Chapter Goals

- Basic simple ideas of data mining algorithms

Lab Goals

- Familiarity with how algorithms generate some outputs given patterns in inputs

General Guidlines

- Visualization
- Modifiable code snippets

```python
# Loading a dataset
# dataset names: "airline", "breast-cancer", "contact-lenses", "cpu",
"cpu.with.vendor", "credit-g", "diabetes", "glass", "hypothyroid",
"ionosphere", "iris.2D", "iris", "labor", "segment-challenge",
"segment-test", "soybean", "supermarket", "unbalanced", "vote",
"weather.nominal", "weather.numeric"
# df = pd.read_csv("data/weather.numeric.csv")
# instances = loader.load_file("data/weather.numeric.arff")
```

Modules & Datasets Setup

```
# @title
!apt-get install default-jdk
!apt install libgraphviz-dev

Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following additional packages will be installed:
  default-jdk-headless default-jre default-jre-headless fonts-dejavu-
core fonts-dejavu-extra
  libatk-wrapper-java libatk-wrapper-java-jni libfontenc1 libice-dev
libsm-dev libxkbfile1
  libxt-dev libxtst6 libxxf86dga1 openjdk-11-jdk openjdk-11-jre x11-
utils
Suggested packages:
  libice-doc libsm-doc libxt-doc openjdk-11-demo openjdk-11-source
visualvm mesa-utils
The following NEW packages will be installed:
```

```
  default-jdk default-jdk-headless default-jre default-jre-headless
fonts-dejavu-core
  fonts-dejavu-extra libatk-wrapper-java libatk-wrapper-java-jni
libfontenc1 libice-dev libsm-dev
  libxkbfile1 libxt-dev libxtst6 libxxf86dga1 openjdk-11-jdk openjdk-
11-jre x11-utils
0 upgraded, 18 newly installed, 0 to remove and 18 not upgraded.
Need to get 5,518 kB of archives.
After this operation, 15.8 MB of additional disk space will be used.
Get:1 http://archive.ubuntu.com/ubuntu jammy/main amd64 default-jre-
headless amd64 2:1.11-72build2 [3,042 B]
Get:2 http://archive.ubuntu.com/ubuntu jammy/main amd64 libxtst6 amd64
2:1.2.3-1build4 [13.4 kB]
Get:3 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64
openjdk-11-jre amd64 11.0.20.1+1-0ubuntu1~22.04 [213 kB]
Get:4 http://archive.ubuntu.com/ubuntu jammy/main amd64 default-jre
amd64 2:1.11-72build2 [896 B]
Get:5 http://archive.ubuntu.com/ubuntu jammy/main amd64 default-jdk-
headless amd64 2:1.11-72build2 [942 B]
Get:6 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64
openjdk-11-jdk amd64 11.0.20.1+1-0ubuntu1~22.04 [1,331 kB]
Get:7 http://archive.ubuntu.com/ubuntu jammy/main amd64 default-jdk
amd64 2:1.11-72build2 [908 B]
Get:8 http://archive.ubuntu.com/ubuntu jammy/main amd64 fonts-dejavu-
core all 2.37-2build1 [1,041 kB]
Get:9 http://archive.ubuntu.com/ubuntu jammy/main amd64 fonts-dejavu-
extra all 2.37-2build1 [2,041 kB]
Get:10 http://archive.ubuntu.com/ubuntu jammy/main amd64 libfontenc1
amd64 1:1.1.4-1build3 [14.7 kB]
Get:11 http://archive.ubuntu.com/ubuntu jammy/main amd64 libxkbfile1
amd64 1:1.1.0-1build3 [71.8 kB]
Get:12 http://archive.ubuntu.com/ubuntu jammy/main amd64 libxxf86dga1
amd64 2:1.1.5-0ubuntu3 [12.6 kB]
Get:13 http://archive.ubuntu.com/ubuntu jammy/main amd64 x11-utils
amd64 7.7+5build2 [206 kB]
Get:14 http://archive.ubuntu.com/ubuntu jammy/main amd64 libatk-
wrapper-java all 0.38.0-5build1 [53.1 kB]
Get:15 http://archive.ubuntu.com/ubuntu jammy/main amd64 libatk-
wrapper-java-jni amd64 0.38.0-5build1 [49.0 kB]
Get:16 http://archive.ubuntu.com/ubuntu jammy/main amd64 libice-dev
amd64 2:1.0.10-1build2 [51.4 kB]
Get:17 http://archive.ubuntu.com/ubuntu jammy/main amd64 libsm-dev
amd64 2:1.2.3-1build2 [18.1 kB]
Get:18 http://archive.ubuntu.com/ubuntu jammy/main amd64 libxt-dev
amd64 1:1.2.1-1 [396 kB]
Fetched 5,518 kB in 1s (5,868 kB/s)
Selecting previously unselected package default-jre-headless.
(Reading database ... 120874 files and directories currently
installed.)
```

```
Preparing to unpack .../00-default-jre-headless_2%3a1.11-
72build2_amd64.deb ...
Unpacking default-jre-headless (2:1.11-72build2) ...
Selecting previously unselected package libxtst6:amd64.
Preparing to unpack .../01-libxtst6_2%3a1.2.3-1build4_amd64.deb ...
Unpacking libxtst6:amd64 (2:1.2.3-1build4) ...
Selecting previously unselected package openjdk-11-jre:amd64.
Preparing to unpack .../02-openjdk-11-jre_11.0.20.1+1-
0ubuntu1~22.04_amd64.deb ...
Unpacking openjdk-11-jre:amd64 (11.0.20.1+1-0ubuntu1~22.04) ...
Selecting previously unselected package default-jre.
Preparing to unpack .../03-default-jre_2%3a1.11-72build2_amd64.deb ...
Unpacking default-jre (2:1.11-72build2) ...
Selecting previously unselected package default-jdk-headless.
Preparing to unpack .../04-default-jdk-headless_2%3a1.11-
72build2_amd64.deb ...
Unpacking default-jdk-headless (2:1.11-72build2) ...
Selecting previously unselected package openjdk-11-jdk:amd64.
Preparing to unpack .../05-openjdk-11-jdk_11.0.20.1+1-
0ubuntu1~22.04_amd64.deb ...
Unpacking openjdk-11-jdk:amd64 (11.0.20.1+1-0ubuntu1~22.04) ...
Selecting previously unselected package default-jdk.
Preparing to unpack .../06-default-jdk_2%3a1.11-72build2_amd64.deb ...
Unpacking default-jdk (2:1.11-72build2) ...
Selecting previously unselected package fonts-dejavu-core.
Preparing to unpack .../07-fonts-dejavu-core_2.37-2build1_all.deb ...
Unpacking fonts-dejavu-core (2.37-2build1) ...
Selecting previously unselected package fonts-dejavu-extra.
Preparing to unpack .../08-fonts-dejavu-extra_2.37-2build1_all.deb ...
Unpacking fonts-dejavu-extra (2.37-2build1) ...
Selecting previously unselected package libfontenc1:amd64.
Preparing to unpack .../09-libfontenc1_1%3a1.1.4-1build3_amd64.deb ...
Unpacking libfontenc1:amd64 (1:1.1.4-1build3) ...
Selecting previously unselected package libxkbfile1:amd64.
Preparing to unpack .../10-libxkbfile1_1%3a1.1.0-1build3_amd64.deb ...
Unpacking libxkbfile1:amd64 (1:1.1.0-1build3) ...
Selecting previously unselected package libxxf86dga1:amd64.
Preparing to unpack .../11-libxxf86dga1_2%3a1.1.5-
0ubuntu3_amd64.deb ...
Unpacking libxxf86dga1:amd64 (2:1.1.5-0ubuntu3) ...
Selecting previously unselected package x11-utils.
Preparing to unpack .../12-x11-utils_7.7+5build2_amd64.deb ...
Unpacking x11-utils (7.7+5build2) ...
Selecting previously unselected package libatk-wrapper-java.
Preparing to unpack .../13-libatk-wrapper-java_0.38.0-
5build1_all.deb ...
Unpacking libatk-wrapper-java (0.38.0-5build1) ...
Selecting previously unselected package libatk-wrapper-java-jni:amd64.
Preparing to unpack .../14-libatk-wrapper-java-jni_0.38.0-
```

```
5build1_amd64.deb ...
Unpacking libatk-wrapper-java-jni:amd64 (0.38.0-5build1) ...
