Machine Learning: Applications and Practices
Lecture 1

Li Chen
University of Louisiana at Lafayette

Welcome!

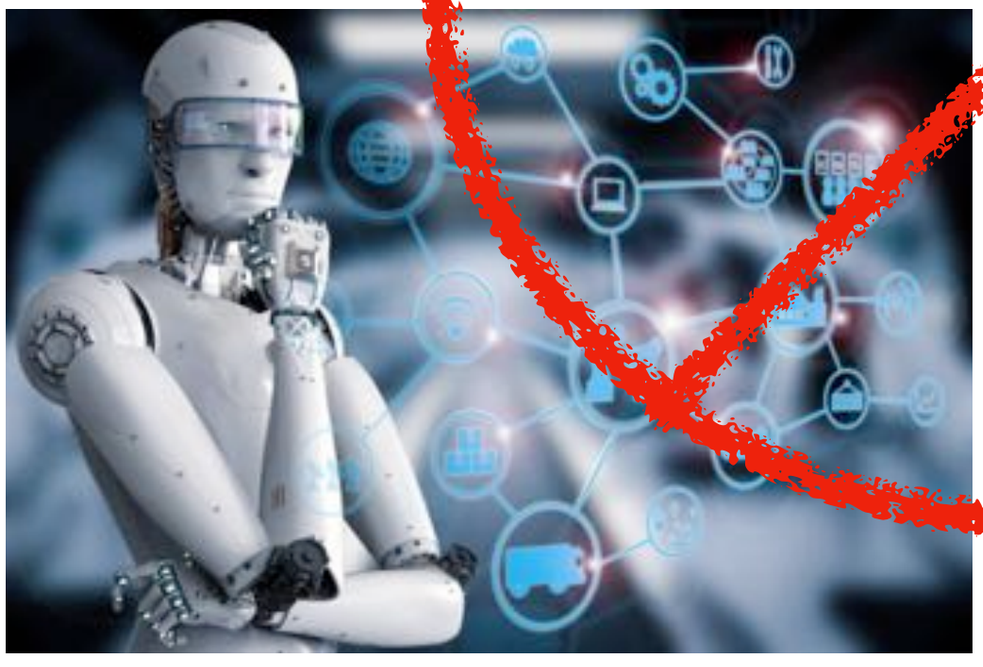
- **Welcome all participants from four universities:**
  - University of Louisiana at Lafayette
  - Southern University
  - University of South Alabama
  - Western Kentucky University
  - Others

Course Information

- **Class Meeting Time:**
  - Wednesday: 10: 00am to 11:15am (Lecture series)
  - Friday: 10: 00am to 11:30am (Hands-on series)
- **Prerequisite:**
  - Have a Windows OS laptop
  - Know the basic of Python programming
- **Course Assistants:**
  - Mr. Yihe Zhang
  - Mr. Jiadong Lou
- **Course Website:**
  - https://people.cmix.louisiana.edu/yuan/2022_Summer_Tutorial_Courses.html
  - *Please don't distribute/spread Twitter Dataset*

What's Our Goals?

We are not ambitious…



Our Goals

This is just an entry level of Machine Learning course!

No credits, no grading!

