Chapter Goals

- Evaluation metrics, their interpretation, and use-cases.

Lab Goals

- Computing, Visualizing, and contrasting metrics.
- Tuning models by them is left to students as an exercise.

General Guidlines

- Visualization.
- Modifiable code snippets.

```
# 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 19 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 5s (1,177 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/libtbb.so.12 is not a symbolic
link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind_2_0.so.3 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 19 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 3s (931 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/libtbb.so.12 is not a symbolic
link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind_2_0.so.3 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 2.6 MB/s eta
0:00:00
etadata (setup.py) ... e=pygraphviz-1.11-cp310-cp310-linux_x86_64.whl
size=175924
sha256=1bc630eb972d7350e3d3c7c7057c0916af680130825c6cd1c7d3ba195254898
3
  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 14.9 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=1743149
sha256=a16a8415f7ad60e2319d93e253f6a5e14562fd57955f39c886b92b4d925eb3b
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 689.9 kB/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=3bd69dc038b82e16729ca7e7a4b8572a09bcb25c863405f3407d99f2d1cb7a2
4
  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=008f323980bf14ab07642468a4cb225271774802b0816cc4790ac2e18fcb902
9
  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=85ef68ae1642350715fd41ffd17ecc63ddb2b4c8c76eebd0f13a2e89ab5ab85
0
  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 1.4 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)
  Downloading sklearn-0.0.post10.tar.gz (3.6 kB)
  Preparing metadata (setup.py) ... ent already satisfied: python-
javabridge>=4.0.0 in /usr/local/lib/python3.10/dist-packages (from
python-weka-wrapper3>=0.2.5->sklearn-weka-plugin) (4.0.3)
Requirement already satisfied: packaging in
/usr/local/lib/python3.10/dist-packages (from python-weka-
wrapper3>=0.2.5->sklearn-weka-plugin) (23.2)
Requirement already satisfied: configurable-objects in
/usr/local/lib/python3.10/dist-packages (from python-weka-
wrapper3>=0.2.5->sklearn-weka-plugin) (0.0.1)
Requirement already satisfied: simple-data-flow in
/usr/local/lib/python3.10/dist-packages (from python-weka-
wrapper3>=0.2.5->sklearn-weka-plugin) (0.0.1)
Building wheels for collected packages: sklearn-weka-plugin, sklearn
  Building wheel for sklearn-weka-plugin (setup.py) ...
e=sklearn_weka_plugin-0.0.7-py3-none-any.whl size=27346
sha256=a7810c5be1c7b0f4bc958e0ddc91a12cbe401a2b26b523ad56d54ac7549a19a
2
  Stored in directory:
/root/.cache/pip/wheels/51/6d/e5/458ea9a1be729f39ed4cf14aab2f87eb51470
47b690402605b
  Building wheel for sklearn (setup.py) ... e=sklearn-0.0.post10-py3-
none-any.whl size=2959
sha256=82d661a1a691ac82003e7a95eabbad6008a58cd6643dbd1a32997b3a31a54f4
f
  Stored in directory:
/root/.cache/pip/wheels/5b/f6/92/0173054cc528db7ffe7b0c7652a96c3102aab
156a6da960387
Successfully built sklearn-weka-plugin sklearn
Installing collected packages: sklearn, sklearn-weka-plugin
Successfully installed sklearn-0.0.post10 sklearn-weka-plugin-0.0.7
```

```
# @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

# @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/weka.jar',
'/usr/local/lib/python3.10/dist-packages/weka/lib/core.jar',
'/usr/local/lib/python3.10/dist-packages/weka/lib/mtj.jar',
'/usr/local/lib/python3.10/dist-packages/weka/lib/python-weka-
wrapper.jar',
'/usr/local/lib/python3.10/dist-packages/weka/lib/arpack_combined.jar'
]
DEBUG:weka.core.jvm:MaxHeapSize=default
DEBUG:weka.core.jvm:Package support enabled
```

```
# @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
```