Selecting previously unselected package libice-dev:amd64.
Preparing to unpack .../15-libice-dev_2%3a1.0.10-1build2_amd64.deb ...
Unpacking libice-dev:amd64 (2:1.0.10-1build2) ...
Selecting previously unselected package libsm-dev:amd64.
Preparing to unpack .../16-libsm-dev_2%3a1.2.3-1build2_amd64.deb ...
Unpacking libsm-dev:amd64 (2:1.2.3-1build2) ...
Selecting previously unselected package libxt-dev:amd64.
Preparing to unpack .../17-libxt-dev_1%3a1.2.1-1_amd64.deb ...
Unpacking libxt-dev:amd64 (1:1.2.1-1) ...
Setting up default-jre-headless (2:1.11-72build2) ...
Setting up libice-dev:amd64 (2:1.0.10-1build2) ...
Setting up libsm-dev:amd64 (2:1.2.3-1build2) ...
Setting up libxtst6:amd64 (2:1.2.3-1build4) ...
Setting up libxxf86dga1:amd64 (2:1.1.5-0ubuntu3) ...
Setting up openjdk-11-jre:amd64 (11.0.20.1+1-0ubuntu1~22.04) ...
Setting up default-jre (2:1.11-72build2) ...
Setting up libfontenc1:amd64 (1:1.1.4-1build3) ...
Setting up default-jdk-headless (2:1.11-72build2) ...
Setting up libxt-dev:amd64 (1:1.2.1-1) ...
Setting up fonts-dejavu-core (2.37-2build1) ...
Setting up fonts-dejavu-extra (2.37-2build1) ...
Setting up openjdk-11-jdk:amd64 (11.0.20.1+1-0ubuntu1~22.04) ...
update-alternatives: using
/usr/lib/jvm/java-11-openjdk-amd64/bin/jconsole to provide
/usr/bin/jconsole (jconsole) in auto mode
Setting up libxkbfile1:amd64 (1:1.1.0-1build3) ...
Setting up default-jdk (2:1.11-72build2) ...
Setting up x11-utils (7.7+5build2) ...
Setting up libatk-wrapper-java (0.38.0-5build1) ...
Setting up libatk-wrapper-java-jni:amd64 (0.38.0-5build1) ...
Processing triggers for hicolor-icon-theme (0.17-2) ...
Processing triggers for libc-bin (2.35-0ubuntu3.1) ...
/sbin/ldconfig.real: /usr/local/lib/libtbbbind_2_0.so.3 is not a
symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbb.so.12 is not a symbolic
link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind.so.3 is not a symbolic
link

/sbin/ldconfig.real: /usr/local/lib/libtbbmalloc_proxy.so.2 is not a
symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind_2_5.so.3 is not a
symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbmalloc.so.2 is not a
```

```
symbolic link

Processing triggers for man-db (2.10.2-1) ...
Processing triggers for fontconfig (2.13.1-4.2ubuntu5) ...
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following additional packages will be installed:
  libgail-common libgail18 libgtk2.0-0 libgtk2.0-bin libgtk2.0-common
libgvc6-plugins-gtk
  librsvg2-common libxdot4
Suggested packages:
  gvfs
The following NEW packages will be installed:
  libgail-common libgail18 libgraphviz-dev libgtk2.0-0 libgtk2.0-bin
libgtk2.0-common
  libgvc6-plugins-gtk librsvg2-common libxdot4
0 upgraded, 9 newly installed, 0 to remove and 18 not upgraded.
Need to get 2,433 kB of archives.
After this operation, 7,694 kB of additional disk space will be used.
Get:1 http://archive.ubuntu.com/ubuntu jammy/main amd64 libgtk2.0-
common all 2.24.33-2ubuntu2 [125 kB]
Get:2 http://archive.ubuntu.com/ubuntu jammy/main amd64 libgtk2.0-0
amd64 2.24.33-2ubuntu2 [2,037 kB]
Get:3 http://archive.ubuntu.com/ubuntu jammy/main amd64 libgail18
amd64 2.24.33-2ubuntu2 [15.9 kB]
Get:4 http://archive.ubuntu.com/ubuntu jammy/main amd64 libgail-common
amd64 2.24.33-2ubuntu2 [132 kB]
Get:5 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libxdot4
amd64 2.42.2-6 [16.4 kB]
Get:6 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libgvc6-
plugins-gtk amd64 2.42.2-6 [22.6 kB]
Get:7 http://archive.ubuntu.com/ubuntu jammy/universe amd64
libgraphviz-dev amd64 2.42.2-6 [58.5 kB]
Get:8 http://archive.ubuntu.com/ubuntu jammy/main amd64 libgtk2.0-bin
amd64 2.24.33-2ubuntu2 [7,932 B]
Get:9 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64
librsvg2-common amd64 2.52.5+dfsg-3ubuntu0.2 [17.7 kB]
Fetched 2,433 kB in 1s (3,330 kB/s)
Selecting previously unselected package libgtk2.0-common.
(Reading database ... 121385 files and directories currently
installed.)
Preparing to unpack .../0-libgtk2.0-common_2.24.33-
2ubuntu2_all.deb ...
Unpacking libgtk2.0-common (2.24.33-2ubuntu2) ...
Selecting previously unselected package libgtk2.0-0:amd64.
Preparing to unpack .../1-libgtk2.0-0_2.24.33-2ubuntu2_amd64.deb ...
Unpacking libgtk2.0-0:amd64 (2.24.33-2ubuntu2) ...
Selecting previously unselected package libgail18:amd64.
```

```
Preparing to unpack .../2-libgail18_2.24.33-2ubuntu2_amd64.deb ...
Unpacking libgail18:amd64 (2.24.33-2ubuntu2) ...
Selecting previously unselected package libgail-common:amd64.
Preparing to unpack .../3-libgail-common_2.24.33-
2ubuntu2_amd64.deb ...
Unpacking libgail-common:amd64 (2.24.33-2ubuntu2) ...
Selecting previously unselected package libxdot4:amd64.
Preparing to unpack .../4-libxdot4_2.42.2-6_amd64.deb ...
Unpacking libxdot4:amd64 (2.42.2-6) ...
Selecting previously unselected package libgvc6-plugins-gtk.
Preparing to unpack .../5-libgvc6-plugins-gtk_2.42.2-6_amd64.deb ...
Unpacking libgvc6-plugins-gtk (2.42.2-6) ...
Selecting previously unselected package libgraphviz-dev:amd64.
Preparing to unpack .../6-libgraphviz-dev_2.42.2-6_amd64.deb ...
Unpacking libgraphviz-dev:amd64 (2.42.2-6) ...
Selecting previously unselected package libgtk2.0-bin.
Preparing to unpack .../7-libgtk2.0-bin_2.24.33-2ubuntu2_amd64.deb ...
Unpacking libgtk2.0-bin (2.24.33-2ubuntu2) ...
Selecting previously unselected package librsvg2-common:amd64.
Preparing to unpack .../8-librsvg2-common_2.52.5+dfsg-
3ubuntu0.2_amd64.deb ...
Unpacking librsvg2-common:amd64 (2.52.5+dfsg-3ubuntu0.2) ...
Setting up libxdot4:amd64 (2.42.2-6) ...
Setting up librsvg2-common:amd64 (2.52.5+dfsg-3ubuntu0.2) ...
Setting up libgtk2.0-common (2.24.33-2ubuntu2) ...
Setting up libgtk2.0-0:amd64 (2.24.33-2ubuntu2) ...
Setting up libgvc6-plugins-gtk (2.42.2-6) ...
Setting up libgail18:amd64 (2.24.33-2ubuntu2) ...
Setting up libgtk2.0-bin (2.24.33-2ubuntu2) ...
Setting up libgail-common:amd64 (2.24.33-2ubuntu2) ...
Setting up libgraphviz-dev:amd64 (2.42.2-6) ...
Processing triggers for libc-bin (2.35-0ubuntu3.1) ...
/sbin/ldconfig.real: /usr/local/lib/libtbbbind_2_0.so.3 is not a
symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbb.so.12 is not a symbolic
link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind.so.3 is not a symbolic
link

/sbin/ldconfig.real: /usr/local/lib/libtbbmalloc_proxy.so.2 is not a
symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind_2_5.so.3 is not a
symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbmalloc.so.2 is not a
symbolic link
```

```