1. Learning the fundamental knowledge

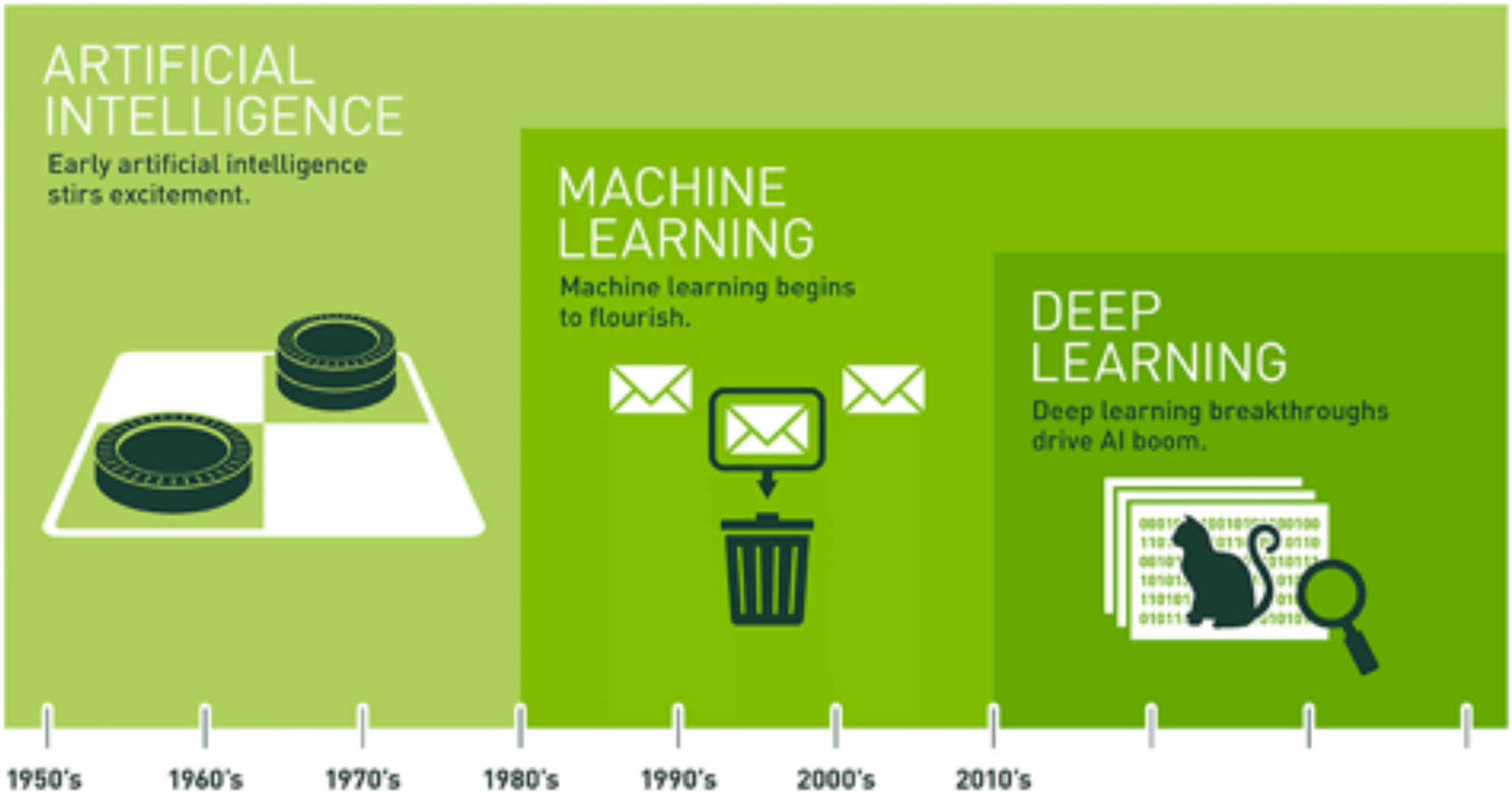
2. Coding practice for Python

3. Practicing on real-world data

My Suggestions

Please attend each lecture and hands-on;
Otherwise, you will be
lost!

AI History



Source from: https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/

AI and ML

- **Artificial Intelligence (AI)**
  - Role of Statistics: Inference from a sample.

- **Machine Learning (ML)**
  - Arthur Samuel (1959): Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed.
  - Tom Mitchell (1998): Well-posed Learning Problem: A computer program is said to learn from experience with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.

What is Machine Learning?

- Study of *Algorithms* that *improve* their *performance* at some *task* with *experience*.

