Principal Component Analysis

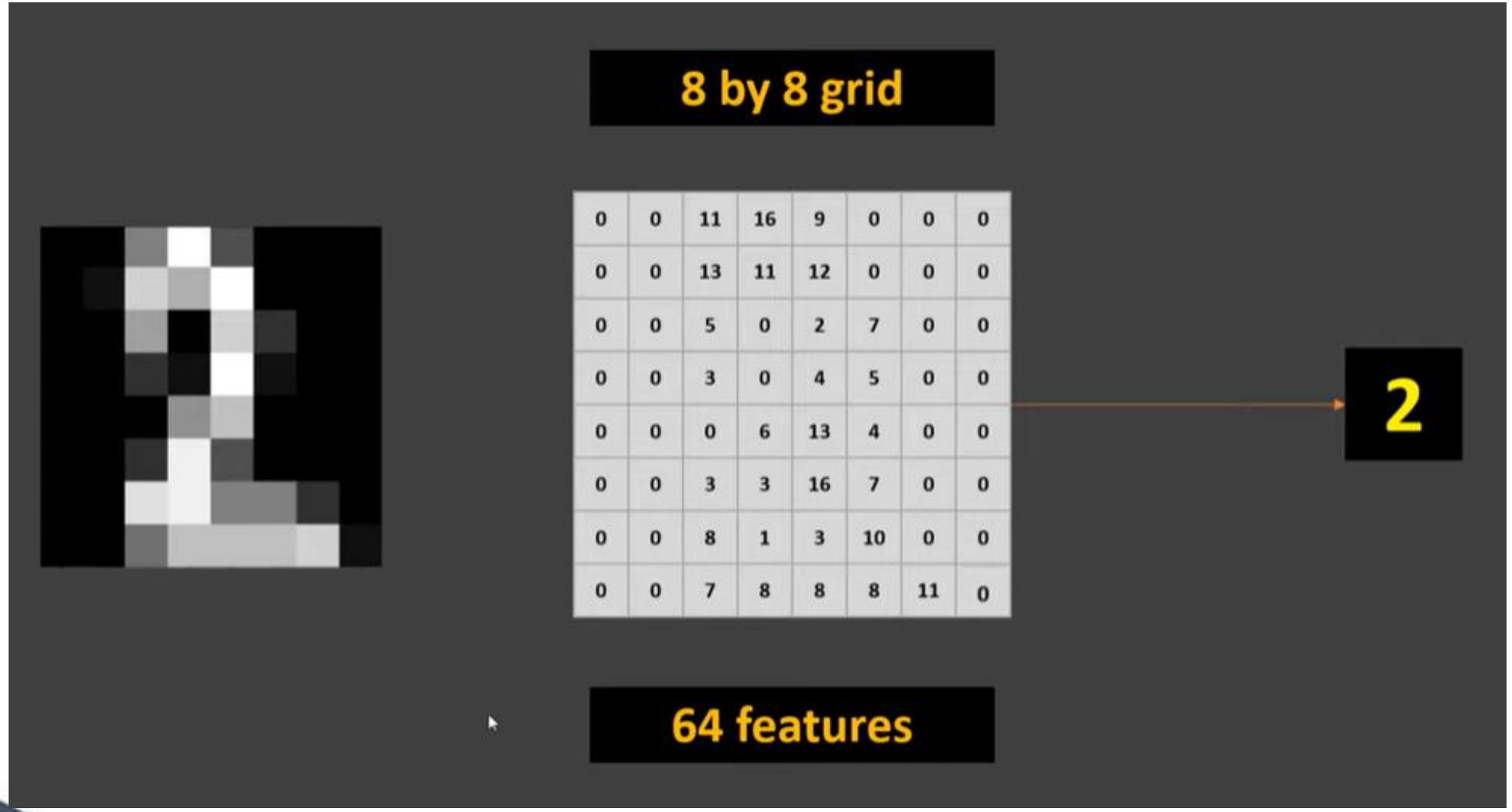
Aksadur Rahman

aksadur@yahoo.com

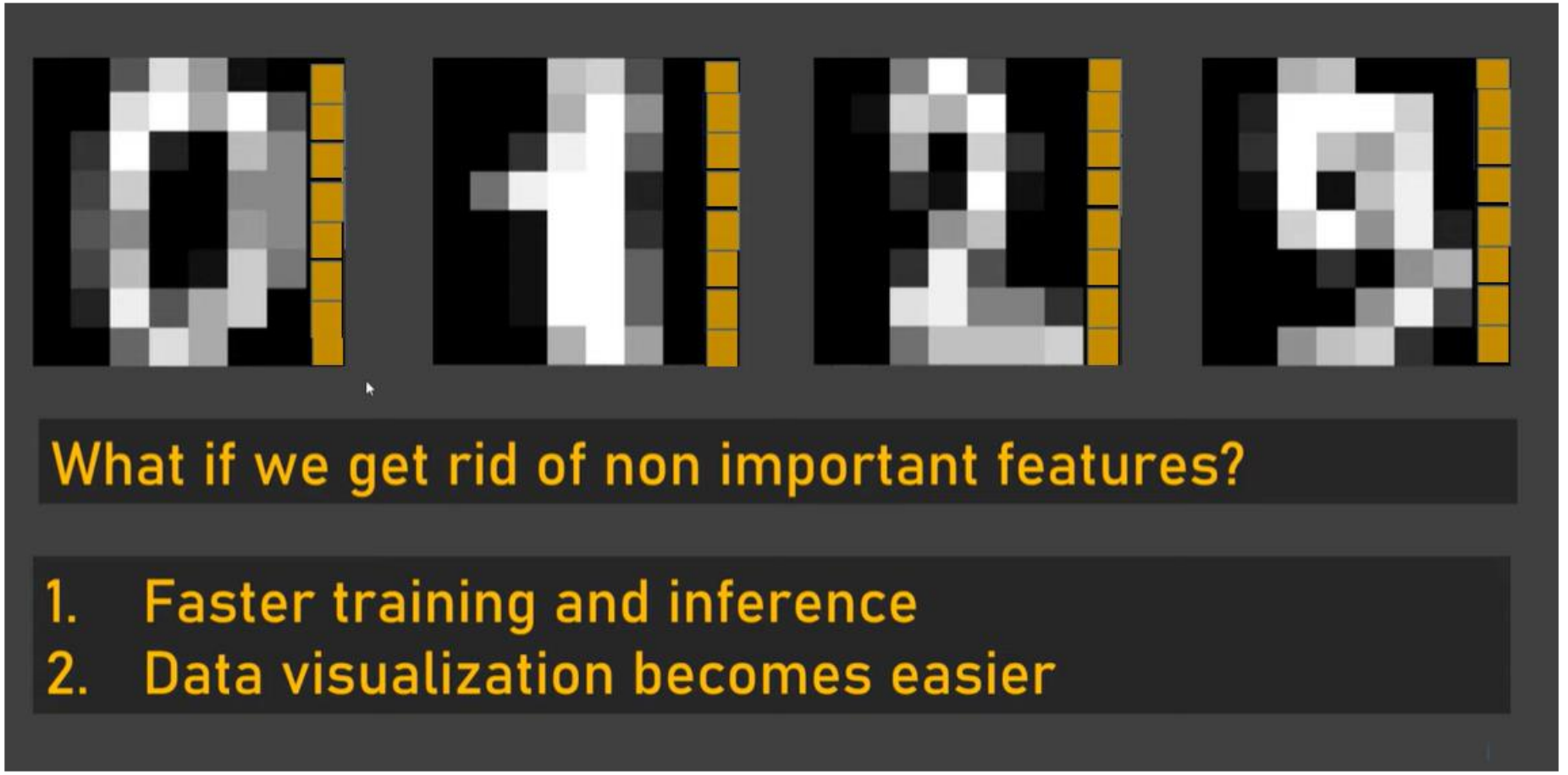
DATA SCIENCE

town	area	bathroom	plot	trees nearby	price
monroe	2600	2	8500	2	550000
monroe	3000	3	9200	2	565000
monroe	3200	3	8750	2	610000
monroe	3600	4	10200	2	680000
monroe	4000	4	15000	2	725000
west windsor	2600	2	7000	2	585000
west windsor	2800	3	9000	2	615000
west windsor	3300	4	10000	1	650000
west windsor	3600	4	10500	1	710000