Processing triggers for man-db (2.10.2-1) ...
Processing triggers for libgdk-pixbuf-2.0-0:amd64 (2.42.8+dfsg-
1ubuntu0.2) ...

# @title
!pip install pygraphviz
!pip install python-javabridge
!pip install python-weka-wrapper3
!pip install sklearn-weka-plugin

Collecting pygraphviz
  Downloading pygraphviz-1.11.zip (120 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 120.8/120.8 kB 1.8 MB/s eta
0:00:00
etadata (setup.py) ... e=pygraphviz-1.11-cp310-cp310-linux_x86_64.whl
size=175926
sha256=f8553fee25a0d4b3091f5149a0d5f17a43564c37a5a1b2d603a6ddfa263aa20
f
  Stored in directory:
/root/.cache/pip/wheels/5b/ee/36/f47a0d35664fbe1a2b5a433ae33c6ad636b00
bb231f68a9aaa
Successfully built pygraphviz
Installing collected packages: pygraphviz
Successfully installed pygraphviz-1.11
Collecting python-javabridge
  Downloading python-javabridge-4.0.3.tar.gz (1.3 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 1.3/1.3 MB 9.4 MB/s eta
0:00:00
etadata (setup.py) ... ent already satisfied: numpy>=1.20.1 in
/usr/local/lib/python3.10/dist-packages (from python-javabridge)
(1.23.5)
Building wheels for collected packages: python-javabridge
  Building wheel for python-javabridge (setup.py) ...
e=python_javabridge-4.0.3-cp310-cp310-linux_x86_64.whl size=1743151
sha256=2a1bf7c136bcf50e71173608bb78d9df5bba8ab21602176a8276b769b58c97c
5
  Stored in directory:
/root/.cache/pip/wheels/35/58/be/c5d71b71a9dd6585f897fa5b2d021e03962eb
30d6b20797396
Successfully built python-javabridge
Installing collected packages: python-javabridge
Successfully installed python-javabridge-4.0.3
Collecting python-weka-wrapper3
  Downloading python-weka-wrapper3-0.2.14.tar.gz (15.9 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 15.9/15.9 MB 17.7 MB/s eta
0:00:00
etadata (setup.py) ... ent already satisfied: python-javabridge>=4.0.0
in /usr/local/lib/python3.10/dist-packages (from python-weka-wrapper3)
(4.0.3)
Requirement already satisfied: numpy in
```

```
/usr/local/lib/python3.10/dist-packages (from python-weka-wrapper3)
(1.23.5)
Requirement already satisfied: packaging in
/usr/local/lib/python3.10/dist-packages (from python-weka-wrapper3)
(23.2)
Collecting configurable-objects (from python-weka-wrapper3)
  Downloading configurable-objects-0.0.1.tar.gz (4.4 kB)
  Preparing metadata (setup.py) ... ple-data-flow (from python-weka-
wrapper3)
  Downloading simple-data-flow-0.0.1.tar.gz (16 kB)
  Preparing metadata (setup.py) ... ple-data-flow
  Building wheel for python-weka-wrapper3 (setup.py) ...
e=python_weka_wrapper3-0.2.14-py3-none-any.whl size=14496261
sha256=64da6c3dbdc3233d6f7ac98bca45b56f6246d346c7813d3e6bd23c784c5ae73
f
  Stored in directory:
/root/.cache/pip/wheels/80/c5/f2/412fa8d3b181151e11b68d46daa52f96e9b83
2a2eca4bc6c88
  Building wheel for configurable-objects (setup.py) ...
e=configurable_objects-0.0.1-py3-none-any.whl size=4695
sha256=f53a3d0bfafb5a357f817a97f731ca0c2ec224a3e4b3692ad5427176cd479e0
a
  Stored in directory:
/root/.cache/pip/wheels/ef/11/bc/75ac8b0592c38dc42412942c37d3947faf0b2
22bad150132a1
  Building wheel for simple-data-flow (setup.py) ... ple-data-flow:
filename=simple_data_flow-0.0.1-py3-none-any.whl size=19063
sha256=60c63a718850e45b3d548b24ffdc0a9a27512a8751fe606f9385e44aa086db7
f
  Stored in directory:
/root/.cache/pip/wheels/b3/02/23/4aec0db3dae7152dd268d6de385905116af55
229c1a8e81303
Successfully built python-weka-wrapper3 configurable-objects simple-
data-flow
Installing collected packages: configurable-objects, simple-data-flow,
python-weka-wrapper3
Successfully installed configurable-objects-0.0.1 python-weka-
wrapper3-0.2.14 simple-data-flow-0.0.1
Collecting sklearn-weka-plugin
  Downloading sklearn-weka-plugin-0.0.7.tar.gz (69 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 69.8/69.8 kB 1.6 MB/s eta
0:00:00
etadata (setup.py) ... ent already satisfied: numpy in
/usr/local/lib/python3.10/dist-packages (from sklearn-weka-plugin)
(1.23.5)
Requirement already satisfied: python-weka-wrapper3>=0.2.5 in
/usr/local/lib/python3.10/dist-packages (from sklearn-weka-plugin)
(0.2.14)
Collecting sklearn (from sklearn-weka-plugin)
```

```python
# @title
#Restart runtime after installing the dependencies

# @title
import os
import glob
import numpy as np
import pandas as pd
import weka.core.jvm as jvm
from weka.core import converters
import matplotlib.pyplot as plt

# @title
data_dir = 'data'

# @title
#!rm -r weka
#!rm -r data
```

```python
# @title
#jvm.stop()
jvm.start(packages=True)
```

```
DEBUG:weka.core.jvm:Adding bundled jars
DEBUG:weka.core.jvm:Classpath=['/usr/local/lib/python3.10/dist-
packages/javabridge/jars/rhino-1.7R4.jar',
'/usr/local/lib/python3.10/dist-packages/javabridge/jars/runnablequeue
.jar',
'/usr/local/lib/python3.10/dist-packages/javabridge/jars/cpython.jar',
'/usr/local/lib/python3.10/dist-packages/weka/lib/mtj.jar',
'/usr/local/lib/python3.10/dist-packages/weka/lib/weka.jar',
'/usr/local/lib/python3.10/dist-packages/weka/lib/arpack_combined.jar'
, '/usr/local/lib/python3.10/dist-packages/weka/lib/python-weka-
wrapper.jar',
'/usr/local/lib/python3.10/dist-packages/weka/lib/core.jar']
DEBUG:weka.core.jvm:MaxHeapSize=default
DEBUG:weka.core.jvm:Package support enabled
```

```python
# @title
# Preparing Datasets
if not os.path.exists(data_dir):
    !mkdir $data_dir
    for file in ['airline.arff', 'breast-cancer.arff', 'contact-
lenses.arff', 'cpu.arff', 'cpu.with.vendor.arff', 'credit-g.arff',
'diabetes.arff', 'glass.arff', 'hypothyroid.arff', 'ionosphere.arff',
'iris.2D.arff', 'iris.arff', 'labor.arff', 'segment-challenge.arff',
'segment-test.arff', 'soybean.arff', 'supermarket.arff',
'unbalanced.arff', 'vote.arff', 'weather.nominal.arff',
'weather.numeric.arff',]:
        url =
'https://git.cms.waikato.ac.nz/weka/weka/-/raw/main/trunk/wekadocs/
data/' + file
        !wget -P $data_dir $url
    loader =
converters.Loader(classname="weka.core.converters.ArffLoader")
    saver =
converters.Saver(classname="weka.core.converters.CSVSaver")
    for file in glob.glob(os.path.join(data_dir, '*.arff')):
        dataset = loader.load_file(file)
        filename, file_extension = os.path.splitext(file)
        saver.save_file(dataset, filename + '.csv')
    !wget -P $data_dir https://raw.githubusercontent.com/Rytuo/ITMO-
CT/master/Others/AdvancedML/data/OpenML/data/1438.arff
    !rm -r weka
```

```
--2023-10-24 14:54:26--
https://git.cms.waikato.ac.nz/weka/weka/-/raw/main/trunk/wekadocs/
data/airline.arff
Resolving git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)...
```