- **Role of Computers:**
  - Having efficient algorithms to solve the optimization problems to learn models
  - Learning Models for unknown and changing worlds
  - Representing and Evaluating the model for inference.

What is Machine Learning?

Experience

This is a “cat”

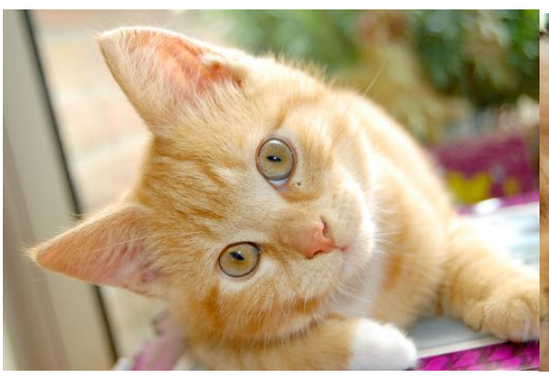
Algorithms

Tasks

Not a “cat”

Not a “cat”

“cat”

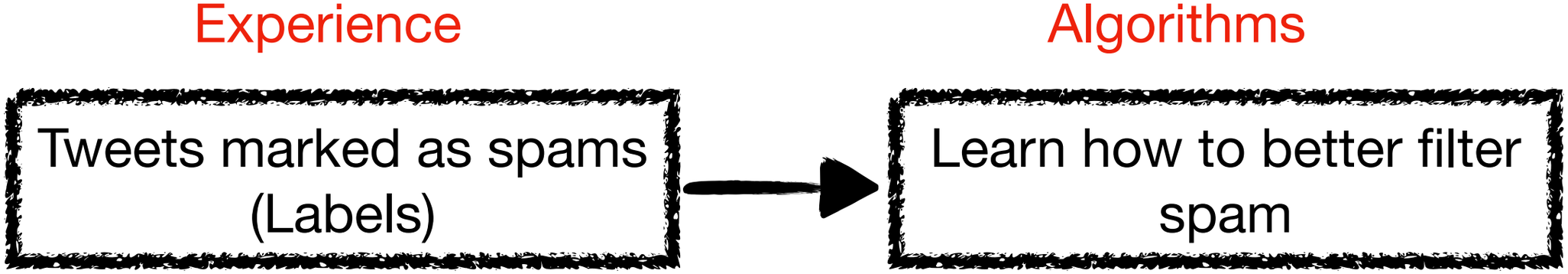


Spam Classification Example

- *Suppose Twitter server watches which tweets marked as spam message. Based on this information, he will learn how to better filter spam.*