```
--2023-10-29 12:57:22--
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

2023-10-29 12:57:24 (1.42 GB/s) - ‘data/airline.arff’ saved
[2357/2357]

--2023-10-29 12:57:24--
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   181KB/s    in
0.2s

2023-10-29 12:57:24 (181 KB/s) - ‘data/breast-cancer.arff’ saved
[29418/29418]

--2023-10-29 12:57:24--
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

2023-10-29 12:57:25 (938 MB/s) - ‘data/contact-lenses.arff’ saved
[2890/2890]
```

```
--2023-10-29 12:57:25--
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

2023-10-29 12:57:26 (93.8 MB/s) - ‘data/cpu.arff’ saved [5561/5561]

--2023-10-29 12:57:26--
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

2023-10-29 12:57:27 (45.8 MB/s) - ‘data/cpu.with.vendor.arff’ saved
[6960/6960]

--2023-10-29 12:57:27--
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   328KB/s    in
0.5s

2023-10-29 12:57:28 (328 KB/s) - ‘data/credit-g.arff’ saved
[162270/162270]
```

```
--2023-10-29 12:57:28--
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   238KB/s    in
0.2s

2023-10-29 12:57:29 (238 KB/s) - ‘data/diabetes.arff’ saved
[37443/37443]

--2023-10-29 12:57:29--
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   110KB/s    in
0.2s

2023-10-29 12:57:30 (110 KB/s) - ‘data/glass.arff’ saved [17850/17850]

--2023-10-29 12:57:30--
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   477KB/s    in
0.6s

2023-10-29 12:57:31 (477 KB/s) - ‘data/hypothyroid.arff’ saved
[310897/310897]
```

```
--2023-10-29 12:57:31--
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   166KB/s    in
0.5s

2023-10-29 12:57:33 (166 KB/s) - ‘data/ionosphere.arff’ saved
[80487/80487]

--2023-10-29 12:57:33--
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

2023-10-29 12:57:33 (51.3 MB/s) - ‘data/iris.2D.arff’ saved
[3492/3492]

--2023-10-29 12:57:33--
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

2023-10-29 12:57:34 (96.3 MB/s) - ‘data/iris.arff’ saved [7486/7486]
```

```
--2023-10-29 12:57:34--
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

2023-10-29 12:57:35 (91.8 MB/s) - ‘data/labor.arff’ saved [8279/8279]

--2023-10-29 12:57:35--
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   308KB/s    in
0.6s

2023-10-29 12:57:36 (308 KB/s) - ‘data/segment-challenge.arff’ saved
[200410/200410]

--2023-10-29 12:57:37--
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   225KB/s    in
0.5s

2023-10-29 12:57:38 (225 KB/s) - ‘data/segment-test.arff’ saved
[109984/109984]
```

```
--2023-10-29 12:57:38--
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   412KB/s    in
0.5s

2023-10-29 12:57:39 (412 KB/s) - ‘data/soybean.arff’ saved
[202935/202935]

--2023-10-29 12:57:39--
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  1.73MB/s    in
1.1s

2023-10-29 12:57:41 (1.73 MB/s) - ‘data/supermarket.arff’ saved
[2025871/2025871]

--2023-10-29 12:57:41--
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   286KB/s    in
0.6s

2023-10-29 12:57:43 (286 KB/s) - ‘data/unbalanced.arff’ saved
```

```
[186360/186360]

--2023-10-29 12:57:43--
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   124KB/s    in
0.3s

2023-10-29 12:57:44 (124 KB/s) - ‘data/vote.arff’ saved [40261/40261]

--2023-10-29 12:57:44--
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

2023-10-29 12:57:44 (507 MB/s) - ‘data/weather.nominal.arff’ saved
[587/587]

--2023-10-29 12:57:44--
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

2023-10-29 12:57:45 (342 MB/s) - ‘data/weather.numeric.arff’ saved
```

```
[495/495]

--2023-10-29 12:57:46--  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, 185.199.109.133, 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
0.006s

2023-10-29 12:57:47 (3.96 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")
```