130.217.218.43
Connecting to git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)|
130.217.218.43|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 2357 (2.3K) [text/plain]
Saving to: 'data/airline.arff'

airline.arff   100%[===================>]   2.30K --.-KB/s   in
0s

2023-10-24 14:54:27 (1.09 GB/s) - 'data/airline.arff' saved
[2357/2357]

--2023-10-24 14:54:27--
https://git.cms.waikato.ac.nz/weka/weka/-/raw/main/trunk/wekadocs/
data/breast-cancer.arff
Resolving git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)...
130.217.218.43
Connecting to git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)|
130.217.218.43|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 29418 (29K) [text/plain]
Saving to: 'data/breast-cancer.arff'

breast-cancer.arff 100%[===================>] 28.73K   188KB/s   in
0.2s

2023-10-24 14:54:28 (188 KB/s) - 'data/breast-cancer.arff' saved
[29418/29418]

--2023-10-24 14:54:28--
https://git.cms.waikato.ac.nz/weka/weka/-/raw/main/trunk/wekadocs/
data/contact-lenses.arff
Resolving git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)...
130.217.218.43
Connecting to git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)|
130.217.218.43|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 2890 (2.8K) [text/plain]
Saving to: 'data/contact-lenses.arff'

contact-lenses.arff 100%[===================>]   2.82K --.-KB/s   in
0s

2023-10-24 14:54:29 (122 MB/s) - 'data/contact-lenses.arff' saved
[2890/2890]

--2023-10-24 14:54:29--
https://git.cms.waikato.ac.nz/weka/weka/-/raw/main/trunk/wekadocs/
data/cpu.arff

```
Resolving git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)...
130.217.218.43
Connecting to git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)|
130.217.218.43|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 5561 (5.4K) [text/plain]
Saving to: 'data/cpu.arff'

cpu.arff              100%[====================>]   5.43K  --.-KB/s    in
0s

2023-10-24 14:54:29 (127 MB/s) - 'data/cpu.arff' saved [5561/5561]

--2023-10-24 14:54:29--
https://git.cms.waikato.ac.nz/weka/weka/-/raw/main/trunk/wekadocs/
data/cpu.with.vendor.arff
Resolving git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)...
130.217.218.43
Connecting to git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)|
130.217.218.43|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 6960 (6.8K) [text/plain]
Saving to: 'data/cpu.with.vendor.arff'

cpu.with.vendor.arf 100%[====================>]   6.80K  --.-KB/s    in
0s

2023-10-24 14:54:30 (127 MB/s) - 'data/cpu.with.vendor.arff' saved
[6960/6960]

--2023-10-24 14:54:30--
https://git.cms.waikato.ac.nz/weka/weka/-/raw/main/trunk/wekadocs/
data/credit-g.arff
Resolving git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)...
130.217.218.43
Connecting to git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)|
130.217.218.43|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 162270 (158K) [text/plain]
Saving to: 'data/credit-g.arff'

credit-g.arff         100%[====================>] 158.47K    257KB/s    in
0.6s

2023-10-24 14:54:31 (257 KB/s) - 'data/credit-g.arff' saved
[162270/162270]

--2023-10-24 14:54:31--
https://git.cms.waikato.ac.nz/weka/weka/-/raw/main/trunk/wekadocs/
data/diabetes.arff
```

```
Resolving git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)...
130.217.218.43
Connecting to git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)|
130.217.218.43|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 37443 (37K) [text/plain]
Saving to: 'data/diabetes.arff'

diabetes.arff        100%[===================>]  36.57K   240KB/s    in
0.2s

2023-10-24 14:54:32 (240 KB/s) - 'data/diabetes.arff' saved
[37443/37443]

--2023-10-24 14:54:32--
https://git.cms.waikato.ac.nz/weka/weka/-/raw/main/trunk/wekadocs/
data/glass.arff
Resolving git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)...
130.217.218.43
Connecting to git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)|
130.217.218.43|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 17850 (17K) [text/plain]
Saving to: 'data/glass.arff'

glass.arff           100%[===================>]  17.43K   114KB/s    in
0.2s

2023-10-24 14:54:33 (114 KB/s) - 'data/glass.arff' saved [17850/17850]

--2023-10-24 14:54:33--
https://git.cms.waikato.ac.nz/weka/weka/-/raw/main/trunk/wekadocs/
data/hypothyroid.arff
Resolving git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)...
130.217.218.43
Connecting to git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)|
130.217.218.43|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 310897 (304K) [text/plain]
Saving to: 'data/hypothyroid.arff'

hypothyroid.arff     100%[===================>] 303.61K   491KB/s    in
0.6s

2023-10-24 14:54:34 (491 KB/s) - 'data/hypothyroid.arff' saved
[310897/310897]

--2023-10-24 14:54:35--
https://git.cms.waikato.ac.nz/weka/weka/-/raw/main/trunk/wekadocs/
data/ionosphere.arff
```

```
Resolving git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)...
130.217.218.43
Connecting to git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)|
130.217.218.43|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 80487 (79K) [text/plain]
Saving to: 'data/ionosphere.arff'

ionosphere.arff      100%[===================>]  78.60K   257KB/s    in
0.3s

2023-10-24 14:54:36 (257 KB/s) - 'data/ionosphere.arff' saved
[80487/80487]

--2023-10-24 14:54:36--
https://git.cms.waikato.ac.nz/weka/weka/-/raw/main/trunk/wekadocs/
data/iris.2D.arff
Resolving git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)...
130.217.218.43
Connecting to git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)|
130.217.218.43|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 3492 (3.4K) [text/plain]
Saving to: 'data/iris.2D.arff'

iris.2D.arff         100%[===================>]   3.41K   --.-KB/s    in
0s

2023-10-24 14:54:36 (59.0 MB/s) - 'data/iris.2D.arff' saved
[3492/3492]

--2023-10-24 14:54:36--
https://git.cms.waikato.ac.nz/weka/weka/-/raw/main/trunk/wekadocs/
data/iris.arff
Resolving git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)...
130.217.218.43
Connecting to git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)|
130.217.218.43|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 7486 (7.3K) [text/plain]
Saving to: 'data/iris.arff'

iris.arff            100%[===================>]   7.31K   --.-KB/s    in
0s

2023-10-24 14:54:37 (115 MB/s) - 'data/iris.arff' saved [7486/7486]

--2023-10-24 14:54:37--
https://git.cms.waikato.ac.nz/weka/weka/-/raw/main/trunk/wekadocs/
data/labor.arff
```

```
Resolving git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)...
130.217.218.43
Connecting to git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)|
130.217.218.43|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 8279 (8.1K) [text/plain]
Saving to: 'data/labor.arff'

labor.arff          100%[====================>]   8.08K  --.-KB/s    in
0s

2023-10-24 14:54:38 (107 MB/s) - 'data/labor.arff' saved [8279/8279]

--2023-10-24 14:54:38--
https://git.cms.waikato.ac.nz/weka/weka/-/raw/main/trunk/wekadocs/
data/segment-challenge.arff
Resolving git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)...
130.217.218.43
Connecting to git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)|
130.217.218.43|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 200410 (196K) [text/plain]
Saving to: 'data/segment-challenge.arff'

segment-challenge.a 100%[====================>] 195.71K   426KB/s    in
0.5s

2023-10-24 14:54:39 (426 KB/s) - 'data/segment-challenge.arff' saved
[200410/200410]

--2023-10-24 14:54:39--
https://git.cms.waikato.ac.nz/weka/weka/-/raw/main/trunk/wekadocs/
data/segment-test.arff
Resolving git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)...
130.217.218.43
Connecting to git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)|
130.217.218.43|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 109984 (107K) [text/plain]
Saving to: 'data/segment-test.arff'

segment-test.arff   100%[====================>] 107.41K   233KB/s    in
0.5s

2023-10-24 14:54:40 (233 KB/s) - 'data/segment-test.arff' saved
[109984/109984]

--2023-10-24 14:54:40--
https://git.cms.waikato.ac.nz/weka/weka/-/raw/main/trunk/wekadocs/
data/soybean.arff
```

```
Resolving git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)...
130.217.218.43
Connecting to git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)|
130.217.218.43|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 202935 (198K) [text/plain]
Saving to: 'data/soybean.arff'

soybean.arff          100%[===================>] 198.18K   324KB/s    in
0.6s

2023-10-24 14:54:42 (324 KB/s) - 'data/soybean.arff' saved
[202935/202935]

--2023-10-24 14:54:42--
https://git.cms.waikato.ac.nz/weka/weka/-/raw/main/trunk/wekadocs/
data/supermarket.arff
Resolving git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)...
130.217.218.43
Connecting to git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)|
130.217.218.43|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 2025871 (1.9M) [text/plain]
Saving to: 'data/supermarket.arff'

supermarket.arff     100%[===================>]   1.93M  1.81MB/s    in
1.1s

2023-10-24 14:54:43 (1.81 MB/s) - 'data/supermarket.arff' saved
[2025871/2025871]

--2023-10-24 14:54:43--
https://git.cms.waikato.ac.nz/weka/weka/-/raw/main/trunk/wekadocs/
data/unbalanced.arff
Resolving git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)...
130.217.218.43
Connecting to git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)|
130.217.218.43|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 186360 (182K) [text/plain]
Saving to: 'data/unbalanced.arff'

unbalanced.arff      100%[===================>] 181.99K   296KB/s    in
0.6s

2023-10-24 14:54:45 (296 KB/s) - 'data/unbalanced.arff' saved
[186360/186360]

--2023-10-24 14:54:45--
https://git.cms.waikato.ac.nz/weka/weka/-/raw/main/trunk/wekadocs/
```

```
data/vote.arff
Resolving git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)...
130.217.218.43
Connecting to git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)|
130.217.218.43|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 40261 (39K) [text/plain]
Saving to: 'data/vote.arff'

vote.arff           100%[===================>]  39.32K   260KB/s    in
0.2s

2023-10-24 14:54:46 (260 KB/s) - 'data/vote.arff' saved [40261/40261]

--2023-10-24 14:54:46--
https://git.cms.waikato.ac.nz/weka/weka/-/raw/main/trunk/wekadocs/
data/weather.nominal.arff
Resolving git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)...
130.217.218.43
Connecting to git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)|
130.217.218.43|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 587 [text/plain]
Saving to: 'data/weather.nominal.arff'

weather.nominal.arf 100%[===================>]     587  --.-KB/s    in
0s

2023-10-24 14:54:46 (172 MB/s) - 'data/weather.nominal.arff' saved
[587/587]

--2023-10-24 14:54:47--
https://git.cms.waikato.ac.nz/weka/weka/-/raw/main/trunk/wekadocs/
data/weather.numeric.arff
Resolving git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)...
130.217.218.43
Connecting to git.cms.waikato.ac.nz (git.cms.waikato.ac.nz)|
130.217.218.43|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 495 [text/plain]
Saving to: 'data/weather.numeric.arff'

weather.numeric.arf 100%[===================>]     495  --.-KB/s    in
0s

2023-10-24 14:54:47 (234 MB/s) - 'data/weather.numeric.arff' saved
[495/495]

--2023-10-24 14:54:50--  https://raw.githubusercontent.com/Rytuo/ITMO-
CT/master/Others/AdvancedML/data/OpenML/data/1438.arff
```

```
Resolving raw.githubusercontent.com (raw.githubusercontent.com)...
185.199.111.133, 185.199.110.133, 185.199.108.133, ...
Connecting to raw.githubusercontent.com (raw.githubusercontent.com)|
185.199.111.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 25381 (25K) [text/plain]
Saving to: 'data/1438.arff'

1438.arff            100%[===================>]  24.79K  --.-KB/s    in
0.001s

2023-10-24 14:54:50 (16.3 MB/s) - 'data/1438.arff' saved [25381/25381]

rm: cannot remove 'weka': No such file or directory

# @title
import weka.core.packages as packages
packages.install_package("simpleEducationalLearningSchemes")

from weka.core.converters import Loader
loader = Loader(classname="weka.core.converters.ArffLoader")
```

