



FIRST GLANCE AT ML

Iris Dataset

IRIS



- Iris is the name of a genus of flowers.
- There are 260-300 species of Iris.
- Can a machine classify the species of Iris, given a dataset?

UNDERSTANDING A DATASET

A	B	C	D	E
1 sepal_length	sepal_width	petal_length	petal_width	species
2 5	2	3.5	1	versicolor
3 6	2.2	4	1	versicolor
4 6	2.2	5	1.5	virginica
5 6.2	2.2	4.5	1.5	versicolor
6 4.5	2.3	1.3	0.3	setosa
7 5	2.3	3.3	1	versicolor
8 5.5	2.3	4	1.3	versicolor
9 6.3	2.3	4.4	1.3	versicolor
10 4.9	2.4	3.3	1	versicolor
11 5.5	2.4	3.8	1.1	versicolor
12 5.5	2.4	3.7	1	versicolor
13 4.9	2.5	4.5	1.7	virginica
14 5.1	2.5	3	1.1	versicolor
15 5.5	2.5	4	1.3	versicolor
16 5.6	2.5	3.9	1.1	versicolor

ATTRIBUTE

- Columns of the dataset
- It describes characteristics of each data.
- There are 2 types of attributes.
 - Numeric
 - Categorical

A	B	C	D	E
1 sepal_length	sepal_width	petal_length	petal_width	species
2 5	2	3.5	1	versicolor
3 6	2.2	4	1	versicolor
4 6	2.2	5	1.5	virginica
5 6.2	2.2	4.5	1.5	versicolor
6 4.5	2.3	1.3	0.3	setosa
7 5	2.3	3.3	1	versicolor
8 5.5	2.3	4	1.3	versicolor
9 6.3	2.3	4.4	1.3	versicolor
10 4.9	2.4	3.3	1	versicolor
11 5.5	2.4	3.8	1.1	versicolor
12 5.5	2.4	3.7	1	versicolor
13 4.9	2.5	4.5	1.7	virginica
14 5.1	2.5	3	1.1	versicolor
15 5.5	2.5	4	1.3	versicolor
16 5.6	2.5	3.9	1.1	versicolor

NUMERIC ATTRIBUTE

- Attributes that have numeric values.
- aka numbers

	A	B	C	D	E
1	sepal_length	sepal_width	petal_length	petal_width	species
2	5	2	3.5		1 versicolor
3	6	2.2	4		1 versicolor
4	6	2.2	5		1.5 virginica
5	6.2	2.2	4.5		1.5 versicolor
6	4.5	2.3	1.3		0.3 setosa
7	5	2.3	3.3		1 versicolor
8	5.5	2.3	4		1.3 versicolor
9	6.3	2.3	4.4		1.3 versicolor
10	4.9	2.4	3.3		1 versicolor
11	5.5	2.4	3.8		1.1 versicolor
12	5.5	2.4	3.7		1 versicolor
13	4.9	2.5	4.5		1.7 virginica
14	5.1	2.5	3		1.1 versicolor
15	5.5	2.5	4		1.3 versicolor
16	5.6	2.5	3.9		1.1 versicolor

NUMERIC ATTRIBUTE

- Attributes that have numeric values.
- aka numbers
- For example, this attribute contains information about sepal width of the flower.
- It contains numeric values of the width of the sepal.

	A	B	C	D	E
1	sepal_length	sepal_width	petal_length	petal_width	species
2	5	2	3.5		1 versicolor
3	6	2.2	4		1 versicolor
4	6	2.2	5		1.5 virginica
5	6.2	2.2	4.5		1.5 versicolor
6	4.5	2.3	1.3		0.3 setosa
7	5	2.3	3.3		1 versicolor
8	5.5	2.3	4		1.3 versicolor
9	6.3	2.3	4.4		1.3 versicolor
10	4.9	2.4	3.3		1 versicolor
11	5.5	2.4	3.8		1.1 versicolor
12	5.5	2.4	3.7		1 versicolor
13	4.9	2.5	4.5		1.7 virginica
14	5.1	2.5	3		1.1 versicolor
15	5.5	2.5	4		1.3 versicolor
16	5.6	2.5	3.9		1.1 versicolor