5.1 Training and Testing

```
# train test split
# accuracy score

# modules
import numpy as np
from sklearn.datasets import load_iris
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score

# Load the Iris dataset
data = load_iris()
X = data.data
y = data.target

# Split the dataset into a training set and a testing set
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.5, random_state=42)

# Create a decision tree classifier
clf = DecisionTreeClassifier()
```

```
# Train the classifier on the training data
clf.fit(X_train, y_train)

# Make predictions on the test data
y_pred = clf.predict(X_test)

# Measure the accuracy of the model
accuracy = accuracy_score(y_test, y_pred)
print(f"Accuracy: {accuracy:.2f}")
```

```
Accuracy: 0.95
```

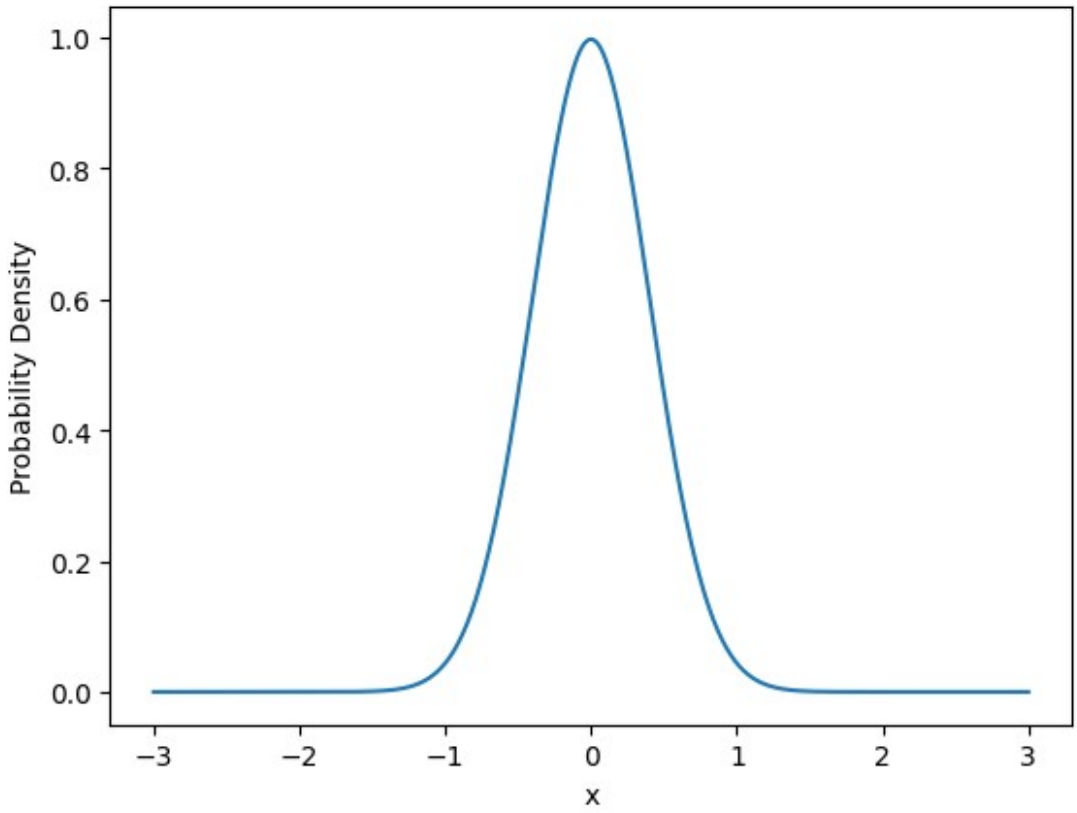
Task 1.1 Experiment by modifying the test-train size, Hypothesize and explain observations.

Task 1.2 What is the best test-train size?