4.1 Inferring Rudimentary Rules

Sources

- [1R by Weka](#)

```
# weather numeric dataset
instances = loader.load_file("data/weather.numeric.arff")
instances

@relation weather

@attribute outlook {sunny,overcast,rainy}
@attribute temperature numeric
@attribute humidity numeric
@attribute windy {TRUE,FALSE}
@attribute play {yes,no}

@data
sunny,85,85,FALSE,no
sunny,80,90,TRUE,no
overcast,83,86,FALSE,yes
rainy,70,96,FALSE,yes
rainy,68,80,FALSE,yes
rainy,65,70,TRUE,no
overcast,64,65,TRUE,yes
sunny,72,95,FALSE,no
```

```
sunny,69,70,FALSE,yes
rainy,75,80,FALSE,yes
sunny,75,70,TRUE,yes
overcast,72,90,TRUE,yes
overcast,81,75,FALSE,yes
rainy,71,91,TRUE,no

# classes indices
# yes, no = 0, 1

# set class attribute to the last attribute, i.e attribute "play"
instances.class_index = instances.num_attributes - 1

# from doc but does not work!
# instances.setClassIndex(instances.numAttributes() - 1);
# use help(instances)

# train model
from weka.classifiers import Classifier
cls = Classifier(classname="weka.classifiers.rules.OneR")
cls.build_classifier(instances)

# model inspection
cls.description

<bound method OptionHandler.description of outlook:
        sunny -> no
        overcast    -> yes
        rainy -> yes
(10/14 instances correct)
>

# instance (row)
instances.get_instance(0)

sunny,85,85,FALSE,no

# prediction on instance
cls.classify_instance( instances.get_instance(0) )

# no

1.0
```

4.2 Simple Probabilistic Model

Sources

- Naive Bayes by Weka
- Naive Bayes by Scikit-Learn (different implementation)

```python
# weather numeric dataset
instances = loader.load_file("data/weather.nominal.arff")
instances
```

```
@relation weather.symbolic

@attribute outlook {sunny,overcast,rainy}
@attribute temperature {hot,mild,cool}
@attribute humidity {high,normal}
@attribute windy {TRUE,FALSE}
@attribute play {yes,no}

@data
sunny,hot,high,FALSE,no
sunny,hot,high,TRUE,no
overcast,hot,high,FALSE,yes
rainy,mild,high,FALSE,yes
rainy,cool,normal,FALSE,yes
rainy,cool,normal,TRUE,no
overcast,cool,normal,TRUE,yes
sunny,mild,high,FALSE,no
sunny,cool,normal,FALSE,yes
rainy,mild,normal,FALSE,yes
sunny,mild,normal,TRUE,yes
overcast,mild,high,TRUE,yes
overcast,hot,normal,FALSE,yes
rainy,mild,high,TRUE,no
```

```python
# classes indices
# yes, no = 0, 1

# set class attribute to the last attribute, i.e attribute "play"
instances.class_index = instances.num_attributes - 1

# train model
from weka.classifiers import Classifier
cls = Classifier(classname="weka.classifiers.bayes.NaiveBayes")
cls.build_classifier(instances)

# model inspection
cls.description
```

```
<bound method OptionHandler.description of Naive Bayes Classifier

                Class
Attribute         yes      no
               (0.63) (0.38)
===========================
outlook
   sunny          3.0     4.0
   overcast       5.0     1.0
```

```
   rainy              4.0     3.0
   [total]           12.0     8.0

temperature
   hot                3.0     3.0
   mild               5.0     3.0
   cool               4.0     2.0
   [total]           12.0     8.0

humidity
   high               4.0     5.0
   normal             7.0     2.0
   [total]           11.0     7.0

windy
   TRUE               4.0     4.0
   FALSE              7.0     3.0
   [total]           11.0     7.0

>
```

```python
# instance (row)
instances.get_instance(0)

sunny,hot,high,FALSE,no

# prediction on instance
cls.classify_instance( instances.get_instance(0) )

# no

1.0

# distribution of classes on the instance
cls.distribution_for_instance( instances.get_instance(0) )

# prob[class = 0] = 0 and prob[class = 1] = 1
# recall 0 = yes and 1 = no

array([0.2957534, 0.7042466])
```

4.3 Divide & Conquer: Decision Trees

Sources

- ID3 by Weka
- Decision Tree by Scikit-Learn (not ID3)

```python
# weather nominal dataset
instances = loader.load_file("data/weather.nominal.arff")
instances

@relation weather.symbolic

@attribute outlook {sunny,overcast,rainy}
@attribute temperature {hot,mild,cool}
@attribute humidity {high,normal}
@attribute windy {TRUE,FALSE}
@attribute play {yes,no}

@data
sunny,hot,high,FALSE,no
sunny,hot,high,TRUE,no
overcast,hot,high,FALSE,yes
rainy,mild,high,FALSE,yes
rainy,cool,normal,FALSE,yes
rainy,cool,normal,TRUE,no
overcast,cool,normal,TRUE,yes
sunny,mild,high,FALSE,no
sunny,cool,normal,FALSE,yes
rainy,mild,normal,FALSE,yes
sunny,mild,normal,TRUE,yes
overcast,mild,high,TRUE,yes
overcast,hot,normal,FALSE,yes
rainy,mild,high,TRUE,no

# classes to indices
# yes, no = 0, 1

# set class attribute to the last attribute, i.e attribute "play"
instances.class_index = instances.num_attributes - 1

# train model
from weka.classifiers import Classifier
cls = Classifier(classname="weka.classifiers.trees.Id3")
cls.build_classifier(instances)

# model inspection
cls.description
```

<bound method OptionHandler.description of Id3

outlook = sunny
| humidity = high: no
| humidity = normal: yes
outlook = overcast: yes
outlook = rainy

```
|   windy = TRUE: no
|   windy = FALSE: yes>
```

```
# instance (row)
instances.get_instance(0)
```

```
sunny,hot,high,FALSE,no
```

```
# prediction on instance
cls.classify_instance( instances.get_instance(0) )
```

```
# no
```

```
1.0
```

