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]:

ID	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	AGE	PAY_0	PAY_2	PAY_3	PAY_4	PAY_5	.
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

Out[4]:

	int
LIMIT_BAL	30000
SEX	30000
EDUCATION	30000
MARRIAGE	30000
AGE	30000
PAY_0	30000
PAY_2	30000
PAY_3	30000
PAY_4	30000
PAY_5	30000
PAY_6	30000
BILL_AMT1	30000
BILL_AMT2	30000
BILL_AMT3	30000
BILL_AMT4	30000
BILL_AMT5	30000
BILL_AMT6	30000
PAY_AMT1	30000
PAY_AMT2	30000
PAY_AMT3	30000
PAY_AMT4	30000
PAY_AMT5	30000
PAY_AMT6	30000
DEFAULT	30000

Part 1: Random test train splits

```
Test score: 0.82
```

Part 2: Cross validation

Out[7]:

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	0.814

```
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 Integr
ity and that I am not in violation.
```