5.2 Predicting Performance

```
# normal distribution

# modules
import scipy.stats as stats
import numpy as np
import matplotlib.pyplot as plt

# Define the parameters of the normal distribution (mean and standard
deviation)
mean = 0.0
std_dev = 0.4

# Create an array of x values for the plot
x_values = np.linspace(-3, 3, 1000)

# Calculate the corresponding probability density function (PDF)
values
pdf_values = stats.norm.pdf(x_values, loc=mean, scale=std_dev)

# Create a figure and axis for the plot
fig, ax = plt.subplots()

# Plot the PDF curve
ax.plot(x_values, pdf_values, label="PDF")

# Set labels and a legend
ax.set_xlabel("x")
ax.set_ylabel("Probability Density")
```

```
# Display the plot
plt.show()
```

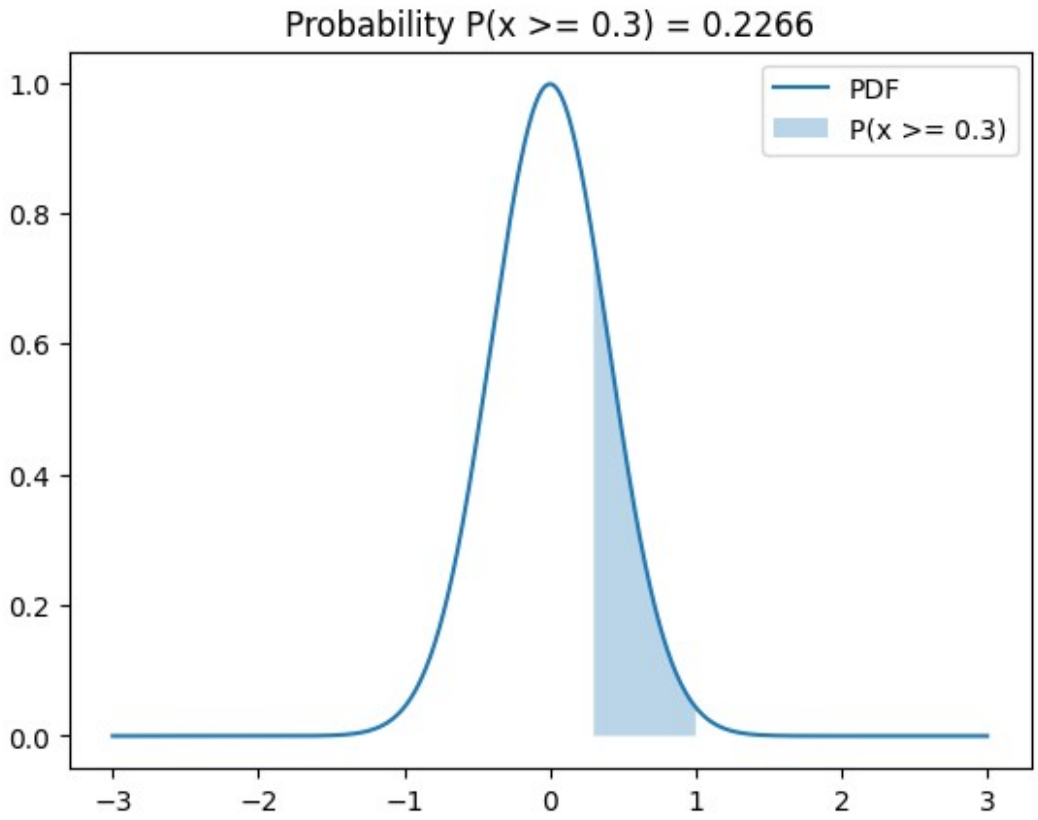



```
# probability on a distribution

# Calculate the probability of x >= 0.3
probability = 1 - stats.norm.cdf(0.3, loc=mean, scale=std_dev)

# Print the result
print(f"Probability(x >= 0.3) = {probability:.4f}")
```

```
Probability(x >= 0.3) = 0.2266
```

```
# Area under the curve filled

# Create a figure and axis for the plot
fig, ax = plt.subplots()

# Plot the PDF curve
ax.plot(x_values, pdf_values, label="PDF")

# Fill the area under the curve for x >= 0.3
x_fill = np.linspace(0.3, 1, 1000)
pdf_fill = stats.norm.pdf(x_fill, loc=mean, scale=std_dev)
```

```
ax.fill_between(x_fill, pdf_fill, alpha=0.3, label="P(x >= 0.3)")

ax.set_title(f"Probability P(x >= 0.3) = {probability:.4f}")
ax.legend()

# Display the plot
plt.show()
```



Task 2.1 Modify parameters of the normal distribution and observe the corresponding graph, and probability of the interval.

