

ML_HW#1

106033212 曾靖桐

1.

a.

Random_state is used to set a seed for the random generator, and it can ensure the results are reproducible.

b.

Random_state	Accuracy
80	0.972
50	0.972
42	1
30	1
20	1
10	0.889
5	0.917
1	1

Since random_state is only the parameter to make sure the data sets are the same in every test, it has no relation between accuracy.

2.

a.

Testset_portion(random_state=42)	Accuracy
0.2	1
0.5	0.989
0.8	0.944
0.9	0.801
0.98	0.394

As the value of testset_portion keep increasing the accuracy drops dramatically.

b.

It is because that if the testset_portion is too large, there are not enough data for training, and the model would be underfitting. Thus, the model can't predict the test data correctly.