

3. Data Science – Machine Learning – Data & ML Algorithm

Contents

1. Data.....	2
2. Data in table	3
3. Row	3
4. Column	3
5. Cell.....	3
6. Statistical Learning Perspective	4
6.1. Example.....	4
7. Why Machine learning algorithms learn this function?.....	6
8. Terminology in statistics,	7
9. Computer Science Perspective.....	8
10. Models and Algorithms	9

3. Data Science – Machine Learning – Data & ML Algorithm

1. Data

- ✓ It is a collection of facts.
- ✓ Facts mean,
 - Alphabets
 - Numbers
 - Alphanumeric
 - Symbol

◇	A	B	C	D
1		Column 1	Column 2	Column 3
2	Row 1	2.2	2.3	1
3	Row 2	2.3	2.6	0
4	Row 3	2.1	2	1
5				

2. Data in table

- ✓ Generally, tables having data like rows, columns and cells
- ✓ Tabular data is very easy to understand.

3. Row

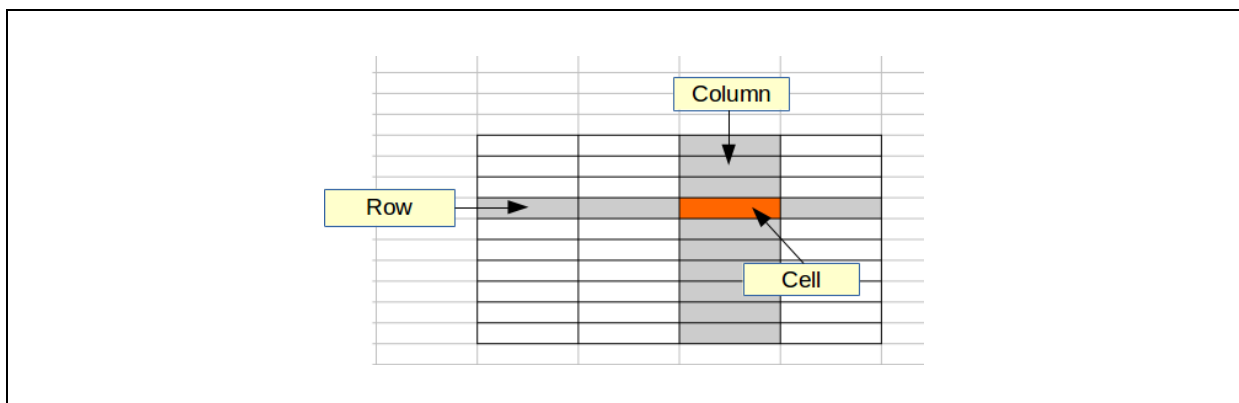
- ✓ A row describes a single entity or observation.
- ✓ A row is also called a record.

4. Column

- ✓ A column is a vertical group of cells within a table.
- ✓ A column having same type of values.
- ✓ In column we can store the data like weights, heights and prices etc info.

5. Cell

- ✓ A cell is a single value in a row and column.
- ✓ It may be a real value (1.5) an integer (2) or a category (red).



6. Statistical Learning Perspective

- ✓ In statistical learning perspective data means,
 - It is the context of a **hypothetical function** (f) that the machine learning algorithm is **trying to learn**.

6.1. Example

- ✓ Assuming that, there is a farmer in village and initially he had 1 acre land.
- ✓ While forming rice it get yields as below,

Area	Rice packets
1	10
2	20
3	30

If former having 4 acres land then how many rice packets it yields?

Answer

- ✓ A non-technical guy also can answer as, we will get **40** packets for **4** acres of land.

Area	Rice packets
1	10
2	20
3	30
4	?

Simple calculation

✓ Formula is,

- | | | | | |
|----------|----|---------------|---|-----------|
| ○ 1 acre | => | 1 x 10 | = | 10 |
| ○ 2 acre | => | 2 x 10 | = | 20 |
| ○ 3 acre | => | 3 x 10 | = | 30 |
| ○ 4 acre | => | 4 x 10 | = | 40 |

- ✓ Simple Formula is,
 - $\text{Rice_yield} = \text{land_size} \times 10$

7. Why Machine learning algorithms learn this function?

- ✓ To predict output for the given input.

$$\text{Output} = f(\text{Input})$$

- ✓ Columns are called as input variables.
- ✓ Predicted result is called as output variable or response variable.

$$\text{Output Variable} = f(\text{Input Variables})$$

◇	A	B	C
1	X1	X2	Y
2	2.2	2.3	1
3	2.3	2.6	0
4	2.1	2	1
5			

- ✓ Input data may have more than one input variable.
- ✓ The group of input variables are called as input vector.

$$\text{Output Variable} = f(\text{Input Vector})$$

8. Terminology in statistics,

- ✓ Input variables are called as Independent variables.
- ✓ Output variable are called as the Dependent variable.
- ✓ Here the output is dependent on a function of input.

$$\text{Dependent Variable} = f(\text{Independent Variables})$$

- ✓ Input variable representing with **X**
- ✓ Output variable representing with **y**

$$Y = f(X)$$

- ✓ If we have multiple input variables then it's represent as input vector x_1, x_2, x_3 for the data

9. Computer Science Perspective

- ✓ A row means it's an entity/observation/instance/object in a table.
- ✓ A Column is called as an attribute.
- ✓ During modeling and predictions,
 - Input is called as input attribute.
 - Output is called as output attributes.

Output Attribute = Program(Input Attributes)

◇	A	B	C	D
1		Attribute 1	Attribute 2	Output Attribute
2	Instance 1	2.2	2.3	1
3	Instance 2	2.3	2.6	0
4	Instance 3	2.1	2	1
5				

Output = Program(Input Features)

Prediction output = Program(Instance)

10. Models and Algorithms

- ✓ A model is a specific representation learned from data and the algorithm as the process of learning it.

$$\text{Model} = \text{Algorithm}(\text{Data})$$