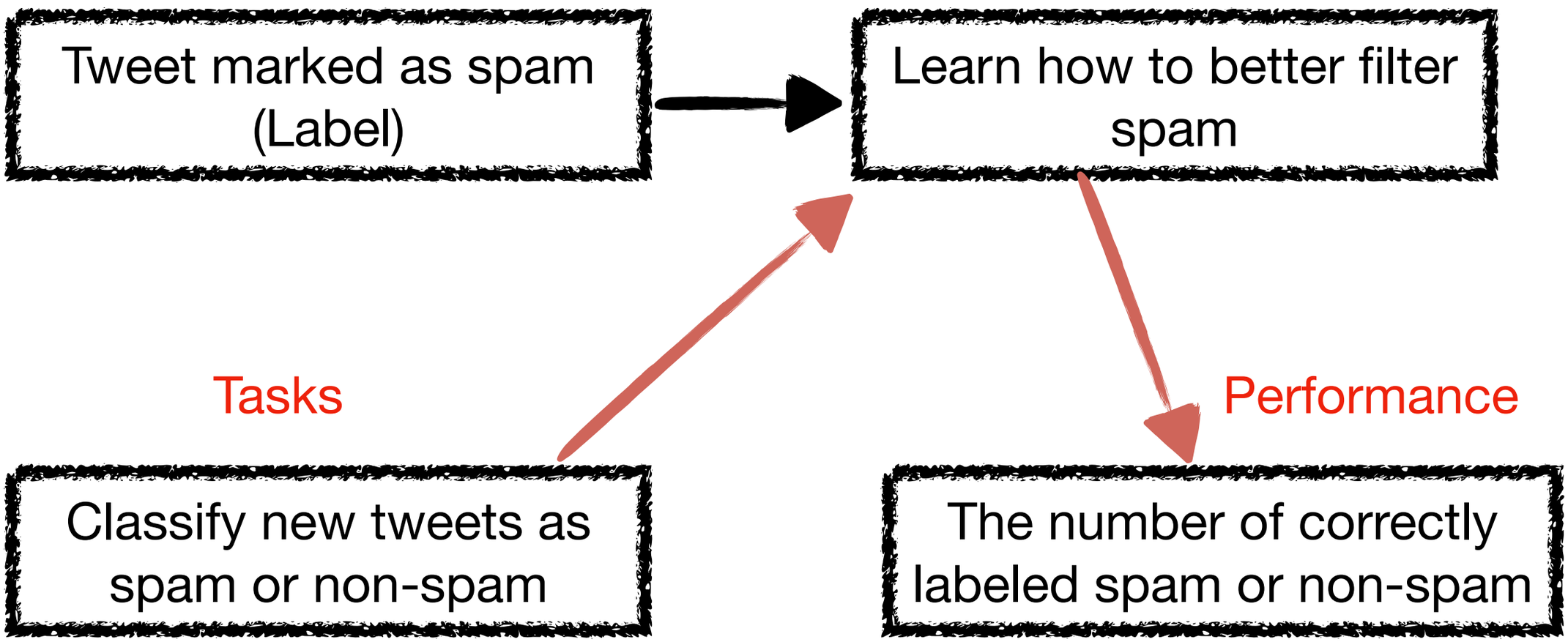
Spam Classification Example

- *Suppose Twitter server watches which tweets marked as spam message. Based on this information, he will learn how to better filter spam.*



Spam Classification Example

- *Suppose a Twitter server watches which tweets are marked as spam messages. Based on this information, it will learn how to better filter spam.*



Weather Prediction Example

- *Suppose a Mesonet station monitors the weather conditions for the past several years, then based on this information, a computer program can learn and predict the weather conditions in next several days.*

Weather Prediction Example

- *Suppose a Mesonet station monitors the weather conditions for the past several years, then based on this information, a computer program can learn and predict the weather conditions in next several days.*

Past several years' observation

Experience

Algorithms

Weather Prediction Example

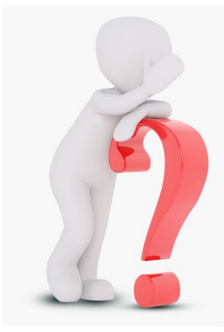
- *Suppose a Mesonet station monitors the weather conditions for the past several years, then based on this information, a computer program can learn and predict the weather conditions in next several days.*

Past several years' observation

Last one week's observation

Tasks

Next week



Machine Learning ~ Looking for a Function

- Image recognition

- Spam classification

- Weather prediction

Machine Learning ~ Training Framework




Training Data

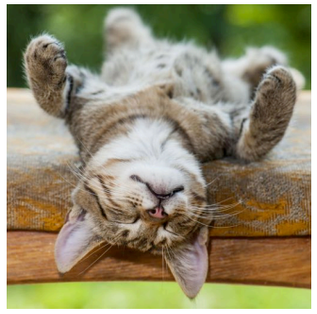
A set of functions (models) $f_1, f_2, \ldots$

