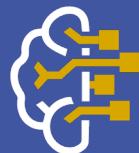


Exploratory Data Analysis

Data Mining: Seminar 4

Dr. Thomas Price

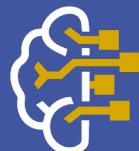


AI Academy

NC STATE UNIVERSITY

Datasets - Supervised vs. Unsupervised Learning

Lesson



AI Academy

Iris Sample Data Set

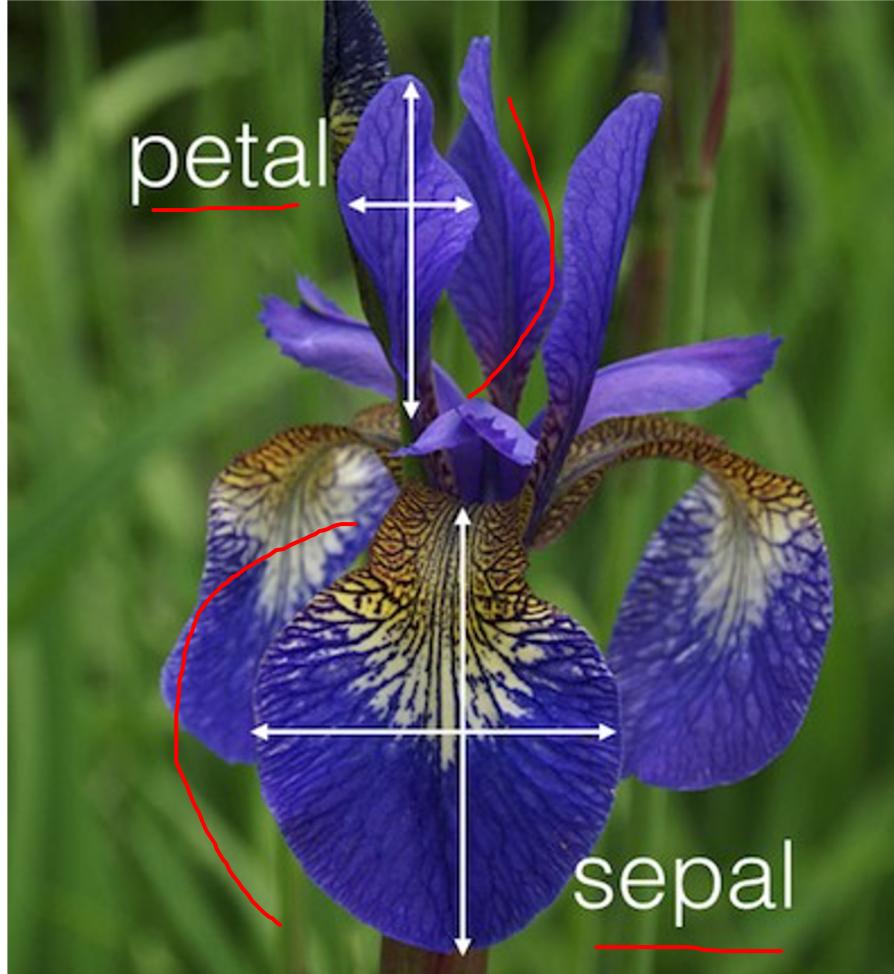
Can be obtained from
the UCI Machine
Learning Repository:

<https://archive.ics.uci.edu/ml/index.php>

Collected by Ronald
Fisher in 1936!



Iris Sample Data Set



Iris Sample Data Set

Three Flowers

- Setosa
- Versicolour
- Virginica

Number of Instances: 150

- 50 instances each



ID	<u>sepal length</u>	<u>sepal width</u>	<u>petal length</u>	<u>petal width</u>	class
1	5.1	3.5	1.4	0.2	Iris-setosa
2	4.9	3	1.4	0.2	Iris-setosa
3	4.7	3.2	1.3	0.2	Iris-setosa
4	4.6	3.1	1.5	0.2	Iris-setosa
5	5	3.6	1.4	0.2	Iris-setosa
...	Iris-setosa
51	7	3.2	4.7	1.4	Iris-versicolor
52	6.4	3.2	4.5	1.5	Iris-versicolor
53	6.9	3.1	4.9	1.5	Iris-versicolor
...	Iris-versicolor
101	6.3	3.3	6	2.5	Iris-virginica
102	5.8	2.7	5.1	1.9	Iris-virginica
103	7.1	3	5.9	2.1	Iris-virginica
104	6.3	2.9	5.6	1.8	Iris-virginica
...	Iris-virginica

Attributes (Columns)

Objects
(Rows)

Sepal length	Sepal width	Petal length	Petal width	Species
5.0	2.0	3.5	1.0	<i>I. versicolor</i>
6.0	2.2	4.0	1.0	<i>I. versicolor</i>
6.2	2.2	4.5	1.5	<i>I. versicolor</i>
6.0	2.2	5.0	1.5	<i>I. virginica</i>
4.5	2.3	1.3	0.3	<i>I. setosa</i>
5.5	2.3	4.0	1.3	<i>I. versicolor</i>

Supervised Learning on Iris Data

“Learning with a teacher”

Two type attributes:

- X: Independent variables
?
- Y: Dependent variables
?
- Task: $X \rightarrow Y$

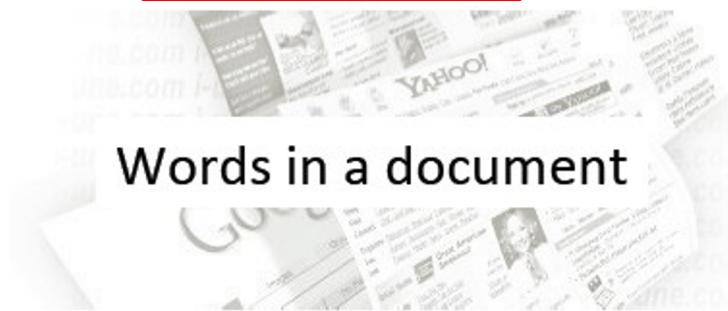
Attributes

Sepal length	Sepal width	Petal length	Petal width	Species
5.0	2.0	3.5	1.0	<i>I. versicolor</i>
6.0	2.2	4.0	1.0	<i>I. versicolor</i>
6.2	2.2	4.5	1.5	<i>I. versicolor</i>
6.0	2.2	5.0	1.5	<i>I. virginica</i>
4.5	2.3	1.3	0.3	<i>I. setosa</i>
5.5	2.3	4.0	1.3	<i>I. versicolor</i>



Supervised Learning

Feature Space \mathcal{X}



Words in a document

Label Space \mathcal{Y}

“Sports”
“News”
“Science”
...

Discrete Labels
Classification



Share Price
“\$ 24.50”

Continuous Labels
Regression

Task: Given $X \in \mathcal{X}$, predict $Y \in \mathcal{Y}$.

See Jupyter Notebook

*Iris Dataset
Unsupervised Learning
Iris Class Data*



Predicting Iris Species: Classification or Regression?

Datasets – Supervised vs. Unsupervised Learning

Exercises



AI Academy

Q&A Forum

Design a project idea involving the iris dataset.