5.3 Cross-Validation

```
# cross-validation

# modules
import numpy as np
from sklearn.datasets import load_iris
from sklearn.model_selection import cross_val_score
from sklearn.tree import DecisionTreeClassifier

# Load the Iris dataset
data = load_iris()
```

```
X = data.data
y = data.target

# Create a decision tree classifier
clf = DecisionTreeClassifier()

# Perform 5-fold cross-validation and calculate the accuracy scores
scores = cross_val_score(clf, X, y, cv=5)

# Print the accuracy scores for each fold and the mean accuracy
for fold, score in enumerate(scores, start=1):
    print(f"Fold {fold}: Accuracy = {score:.2f}")

mean_accuracy = np.mean(scores)
print(f"Mean Accuracy: {mean_accuracy:.2f}")
```

```
Fold 1: Accuracy = 0.97
Fold 2: Accuracy = 0.97
Fold 3: Accuracy = 0.90
Fold 4: Accuracy = 0.97
Fold 5: Accuracy = 1.00
Mean Accuracy: 0.96
```

Task 2.1 Experiment by modifying folds size, Hypothesize and explain observations.

Task 2.2 What is the distinguishing feature of cross validation?

5.4 Other Estimators

Leave One Out

```
# leave one out

from sklearn.model_selection import LeaveOneOut

X = [1, 2, 3, 4]
loo = LeaveOneOut()
for train, test in loo.split(X):
    print("%s %s" % (train, test))
```

```
[1 2 3] [0]
[0 2 3] [1]
[0 1 3] [2]
[0 1 2] [3]
```

```
from sklearn.datasets import load_iris
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score
```

```
# Load the Iris dataset
data = load_iris()
X = data.data
y = data.target

# Create a decision tree classifier
clf = DecisionTreeClassifier()

# Initialize Leave-One-Out cross-validator
loo = LeaveOneOut()

# Initialize variables to keep track of total accuracy and the number
of iterations
total_accuracy = 0
num_iterations = 0

# Perform Leave-One-Out cross-validation
for train_index, test_index in loo.split(X):
    X_train, X_test = X[train_index], X[test_index]
    y_train, y_test = y[train_index], y[test_index]

    # Train the classifier on the training data
    clf.fit(X_train, y_train)

    # Make predictions on the test data
    y_pred = clf.predict(X_test)

    # Measure the accuracy for this iteration
    accuracy = accuracy_score(y_test, y_pred)

    # Update the total accuracy and the number of iterations
    total_accuracy += accuracy
    num_iterations += 1

# Calculate the mean accuracy over all iterations
mean_accuracy = total_accuracy / num_iterations

print(f"Mean Accuracy: {mean_accuracy:.2f}")
```

```
Mean Accuracy: 0.94
```

Task 4.1 In which case leave-one-out is most useful?