NOMINAL ATTRIBUTE

- Attributes which values are grouped.
- aka discrete value

	A	B	C	D	E
1	sepal_length	sepal_width	petal_length	petal_width	species
2	5	2	3.5	1	versicolor
3	6	2.2	4	1	versicolor
4	6	2.2	5	1.5	virginica
5	6.2	2.2	4.5	1.5	versicolor
6	4.5	2.3	1.3	0.3	setosa
7	5	2.3	3.3	1	versicolor
8	5.5	2.3	4	1.3	versicolor
9	6.3	2.3	4.4	1.3	versicolor
10	4.9	2.4	3.3	1	versicolor
11	5.5	2.4	3.8	1.1	versicolor
12	5.5	2.4	3.7	1	versicolor
13	4.9	2.5	4.5	1.7	virginica
14	5.1	2.5	3	1.1	versicolor
15	5.5	2.5	4	1.3	versicolor
16	5.6	2.5	3.9	1.1	versicolor

NOMINAL ATTRIBUTE

- Attributes which values are grouped.
- aka discrete value
- For example, this attribute describes the species of each data point.
- There are 3 groups (categories) of Iris: versicolor, virginica, and setosa.

	A	B	C	D	E
1	sepal_length	sepal_width	petal_length	petal_width	species
2	5	2	3.5	1	versicolor
3	6	2.2	4	1	versicolor
4	6	2.2	5	1.5	virginica
5	6.2	2.2	4.5	1.5	versicolor
6	4.5	2.3	1.3	0.3	setosa
7	5	2.3	3.3	1	versicolor
8	5.5	2.3	4	1.3	versicolor
9	6.3	2.3	4.4	1.3	versicolor
10	4.9	2.4	3.3	1	versicolor
11	5.5	2.4	3.8	1.1	versicolor
12	5.5	2.4	3.7	1	versicolor
13	4.9	2.5	4.5	1.7	virginica
14	5.1	2.5	3	1.1	versicolor
15	5.5	2.5	4	1.3	versicolor
16	5.6	2.5	3.9	1.1	versicolor

CLASS ATTRIBUTE

	A	B	C	D	E
1	sepal_length	sepal_width	petal_length	petal_width	species
2	5	2	3.5	1	versicolor
3	6	2.2	4	1	versicolor
4	6	2.2	5	1.5	virginica
5	6.2	2.2	4.5	1.5	versicolor
6	4.5	2.3	1.3	0.3	setosa
7	5	2.3	3.3	1	versicolor
8	5.5	2.3	4	1.3	versicolor
9	6.3	2.3	4.4	1.3	versicolor
10	4.9	2.4	3.3	1	versicolor
11	5.5	2.4	3.8	1.1	versicolor
12	5.5	2.4	3.7	1	versicolor
13	4.9	2.5	4.5	1.7	virginica
14	5.1	2.5	3	1.1	versicolor
15	5.5	2.5	4	1.3	versicolor
16	5.6	2.5	3.9	1.1	versicolor

- Attributes which classify the data.
- It tells the class of each data point.

INSTANCES

	A	B	C	D	E
1	sepal_length	sepal_width	petal_length	petal_width	species
2	5	2	3.5	1	versicolor
3	6	2.2	4	1	versicolor
4	6	2.2	5	1.5	virginica
5	6.2	2.2	4.5	1.5	versicolor
6	4.5	2.3	1.3	0.3	setosa
7	5	2.3	3.3	1	versicolor
8	5.5	2.3	4	1.3	versicolor
9	6.3	2.3	4.4	1.3	versicolor
10	4.9	2.4	3.3	1	versicolor
11	5.5	2.4	3.8	1.1	versicolor
12	5.5	2.4	3.7	1	versicolor
13	4.9	2.5	4.5	1.7	virginica
14	5.1	2.5	3	1.1	versicolor
15	5.5	2.5	4	1.3	versicolor
16	5.6	2.5	3.9	1.1	versicolor

- Rows of the dataset.
- Each row describes one data point.
- It tells the class of each data point.



LEARN THE SPECIES

- Given the dataset, can a machine learn the species of Iris?

- Lab 1: First look at Machine Learning
 - <https://github.com/BawornsakS/Machine-Learning-Lesson/>