Goodness of function f

Pick the "best" function f^*

Trained Model

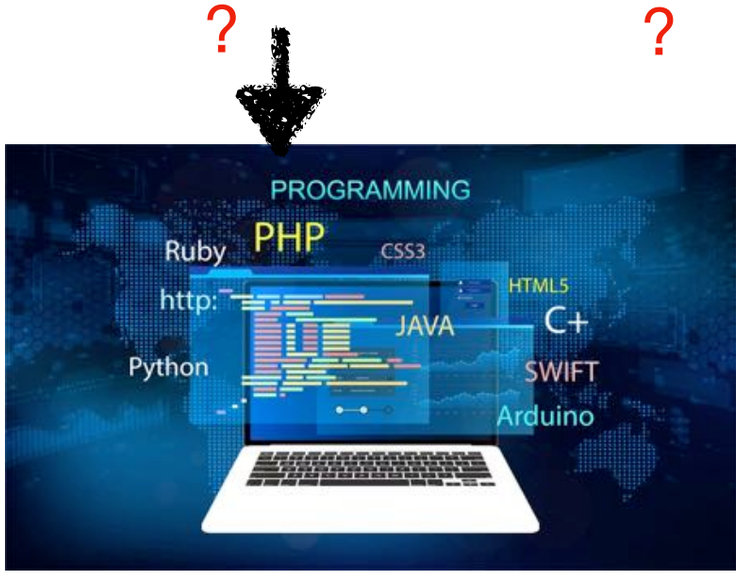
Machine Learning ~ Testing Framework

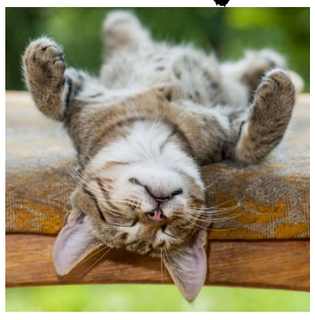


?

?

?





“Cat” (95%)

“Cat” (95%)

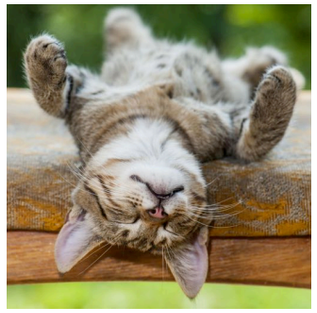
“Cat” (85%)

Testing Data

Trained Model (f)

Labels

Machine Learning ~ Testing Framework

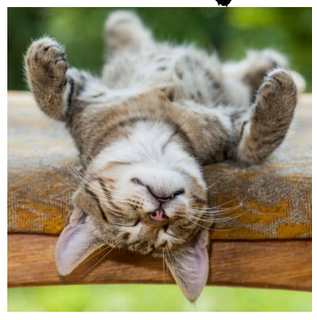


?

?

?

?



"Cat" (95%)

"Cat" (95%)

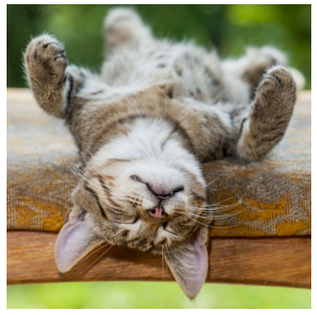
"Cat" (85%)

Testing Data

Trained Model (f)

Labels

Machine Learning ~ Testing Framework



?

?

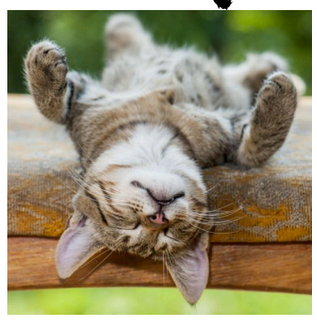
?

?