Bootstrap

```
# Mean

import numpy as np

# Original dataset
data = np.array([15, 20, 21, 22, 24, 25, 28, 30, 31, 35, 40, 42, 45,
```

```
50])

# Number of bootstraps
num_bootstraps = 10

# Initialize an array to store bootstrapped sample means
bootstrapped_means = np.zeros(num_bootstraps)

# Perform bootstrapping
for i in range(num_bootstraps):
    # Generate a random sample with replacement from the original data
    bootstrap_sample = np.random.choice(data, size=len(data),
replace=True)

    # Calculate the mean of the bootstrapped sample
    bootstrapped_means[i] = np.mean(bootstrap_sample)

# Print the results
print(f"Original Data Mean: {np.mean(data):.2f}")
print(f"Bootstrap Mean: {np.mean(bootstrapped_means):.2f}")
```

```
Original Data Mean: 30.57
Bootstrap Mean: 29.56
```

```
# Linear Regression

import numpy as np
from sklearn.linear_model import LinearRegression

# Generate some synthetic data for demonstration
np.random.seed(0)
X = np.random.rand(100, 1)
y = 2 * X + 1 + 0.1 * np.random.randn(100, 1)

# Number of bootstraps
num_bootstraps = 10

# Initialize arrays to store bootstrapped regression coefficients
bootstrapped_coeffs = np.zeros((num_bootstraps, 2))  # Two
coefficients: intercept and slope

# Perform bootstrapping
for i in range(num_bootstraps):
    # Generate a random sample with replacement
    indices = np.random.choice(len(X), size=len(X), replace=True)
    X_bootstrap = X[indices]
    y_bootstrap = y[indices]

    # Fit a linear regression model to the bootstrapped sample
    model = LinearRegression()
    model.fit(X_bootstrap, y_bootstrap)
```

```
    # Store the intercept and coefficient (slope)
    bootstrapped_coeffs[i, 0] = model.intercept_
    bootstrapped_coeffs[i, 1] = model.coef_

# Calculate the confidence intervals for the intercept and coefficient
(slope)
confidence_intervals = np.percentile(bootstrapped_coeffs, [2.5, 97.5],
axis=0)

# Print the results
print("Original Regression Coefficients:")
print(f"Intercept: {1:.2f}") # hard-coded in equation above
print(f"Slope: {2:.2f}")

print("\nBootstrap Results:")
print(f"Intercept 95% Confidence Interval: ({confidence_intervals[0,
0]:.2f}, {confidence_intervals[1, 0]:.2f})")
print(f"Slope 95% Confidence Interval: ({confidence_intervals[0,
1]:.2f}, {confidence_intervals[1, 1]:.2f})")
```

```
Original Regression Coefficients:
Intercept: 1.00
Slope: 2.00

Bootstrap Results:
Intercept 95% Confidence Interval: (1.00, 1.06)
Slope 95% Confidence Interval: (1.94, 2.05)
```

Task 4.2 Construct a dummy data with high variance (i.e many outliers) and apply the same pipeline. In light of results, What is the value of bootstrap?

5.5 Hyper-parameter Selection

```
# train, validate, test splitting

import numpy as np
from sklearn.model_selection import train_test_split

# Generate synthetic data for demonstration
np.random.seed(0)
X = np.random.rand(100, 2)
y = np.random.randint(0, 2, 100)

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)
# Split training data into train and validation sets
X_train, X_val, y_train, y_val = train_test_split(X_train, y_train,
```

```
test_size=0.25, random_state=42)

# Print the sizes of the sets
print(f"Training Set Size: {len(X_train)}")
print(f"Validation Set Size: {len(X_val)}")
print(f"Testing Set Size: {len(X_test)}")
```

```
Training Set Size: 60
Validation Set Size: 20
Testing Set Size: 20
```

Task 5.1 Select a model of your choice. Tune it by train-validate-test splitting procedure above. Compare it with tuning only by train-test splitting. Compare accuracies.

Task 5.2 Experiment and observe whether the above procedure had saved you from over-fitting the model.

5.6 Comparing Data Mining Schemes

```
import numpy as np
from scipy import stats

# Simulated performance metrics for two models (replace with your
actual data)
model1_metrics = np.array([0.85, 0.88, 0.82, 0.90, 0.87])
model2_metrics = np.array([0.78, 0.80, 0.75, 0.79, 0.81])

# Perform a two-sample t-test
t_statistic, p_value = stats.ttest_ind(model1_metrics, model2_metrics)

# Set the significance level
alpha = 0.05

# Check if the p-value is less than the significance level
if p_value < alpha:
    print("Statistically significant difference")
else:
    print("No statistically significant difference")
```

```
Statistically significant difference
```

Task 6.1 Why do we need to use t-test? Why don't we go with the highest model accuracy?

