

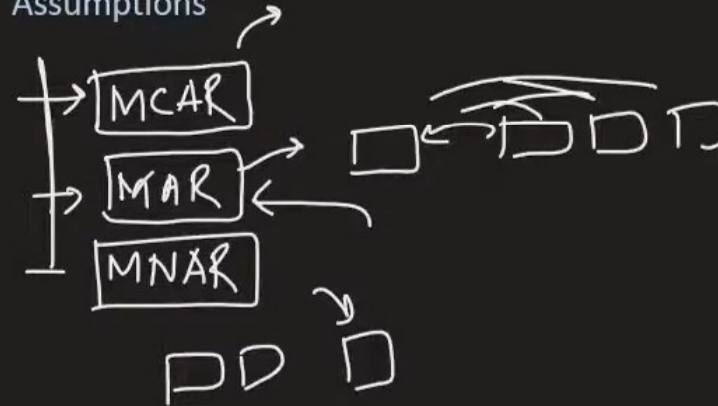


Iterative Imputer/MICE

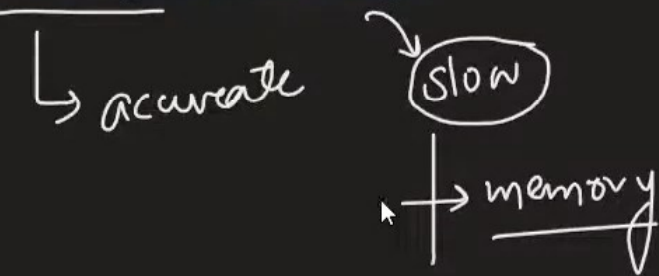
Wednesday, April 28, 2021 11:40 AM

MICE stands for **Multivariate Imputation by Chained Equations**

Assumptions



Advantages & Disadvantage



- 100 Days of ML
- Day 22 - Pandas Pr...
- Day 23 - Feature E...
- Day 24 - Standardi...
- Day 25 - Normaliza...
- Day 26 - Ordinal E...
- Day 27 - One Hot E...
- Day 28 - ColumnTr...
- Day 29 - Pipelines
- Day 30 - Function T...
- Day 31 - Power Tra...
- Day 32 - Discrtizati...
- Day 33 - Working-...
- Day 34 - Working...
- Day 35 - Complete...
- Day 36 - Handling...
- Day 37 - Handling...
- Day38-Missing Indi...
- Day39 - KNN Impu...
- Day40 - Iterative I...

+ Add section

+ Add page



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

OneNote for Windows 10

37	<u>NaN</u>	5.0	20.0
2	15.0	10.0	41.0
14	12.0	<u>NaN</u>	26.0
44	2.0	15.0	<u>NaN</u>



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	<u>9.25</u>	5.00	20.00
2	15.00	10.00	41.00
14	12.00	<u>11.25</u>	26.00
44	2.00	15.00	<u>29.25</u>

Step 2 - Remove all col1 missing values

Step 3 - Predict the missing values of col1 using

OneNote for Windows 10

Home Insert Draw View Help

44 2.00 15.00 29.25



Step 2 - Remove all col1 missing values

	<u>R&D Spend</u>	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

	<u>Administration</u>	<u>Marketing Spend</u>
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

	<u>R&D Spend</u>	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

	<u>R&D Spend</u>	Administration	Marketing Spend
21	8.00	15.0	30.00
37	23.14	5.0	20.00

	<u>R&D Spend</u>	<u>Marketing Spend</u>
21	8.00	30.00

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

	<u>R&D Spend</u>	Administration	Marketing Spend
21	8.00	15.00	30.00



	<u>R&D Spend</u>	<u>Administration</u>	<u>Marketing Spend</u>
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

	<u>Administration</u>	<u>Marketing Spend</u>
21	15.00	30.00
2	10.00	41.00
14	11.25	26.00
44	15.00	29.25

	<u>R&D Spend</u>	<u>Administration</u>	<u>Marketing Spend</u>
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

	<u>R&D Spend</u>	<u>Administration</u>	<u>Marketing Spend</u>
21	8.00	15.00	30.00
37	23.14	5.00	20.00
2	15.00	10.00	41.00
14	12.00	NaN	26.00
44	2.00	15.00	29.25

	<u>R&D Spend</u>	<u>Marketing Spend</u>
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

	<u>R&D Spend</u>	<u>Administration</u>	<u>Marketing Spend</u>
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

← → OneNote for Windows 10

Home Insert Draw View Help

↶ ↷ ↻ ↺ ↻ ↺ ↻ ↺ ↻ ↺ ↻ ↺ ↻ ↺

44 2.00 15.0 29.25 44 2.00

Shapes Ink to Shape Ink to Text



Step 6 - Remove all col3 values

	R&D Spend	Administration	Marketing Spend
21	8.00	15.00	30.0
37	23.14	5.00	20.0
2	15.00	10.00	41.0
14	12.00	11.06	26.0
44	2.00	15.00	NaN

	R&D Spend	Administration
21	8.00	15.00
37	23.14	5.00
2	15.00	10.00
14	12.00	11.06

	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 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

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

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



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

2	15.00
14	12.00
44	2.00
	15.00
	31.56

Iteration 0

Iteration 1

Difference

	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

	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

	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

Iteration 2

Difference

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

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

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



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	8.00	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
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
--	-----------	----------------	-----------------