Iterative Imputer/MICE
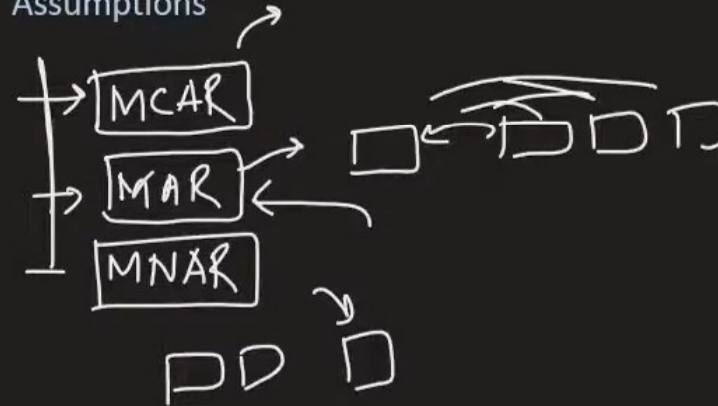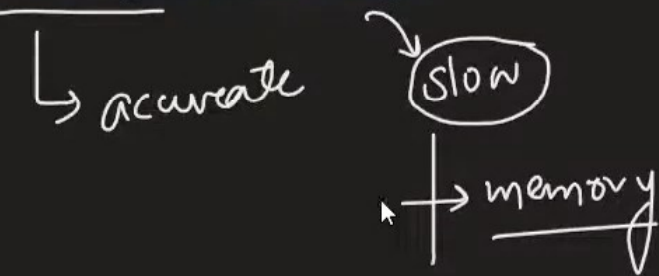
Wednesday, April 28, 2021   11:40 AM

MICE stands for Multivariate Imputation by Chained Equations

Assumptions

MCAR
MAR
MNAR

Advantages & Disadvantage

accurate   slow

memory

1. Actual Dataset

	R&D Spend	Administration	Marketing Spend	Profit
21	8.0	15.0	30.0	11.0
37	4.0	5.0	20.0	9.0
2	15.0	10.0	41.0	19.0
14	12.0	16.0	26.0	13.0
44	2.0	15.0	3.0	7.0

2. Removing the target column

	R&D Spend	Administration	Marketing Spend
21	8.0	15.0	30.0
37	4.0	5.0	20.0
2	15.0	10.0	41.0
14	12.0	16.0	26.0
44	2.0	15.0	3.0

2. Removing the target column

	R&D Spend	Administration	Marketing Spend
21	8.0	15.0	30.0
37	4.0	5.0	20.0
2	15.0	10.0	41.0
14	12.0	16.0	26.0
44	2.0	15.0	3.0

3. Introduced some fake nan values

	R&D Spend	Administration	Marketing Spend
21	8.0	15.0	30.0
37	NaN	5.0	20.0
2	15.0	10.0	41.0
14	12.0	NaN	26.0

21	8.0	15.0	30.0
37	4.0	5.0	20.0
2	15.0	10.0	41.0
14	12.0	16.0	26.0
44	2.0	15.0	3.0

3. Introduced some fake nan values

	R&D Spend	Administration	Marketing Spend
21	8.0	15.0	30.0
37	NaN	5.0	20.0
2	15.0	10.0	41.0
14	12.0	NaN	26.0
44	2.0	15.0	NaN

Step 1 - Fill all the NaN values with mean of

37	NaN	5.0	20.0
2	15.0	10.0	41.0
14	12.0	NaN	26.0
44	2.0	15.0	NaN

Step 1 - Fill all the NaN values with mean of respective cols

	R&D Spend	Administration	Marketing Spend
21	8.00	15.00	30.00
37	9.25	5.00	20.00
2	15.00	10.00	41.00
14	12.00	11.25	26.00
44	2.00	15.00	29.25

Step 2 - Remove all col1 missing values

Step 3 - Predict the missing values of col1 using

| 44 | 2.00 | 15.00 | 29.25 |

Step 2 - Remove all col1 missing values

	R&D Spend	Administration	Marketing Spend
21	8.0	15.00	30.00
37	NaN	5.00	20.00
2	15.0	10.00	41.00
14	12.0	11.25	26.00
44	2.0	15.00	29.25

	Administration	Marketing Spend
21	15.00	30.00
2	10.00	41.00
14	11.25	26.00
44	15.00	29.25

Step 3 - Predict the missing values of col1 using other cols

	R&D Spend	Administration	Marketing Spend
21	8.00	15.00	30.00
37	23.14	5.00	20.00
2	15.00	10.00	41.00
14	12.00	11.25	26.00
44	2.00	15.00	29.25