5.7 Predicting Probabilities

```
# Quadratic Loss Function
```

```
# modules
import numpy as np
from sklearn.metrics import mean_squared_error

actual_probabilities = [0.4, 0.4, 0.9]  # Ground truth
predicted_probabilities = [0.65, 0.25, 0.78]  # Model's predictions

mse = mean_squared_error(actual_probabilities,
predicted_probabilities)

print(f"Mean Squared Error: {mse:.2f}")
```

```
Mean Squared Error: 0.03
```

```
# Logistic Loss Function

# modules
from sklearn.metrics import log_loss

y_true = [0, 0, 1, 1] # ground truth
y_pred = [[.9, .1], [.8, .2], [.3, .7], [.01, .99]] # model's
predictions

logloss = log_loss(y_true, y_pred)

print(f"Logistic Loss: {logloss:.2f}")
```

```
Logistic Loss: 0.17
```

Task 7.1 Tune a model based on these metrics. Contrast with your previous approaches.

5.8 Counting The Cost

```
# TP, TN, FP, FN

# modules
from sklearn.metrics import confusion_matrix

# Actual labels (ground truth)
actual_labels = [1, 0, 1, 0, 1, 1, 0, 0, 1, 0]

# Predicted labels by your model
predicted_labels = [1, 1, 1, 0, 1, 0, 0, 1, 1, 0]

# Compute the confusion matrix
confusion = confusion_matrix(actual_labels, predicted_labels)

# Extract TP, TN, FP, FN from the confusion matrix
TP = confusion[1, 1]
```

```
TN = confusion[0, 0]
FP = confusion[0, 1]
FN = confusion[1, 0]

# Print the results
print(f"True Positives (TP): {TP}")
print(f"True Negatives (TN): {TN}")
print(f"False Positives (FP): {FP}")
print(f"False Negatives (FN): {FN}")
```

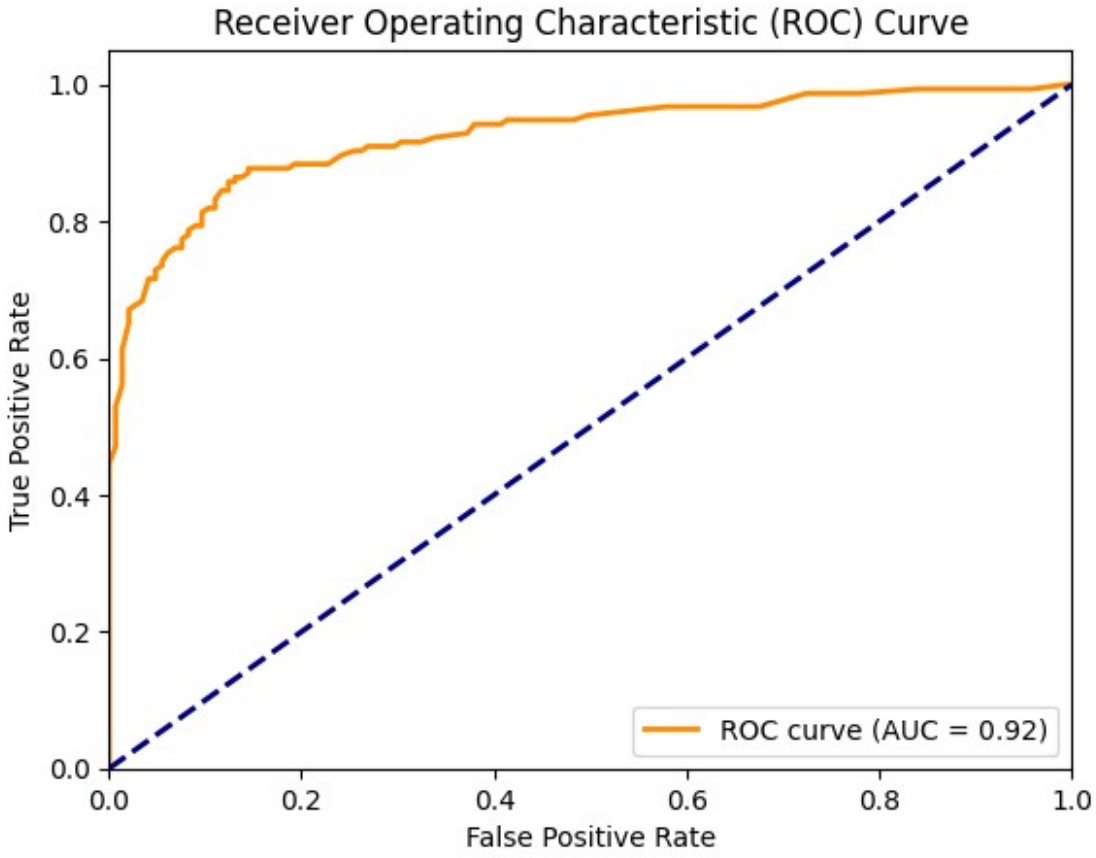
```
True Positives (TP): 4
True Negatives (TN): 3
False Positives (FP): 2
False Negatives (FN): 1
```