Testing Data

Trained Model (f)

Labels



“Cat” (95%)

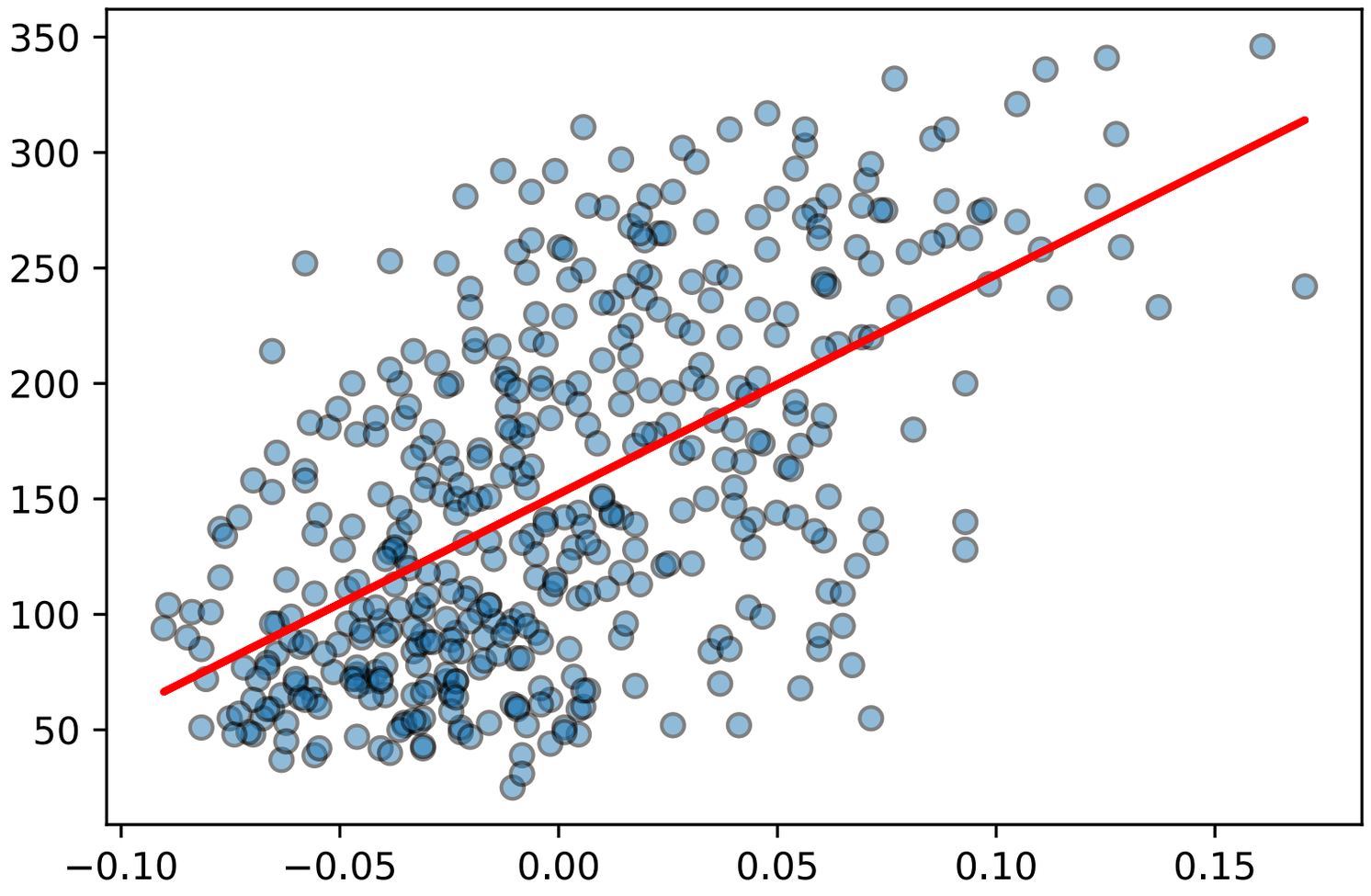
“Cat” (95%)

“Cat” (85%)

“Unknown” (what’s this guy?)