Step 4 - Remove all col2 missing values

	R&D Spend	Administration	Marketing Spend
21	8.00	15.0	30.00
37	23.14	5.0	20.00

	R&D Spend	Marketing Spend
21	8.00	30.00

Step 5 - Predict the missing values of col2 using other cols

	R&D Spend	Administration	Marketing Spend
21	8.00	15.00	30.00

	R&D Spend	Administration	Marketing Spend
21	8.0	15.00	30.00
37	NaN	5.00	20.00
2	15.0	10.00	41.00
14	12.0	11.25	26.00
44	2.0	15.00	29.25

	Administration	Marketing Spend
21	15.00	30.00
2	10.00	41.00
14	11.25	26.00
44	15.00	29.25

	R&D Spend	Administration	Marketing Spend
21	8.00		
37	23.14	5.00	20.00
2	15.00	10.00	41.00
14	12.00	11.25	26.00
44	2.00	15.00	29.25

Step 4 - Remove all col2 missing values

	R&D Spend	Administration	Marketing Spend
21	8.00	15.0	30.00
37	23.14	5.0	20.00
2	15.00	10.0	41.00
14	12.00	NaN	26.00
44	2.00	15.0	29.25

	R&D Spend	Marketing Spend
21	8.00	30.00
37	23.14	20.00
2	15.00	41.00
44	2.00	29.25

Step 5 - Predict the missing values of col2 using other cols

	R&D Spend	Administration	Marketing Spend
21	8.00	15.00	30.00
37	23.14	5.00	20.00
2	15.00	10.00	41.00
14	12.00	11.06	26.00
44	2.00	15.00	29.25

Step 6 - Remove all col3 values

2	15.00	10.00		41.0
14	12.00	11.06		26.0
44	2.00	15.00		NaN

2	15.00	10.00
14	12.00	11.06
44	2.00	15.00

2	15.00		
14	12.00		
44	2.00	15.00	31.56

$X \to U2$

Iteration 0 —

	R&D Spend	Administration	Marketing Spend
21	8.00	15.00	30.00
37	9.25	5.00	20.00
2	15.00	10.00	41.00
14	12.00	11.25	26.00
44	2.00	15.00	29.25

Iteration 1

	R&D Spend	Administration	Marketing Spend
21	8.00	15.00	30.00
37	23.	5.00	20.00
2	15.00	10.00	41.00
14	12.00	11.06	26.00
44	2.00	15.00	31.56

Difference

	R&D Spend	Administration	Marketing Spend
21	0.00	0.00	0.00
37	13.89	0.00	0.00
2	0.00	0.00	0.00
14	0.00	-0.19	0.00
44	0.00	0.00	2.31

Iteration 1

	R&D Spend	Administration	Marketing Spend
21	8.00	15.00	30.00
37	23.14	5.00	20.00

Iteration 2

	R&D Spend	Administration	Marketing Spend
21	8.00	15.00	30.00
37	23.78	5.00	20.00

Difference

	R&D Spend	Administration	Marketing Spend
21	0.00	0.00	0.0
37	0.64	0.00	0.0

Iteration 0

	R&D Spend	Administration	Marketing Spend
21	8.00	15.00	30.00
37	9.25	5.00	20.00
2	15.00	10.00	41.00
14	12.00	11.25	26.00
44	2.00	15.00	29.25

Iteration 1

	R&D Spend	Administration	Marketing Spend
21	8.00	15.00	30.00
37		5.00	20.00
2	15.00	10.00	41.00
14	12.00	11.06	26.00
44	2.00	15.00	31.56

Difference

	R&D S		
21	0.00	0.00	0.00
37	13.89	0.00	0.00
2	0.00	0.00	0.00
14	0.00	-0.19	0.00
44	0.00	0.00	2.31

Iteration 1

	R&D Spend	Administration	Marketing Spend
21	8.00	15.00	30.00
37	23.14	5.00	20.00
2	15.00	10.00	41.00
14	12.00	11.06	26.00
44	2.00	15.00	31.56

Iteration 2

	R&D Spend	Administration	Marketing Spend
21	8.00	15.00	30.00
37	23.78	5.00	20.00
2	15.00	10.00	41.00
14	12.00	11.22	26.00
44	2.00	15.00	31.56

Difference

	R&D Spend	Administration	Marketing Spend
21	0.00	0.00	0.0
37	0.64	0.00	0.0
2	0.00	0.00	0.0
14	0.00	0.16	0.0
44	0.00	0.00	0.0

Iteration 2

R&D Spend	Administration	Marketing Spend

Iteration 3

R&D Spend	Administration	Marketing Spend

Difference

R&D Spend	Administration	Marketing Spend