Task 8.1 Give a use-case in which these measurements are critical.

Task 8.2 Which of these metrics do you need to critically optimize?

```
# modules
import numpy as np
from sklearn.datasets import make_classification
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import roc_curve, auc
import matplotlib.pyplot as plt

# Generate dummy data
X, y = make_classification(n_samples=1000, n_features=20,
random_state=42)

# Split the data into a training and testing set
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.3, random_state=42)

# Create a random forest classifier
clf = RandomForestClassifier(n_estimators=100, random_state=42)
clf.fit(X_train, y_train)

# Get predicted probabilities for the positive class
y_prob = clf.predict_proba(X_test)[:, 1]

# Compute ROC curve
fpr, tpr, thresholds = roc_curve(y_test, y_prob)

# Calculate the area under the ROC curve (AUC)
roc_auc = auc(fpr, tpr)

# Visualize the ROC curve
```

```
plt.figure()
plt.plot(fpr, tpr, color='darkorange', lw=2, label=f'ROC curve (AUC =
{roc_auc:.2f})')
plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('Receiver Operating Characteristic (ROC) Curve')
plt.legend(loc='lower right')
plt.show()
```



Challenge 8.3 Learn more from here about other methods. Observe there is a whole new world besides accuracy

Challenge 8.4 Learn about regularized learning and use above metrics to do it. See this for example.

5.9 Evaluating Numeric Prediction

```
import numpy as np
from sklearn.metrics import mean_squared_error

# Generate some synthetic data for demonstration
np.random.seed(0)
y_true = np.random.rand(20)  # Actual values
y_pred = y_true + np.random.randn(20) * 0.1  # Predicted values with
some noise

# Calculate Mean Squared Error (MSE)
mse = mean_squared_error(y_true, y_pred)

# Calculate Relative Absolute Error (RAE)
rae = np.mean(np.abs(y_true - y_pred)) / np.mean(np.abs(y_true -
np.mean(y_true)))

# Print the results
print(f"Mean Squared Error (MSE): {mse:.4f}")
print(f"Relative Absolute Error (RAE): {rae:.4f}")

# Compare MSE and RAE
if mse < rae:
    print("MSE is lower, indicating better model performance in terms
of mean squared error.")
elif mse > rae:
    print("RAE is lower, indicating better model performance in terms
of relative absolute error.")
else:
    print("MSE and RAE are equal, and the model's performance is
identical in both metrics.")
```

```
Mean Squared Error (MSE): 0.0141
Relative Absolute Error (RAE): 0.4211
MSE is lower, indicating better model performance in terms of mean
squared error.
```

Challenge 9.1 Check out *table 5.8* in page 195 from the book. Draw cases where one metric is preferable over another.