4.4 Covering Algorithms: Constructing Rules

Sources

- [PRISM by Weka](#)

```
# weather nominal dataset
instances = loader.load_file("data/contact-lenses.arff")
instances

@relation contact-lenses

@attribute age {young,pre-presbyopic,presbyopic}
@attribute spectacle-prescrip {myope,hypermetrope}
@attribute astigmatism {no,yes}
@attribute tear-prod-rate {reduced,normal}
@attribute contact-lenses {soft,hard,none}

@data
young,myope,no,reduced,none
young,myope,no,normal,soft
young,myope,yes,reduced,none
young,myope,yes,normal,hard
young,hypermetrope,no,reduced,none
young,hypermetrope,no,normal,soft
young,hypermetrope,yes,reduced,none
young,hypermetrope,yes,normal,hard
pre-presbyopic,myope,no,reduced,none
pre-presbyopic,myope,no,normal,soft
pre-presbyopic,myope,yes,reduced,none
pre-presbyopic,myope,yes,normal,hard
pre-presbyopic,hypermetrope,no,reduced,none
pre-presbyopic,hypermetrope,no,normal,soft
pre-presbyopic,hypermetrope,yes,reduced,none
```

```
pre-presbyopic,hypermetrope,yes,normal,none
presbyopic,myope,no,reduced,none
presbyopic,myope,no,normal,none
presbyopic,myope,yes,reduced,none
presbyopic,myope,yes,normal,hard
presbyopic,hypermetrope,no,reduced,none
presbyopic,hypermetrope,no,normal,soft
presbyopic,hypermetrope,yes,reduced,none
presbyopic,hypermetrope,yes,normal,none

# classes to indices
# soft, hard, none = 0 ,1, 2

# set class attribute to the last attribute, i.e attribute "contact-
lenses"
instances.class_index = instances.num_attributes - 1

# train model
from weka.classifiers import Classifier
cls = Classifier(classname="weka.classifiers.rules.Prism")
cls.build_classifier(instances)

# model inspection
cls.description

<bound method OptionHandler.description of Prism rules
----------
If astigmatism = no
    and tear-prod-rate = normal
    and spectacle-prescrip = hypermetrope then soft
If astigmatism = no
    and tear-prod-rate = normal
    and age = young then soft
If age = pre-presbyopic
    and astigmatism = no
    and tear-prod-rate = normal then soft
If astigmatism = yes
    and tear-prod-rate = normal
    and spectacle-prescrip = myope then hard
If age = young
    and astigmatism = yes
    and tear-prod-rate = normal then hard
If tear-prod-rate = reduced then none
If age = presbyopic
    and tear-prod-rate = normal
    and spectacle-prescrip = myope
    and astigmatism = no then none
If spectacle-prescrip = hypermetrope
    and astigmatism = yes
    and age = pre-presbyopic then none
```

```
If age = presbyopic
    and spectacle-prescrip = hypermetrope
    and astigmatism = yes then none
>

# instance (row)
instances.get_instance(0)

# age=young, spectacle-prescrip=myope, astigmatism=no, tear-prod-
rate=reduced, contact-lenses=:none
# by rule: If tear-prod-rate = reduced then none

young,myope,no,reduced,none

# prediction on instance
cls.classify_instance( instances.get_instance(0) )

# none

2.0
```

4.5 Mining Association Rules

Sources

- [Apriori by Weka](#)

```
# weather nominal dataset
instances = loader.load_file("data/weather.nominal.arff")
instances

@relation weather.symbolic

@attribute outlook {sunny,overcast,rainy}
@attribute temperature {hot,mild,cool}
@attribute humidity {high,normal}
@attribute windy {TRUE,FALSE}
@attribute play {yes,no}

@data
sunny,hot,high,FALSE,no
sunny,hot,high,TRUE,no
overcast,hot,high,FALSE,yes
rainy,mild,high,FALSE,yes
rainy,cool,normal,FALSE,yes
rainy,cool,normal,TRUE,no
overcast,cool,normal,TRUE,yes
sunny,mild,high,FALSE,no
sunny,cool,normal,FALSE,yes
```

```
rainy,mild,normal,FALSE,yes
sunny,mild,normal,TRUE,yes
overcast,mild,high,TRUE,yes
overcast,hot,normal,FALSE,yes
rainy,mild,high,TRUE,no

# classes to indices
# yes, no = 0, 1

# set class attribute to the last attribute, i.e attribute "play"
instances.class_index = instances.num_attributes - 1

# train model
from weka.associations import Associator
associator = Associator(classname="weka.associations.Apriori")
associator.build_associations(instances)

# model inspection
associator.description
```

<bound method OptionHandler.description of
Apriori
=======

Minimum support: 0.15 (2 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 17

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12

Size of set of large itemsets L(2): 47

Size of set of large itemsets L(3): 39

Size of set of large itemsets L(4): 6

Best rules found:

 1. outlook=overcast 4 ==> play=yes 4   <conf:(1)> lift:(1.56) lev:
(0.1) [1] conv:(1.43)
 2. temperature=cool 4 ==> humidity=normal 4   <conf:(1)> lift:(2)
lev:(0.14) [2] conv:(2)
 3. humidity=normal windy=FALSE 4 ==> play=yes 4   <conf:(1)> lift:
(1.56) lev:(0.1) [1] conv:(1.43)
 4. outlook=sunny play=no 3 ==> humidity=high 3   <conf:(1)> lift:(2)
lev:(0.11) [1] conv:(1.5)
 5. outlook=sunny humidity=high 3 ==> play=no 3   <conf:(1)> lift:
(2.8) lev:(0.14) [1] conv:(1.93)
 6. outlook=rainy play=yes 3 ==> windy=FALSE 3   <conf:(1)> lift:

```
(1.75) lev:(0.09) [1] conv:(1.29)
 7. outlook=rainy windy=FALSE 3 ==> play=yes 3     <conf:(1)> lift:
(1.56) lev:(0.08) [1] conv:(1.07)
 8. temperature=cool play=yes 3 ==> humidity=normal 3     <conf:(1)>
lift:(2) lev:(0.11) [1] conv:(1.5)
 9. outlook=sunny temperature=hot 2 ==> humidity=high 2     <conf:(1)>
lift:(2) lev:(0.07) [1] conv:(1)
10. temperature=hot play=no 2 ==> outlook=sunny 2     <conf:(1)> lift:
(2.8) lev:(0.09) [1] conv:(1.29)
>

# instance (row)
instances.get_instance(0)

# age=young, spectacle-prescrip=myope, astigmatism=no, tear-prod-
rate=reduced, contact-lenses=:none
# by rule: If tear-prod-rate = reduced then none

sunny,hot,high,FALSE,no

# From `help(associator)` it seems we cannot infer based on a given
instance
```