DATA SCIENCE



DATA SCIENCE

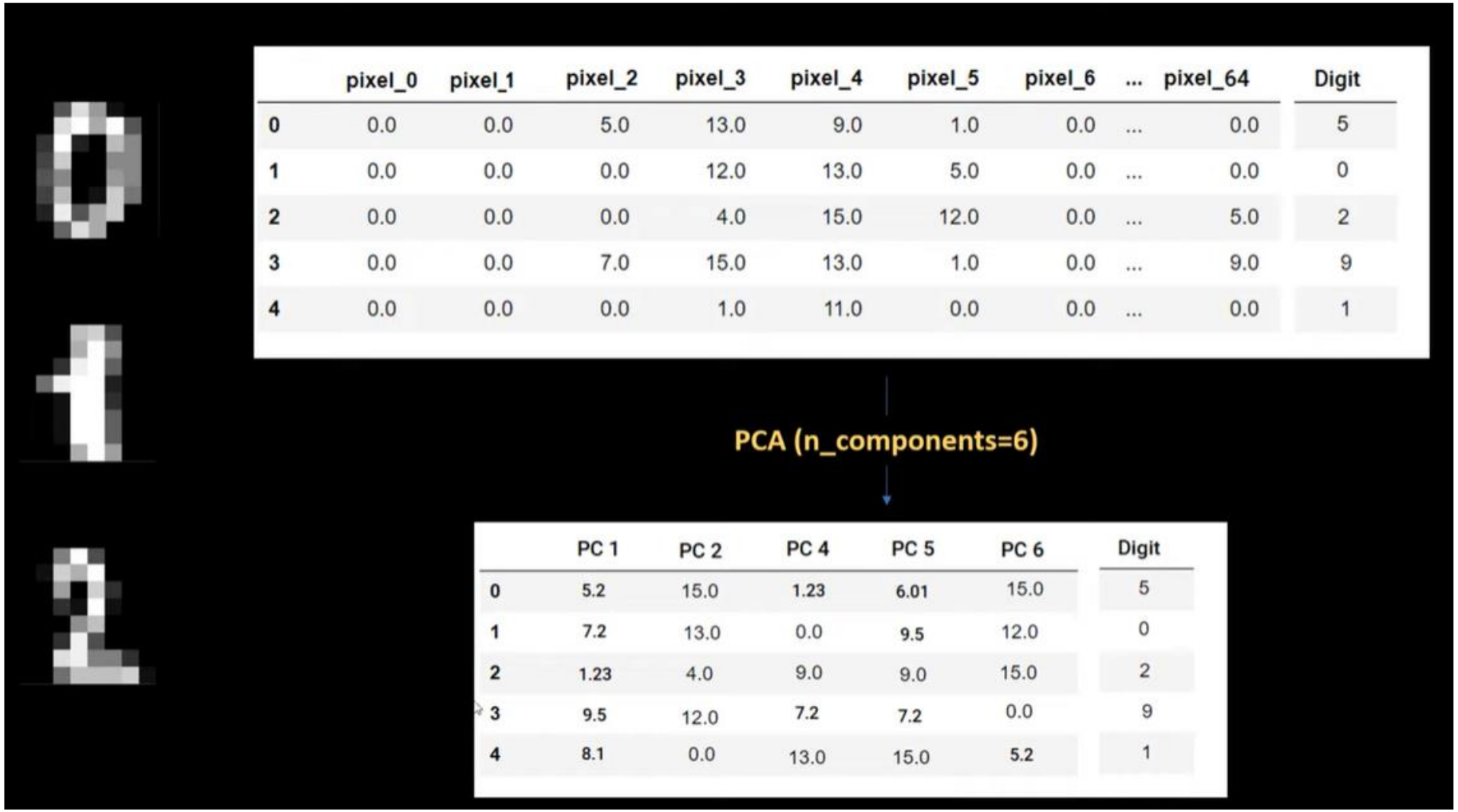


DATA SCIENCE

PCA: Principal Component Analysis

Is a process of figuring out most important features or principal components that has the most impact on the target variable

DATA SCIENCE



	pixel_0	pixel_1	pixel_2	pixel_3	pixel_4	pixel_5	pixel_6	...	pixel_64	Digit
0	0.0	0.0	5.0	13.0	9.0	1.0	0.0	...	0.0	5
1	0.0	0.0	0.0	12.0	13.0	5.0	0.0	...	0.0	0
2	0.0	0.0	0.0	4.0	15.0	12.0	0.0	...	5.0	2
3	0.0	0.0	7.0	15.0	13.0	1.0	0.0	...	9.0	9
4	0.0	0.0	0.0	1.0	11.0	0.0	0.0	...	0.0	1

PCA (n_components=6)

	PC 1	PC 2	PC 4	PC 5	PC 6	Digit
0	5.2	15.0	1.23	6.01	15.0	5
1	7.2	13.0	0.0	9.5	12.0	0
2	1.23	4.0	9.0	9.0	15.0	2
3	9.5	12.0	7.2	7.2	0.0	9
4	8.1	0.0	13.0	15.0	5.2	1

DATA SCIENCE

Few things to keep in mind before using PCA

- Scale Features Before Applying PCA
- Accuracy might drop