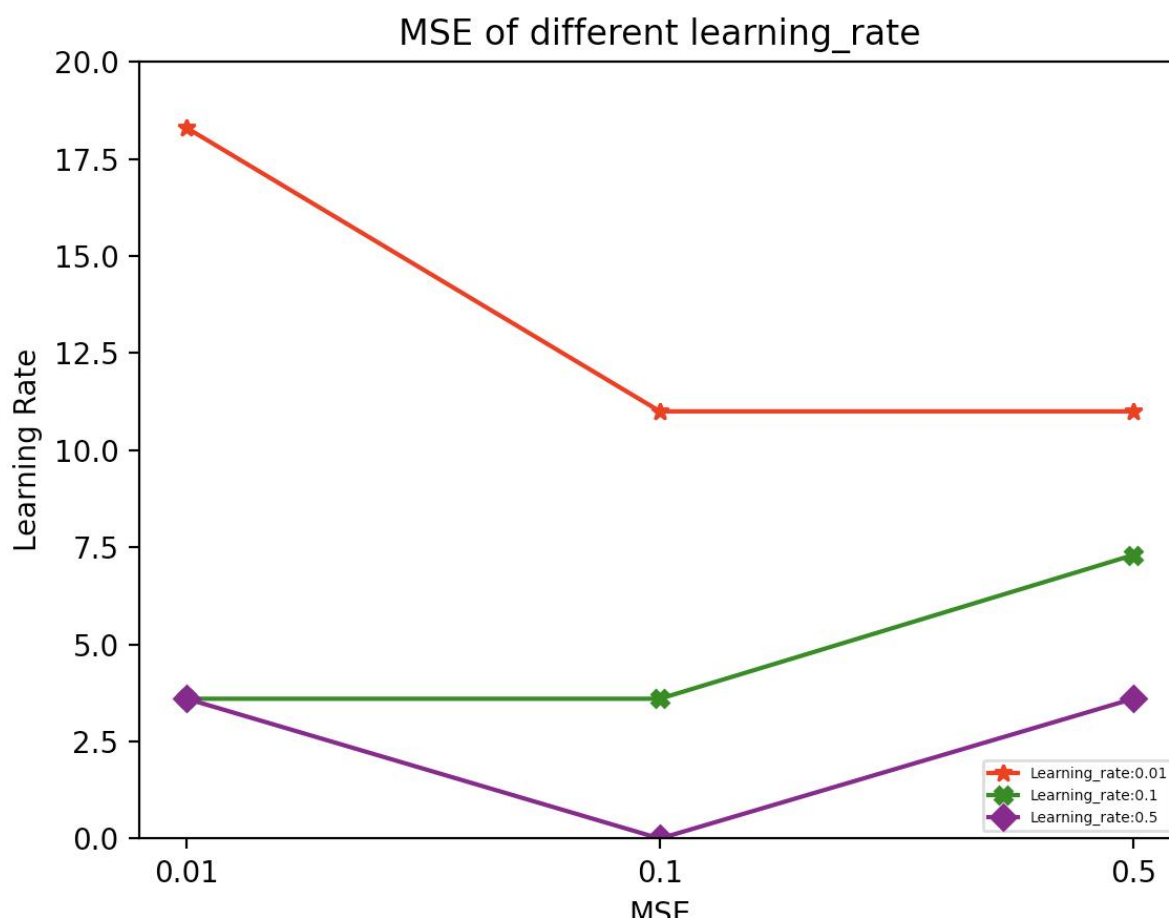
MSE: 0.2426470588235294

precision	recall	f1-score	support
-----------	--------	----------	---------

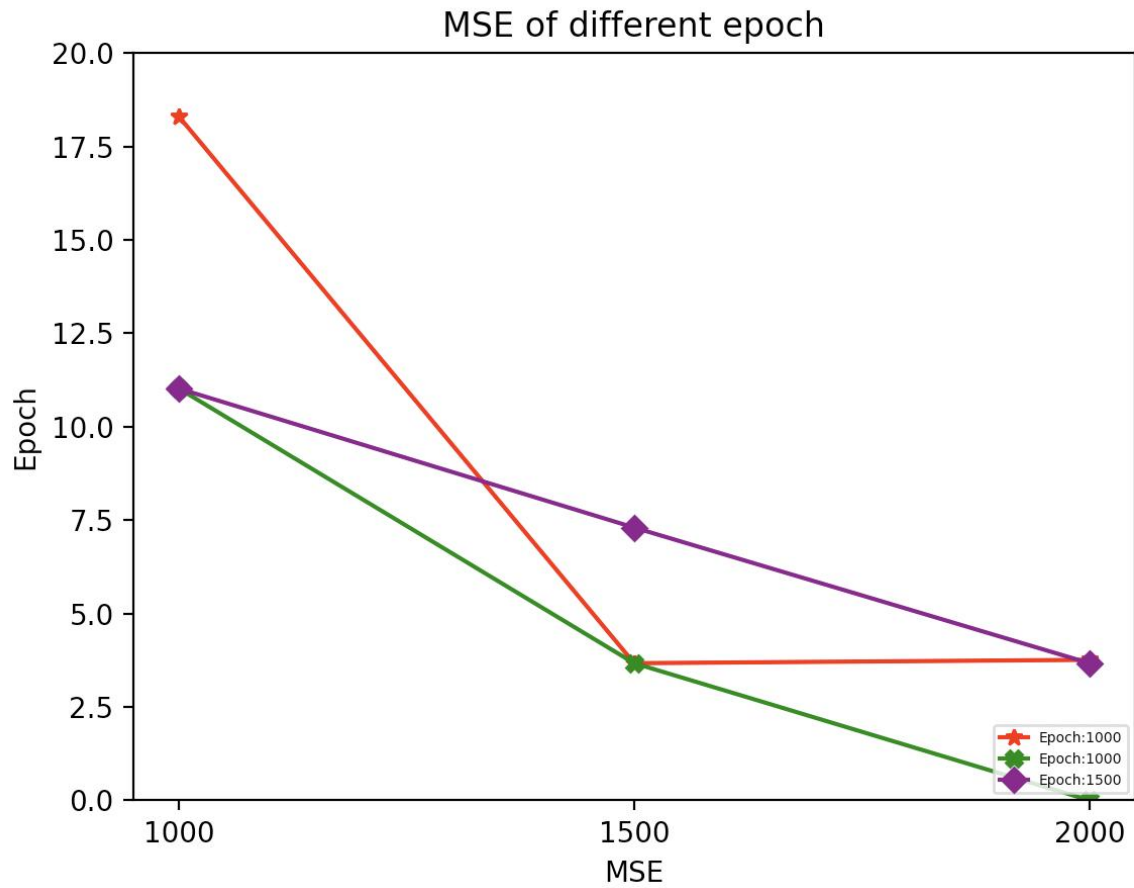
0	0.71	0.71	0.71	113
1	0.79	0.79	0.79	159

accuracy			0.76	272
macro avg	0.75	0.75	0.75	272
weighted avg	0.76	0.76	0.76	272

With this setting we got 76% accuracy on the dataset



Here is the plot of MSE of different learning_rate



This plot shows the MSE of different epoch