4.6 Linear Models

Linear Regression

```
df = pd.read_csv("data/cpu.csv")
df

      MYCT   MMIN    MMAX   CACH   CHMIN   CHMAX   class
0      125    256    6000    256      16     128     198
1       29   8000   32000     32       8      32     269
2       29   8000   32000     32       8      32     220
3       29   8000   32000     32       8      32     172
4       29   8000   16000     32       8      16     132
..     ...    ...     ...    ...     ...     ...     ...
204    124   1000    8000      0       1       8      42
205     98   1000    8000     32       2       8      46
206    125   2000    8000      0       2      14      52
207    480    512    8000     32       0       0      67
208    480   1000    4000      0       0       0      45

[209 rows x 7 columns]

from sklearn import linear_model

reg = linear_model.LinearRegression()
```
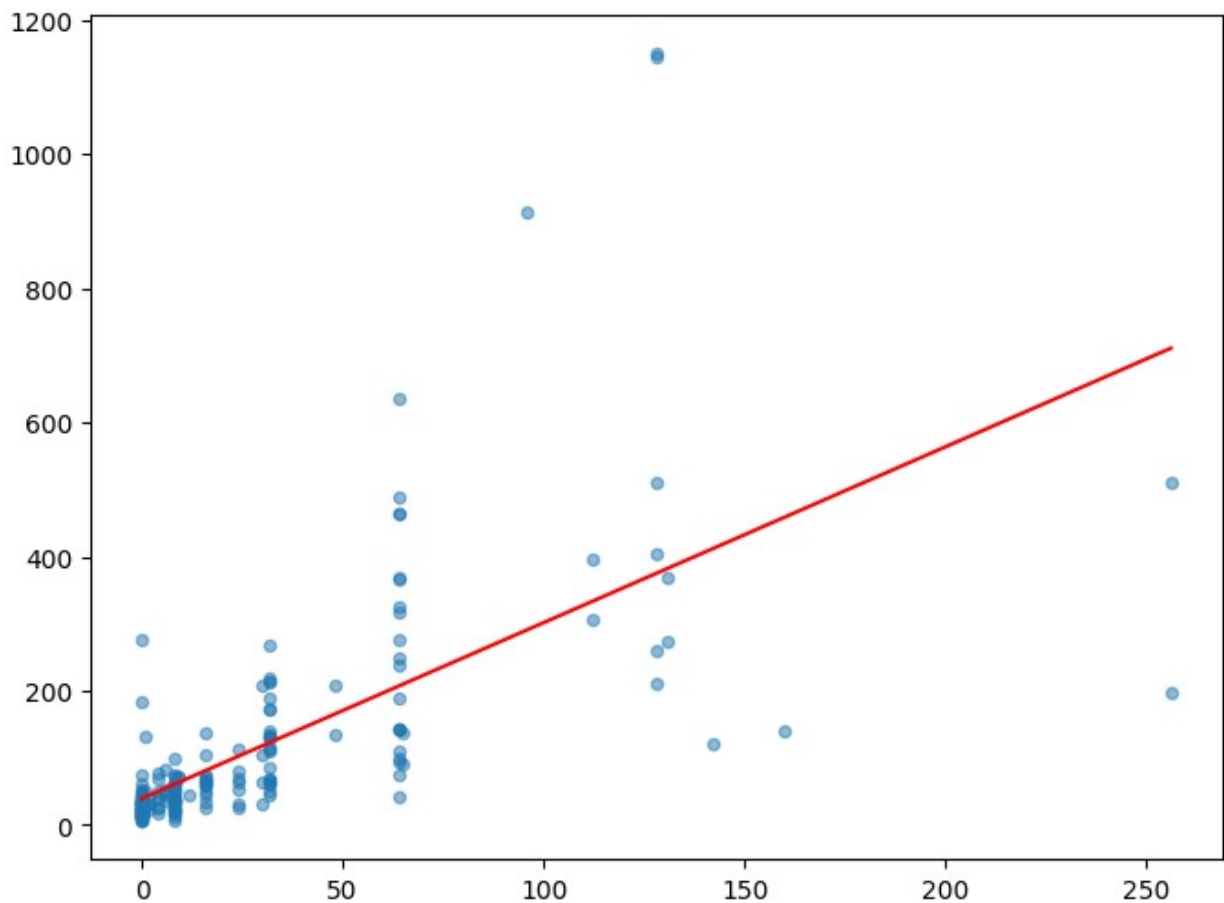
```
reg.fit( df[["CACH"]], y = df["class"] )

reg.coef_, reg.intercept_

(array([2.62309769]), 39.50488702952755)

plt.figure(figsize=(8, 6))
plt.plot( df[["CACH"]], reg.predict(df[["CACH"]]), c='r' )
plt.scatter( df[["CACH"]], y=df["class"], s=20, alpha=0.5 )

<matplotlib.collections.PathCollection at 0x7f5f39fcfb80>
```


Logistic Regression

```
# Dummy Dataframe
#####

df = pd.DataFrame([
    [-2,0],
    [-1,0],
    [0,1],
    [1,1],
```

```
        [2,1],
    ],
    columns = ["X", "class"]
)
df

     X  class
0 -2      0
1 -1      0
2  0      1
3  1      1
4  2      1

# Fit the model

from sklearn.linear_model import LogisticRegression

clf = LogisticRegression()
clf.fit(df[["X"]].to_numpy(), df["class"].to_numpy())

LogisticRegression()

# Visualize the logit function

# Define the logistic function
def logistic_function(x, coef, intercept):
    # dot product + intercept
    logit = np.dot(coef, x) + intercept
    return 1 / (1 + np.exp(-logit))

# Define the range of x values for the plot
X_tem = np.linspace(-4, 4, 100)

Y_tem = logistic_function(X_tem, clf.coef_[0][0], clf.intercept_[0])

# Create a plot
plt.figure(figsize=(8, 6))
plt.plot(X_tem, Y_tem)

[<matplotlib.lines.Line2D at 0x7f7e01ceb010>]
```
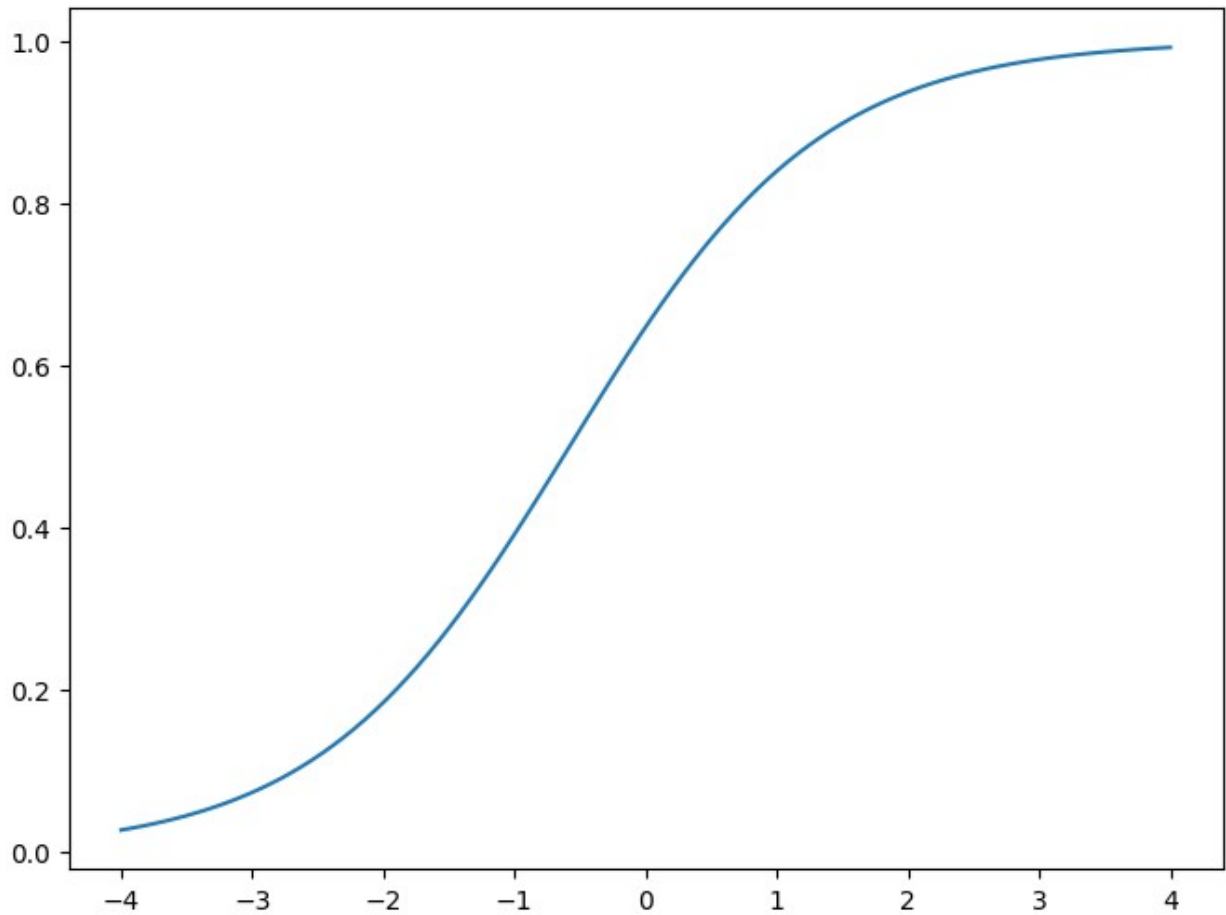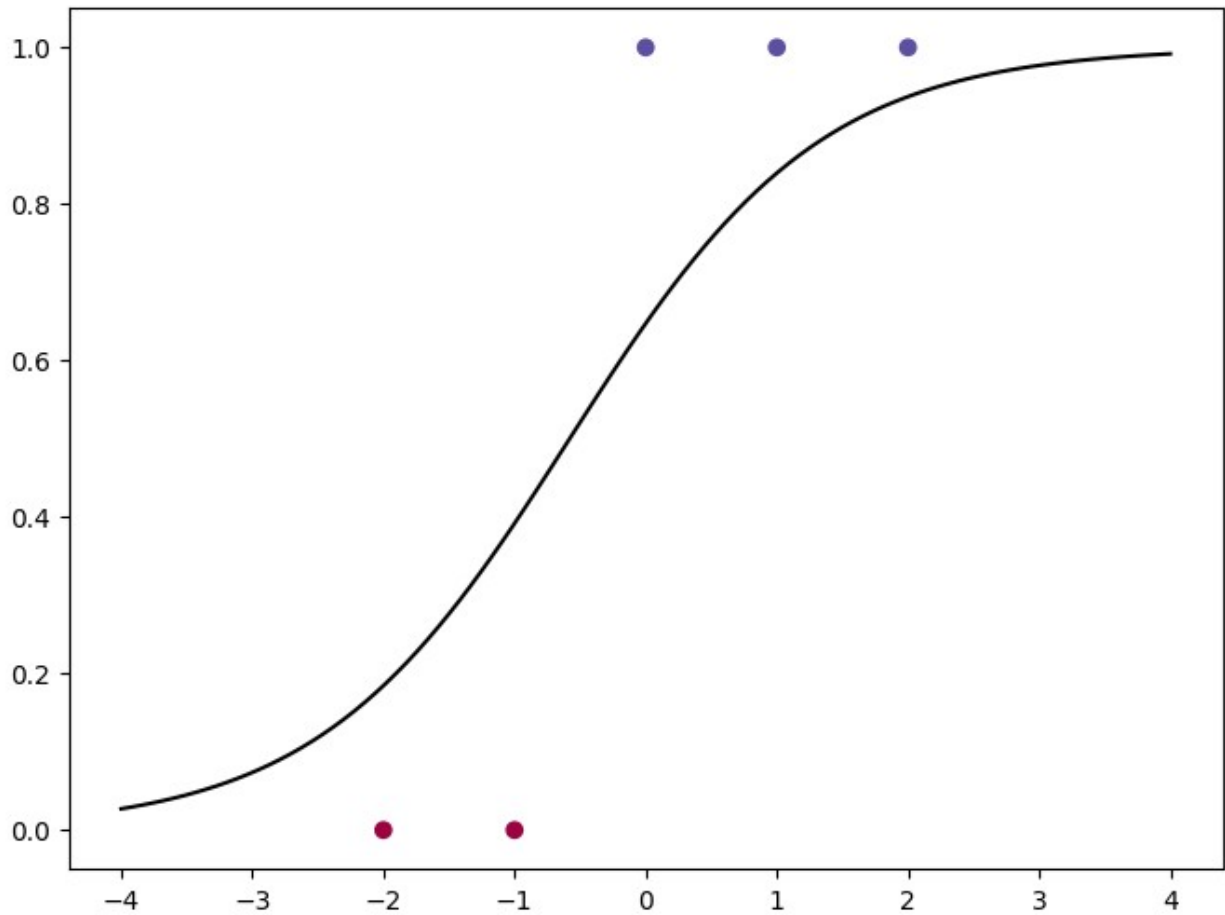
```
plt.figure(figsize=(8, 6))
plt.scatter( df["X"], y = df["class"], c=df["class"],
cmap=plt.cm.Spectral )
plt.plot(X_tem, Y_tem, c="black")
```

```
[<matplotlib.lines.Line2D at 0x7f7e0161a3e0>]
```

```
# predict an unseen instance
clf.predict([ [-4] ])

array([0])
```

```
# probabilities of classes
clf.predict_proba([
    [0.5]
])

array([[0.24404677, 0.75595323]])
```

```
clf.predict_proba([
    [-1]
])

array([[0.60821662, 0.39178338]])
```

