

Saranpat Prasertthum (655667271)

IE517 ML in Fin Lab

Module 6 Homework (Cross validation)

Use the Treasury Yield Curve dataset

Out[1]:

Click here to toggle on/off the raw code.

Load Data

Out[3]:

	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	AGE	PAY_0	PAY_2	PAY_3	PAY_4	PAY_5
ID										
1	20000	2	2	1	24	2	2	-1	-1	-2
2	120000	2	2	2	26	-1	2	0	0	0
3	90000	2	2	2	34	0	0	0	0	0
4	50000	2	2	1	37	0	0	0	0	0
5	50000	1	2	1	57	-1	0	-1	0	0

5 rows × 24 columns

Part 1: Random test train splits

Out[6]:

Random State	1	2	3	4	5	6	7	8	
Train Score	0.824704	0.824333	0.823037	0.825333	0.823593	0.825296	0.824444	0.823444	0.
Test Score	0.820000	0.824333	0.835000	0.809667	0.830000	0.810333	0.817667	0.828000	0.

Avg Train Score: 0.82
Std Train Score: 0.0008
Avg Test Score: 0.82
Std Test Score: 0.0086

Part 2: Cross validation

Out[8]:

K	1	2	3	4	5	6	7	8	9	
Test Score	0.817667	0.821667	0.812	0.822333	0.817333	0.817	0.823333	0.829	0.821333	0.81

Avg Test Score: 0.82
Std Test Score: 0.0047

Part 3: Conclusions

The results of both non-cross validation and cross validation are similar, but using cross validation provides a more reliable estimate of model performance. Non-cross validation only trains and tests the model on a single train and test set, which can be faster and more efficient. However, cross validation trains the model 10 times and generates 10 scores, which helps to reduce any bias in the train and test sets. Although cross validation can be more computationally intensive, the scores are more reliable and provide a better representation of the model's ability to generalize to new data. Therefore, it is recommended to use cross validation when evaluating the performance of a model.

Part 4: Appendix

My name is Saranpat Prasertthum
My NetID is: 655667271
I hereby certify that I have read the University policy on Academic Integrity and that I am not in violation.