4.7 Instance-Based Learning (in development)

```python
# kd-tree
#
https://weka.sourceforge.io/doc.dev/weka/core/neighboursearch/KDTree.h
tml
```

4.8 Clustering

```python
# Dummy Dataframe


# Set a random seed for reproducibility
#np.random.seed(0)

# Number of data points for each class
numSamples = 100

# dataset of class 0
df_tem0 = pd.DataFrame({
    "X": np.random.randn(numSamples) + 2, "Y":
np.random.randn(numSamples) + 2
  })

# dataset of class 1
df_tem1 = pd.DataFrame({
    "X": np.random.randn(numSamples) - 2, "Y":
np.random.randn(numSamples) - 2
  })

# dataframe, concatenates both
df = pd.concat([df_tem0, df_tem1])

# Plot the generated data
plt.figure(figsize=(8, 6))
plt.scatter( df["X"], y = df["Y"] )
plt.xlabel("X"); plt.ylabel("Y")
# plt.show()

Text(0, 0.5, 'Y')
```

```
# Fit K-Means

from sklearn.cluster import KMeans
kmeans = KMeans(n_clusters=2, random_state=0,
n_init="auto").fit(df[["X","Y"]])

#kmeans.predict([[0, 0], [12, 3]])
#kmeans.cluster_centers_

array([1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1,
1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
```
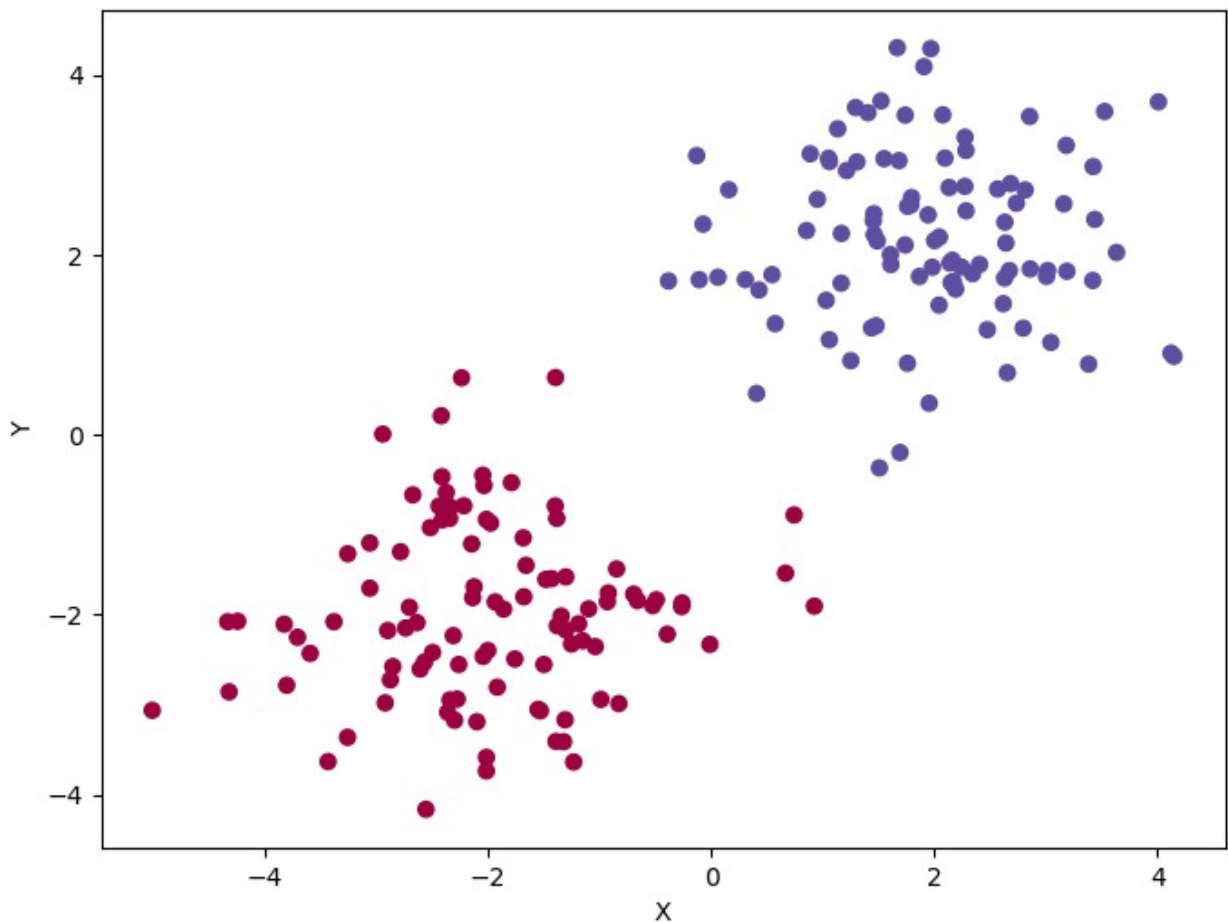
```
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
        0, 0], dtype=int32)
```

```python
# Visualize clusters
plt.figure(figsize=(8, 6))
plt.scatter( df["X"], y = df["Y"], c=kmeans.labels_,
cmap=plt.cm.Spectral )
plt.xlabel("X"); plt.ylabel("Y")
```

```
Text(0, 0.5, 'Y')
```


Lab Tasks

- Construct a dummy dataframe using randomization as we illustrated, or alternatively by the following snippet:

```
pd.DataFrame([
    [1,1],
    [1,2],
    [4,4],
    [5,6]
  ],
  columns=["X","Y"]
)
```

- Select an algorithm.
- Experiment with different inputs, comparing and contrasting outputs and visualizations.
- Decide on an output pattern, then construct a dataframe conforming to it.

P.S. We delegate to students to express what they wish to learn the most, and accordingly solve the lab task.