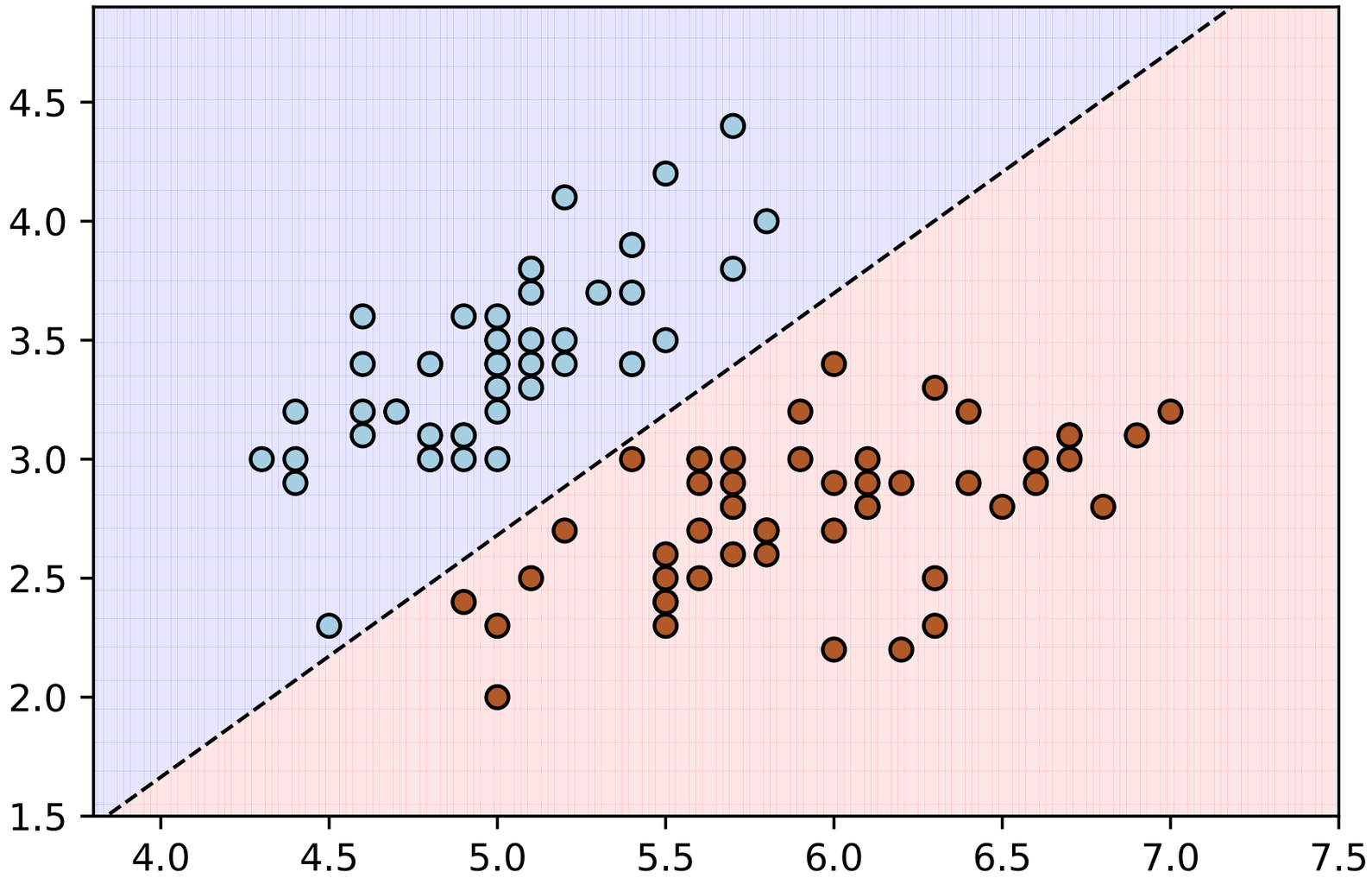
So far,
you can see finding a suitable function is the
core of machine learning

Linear Regression



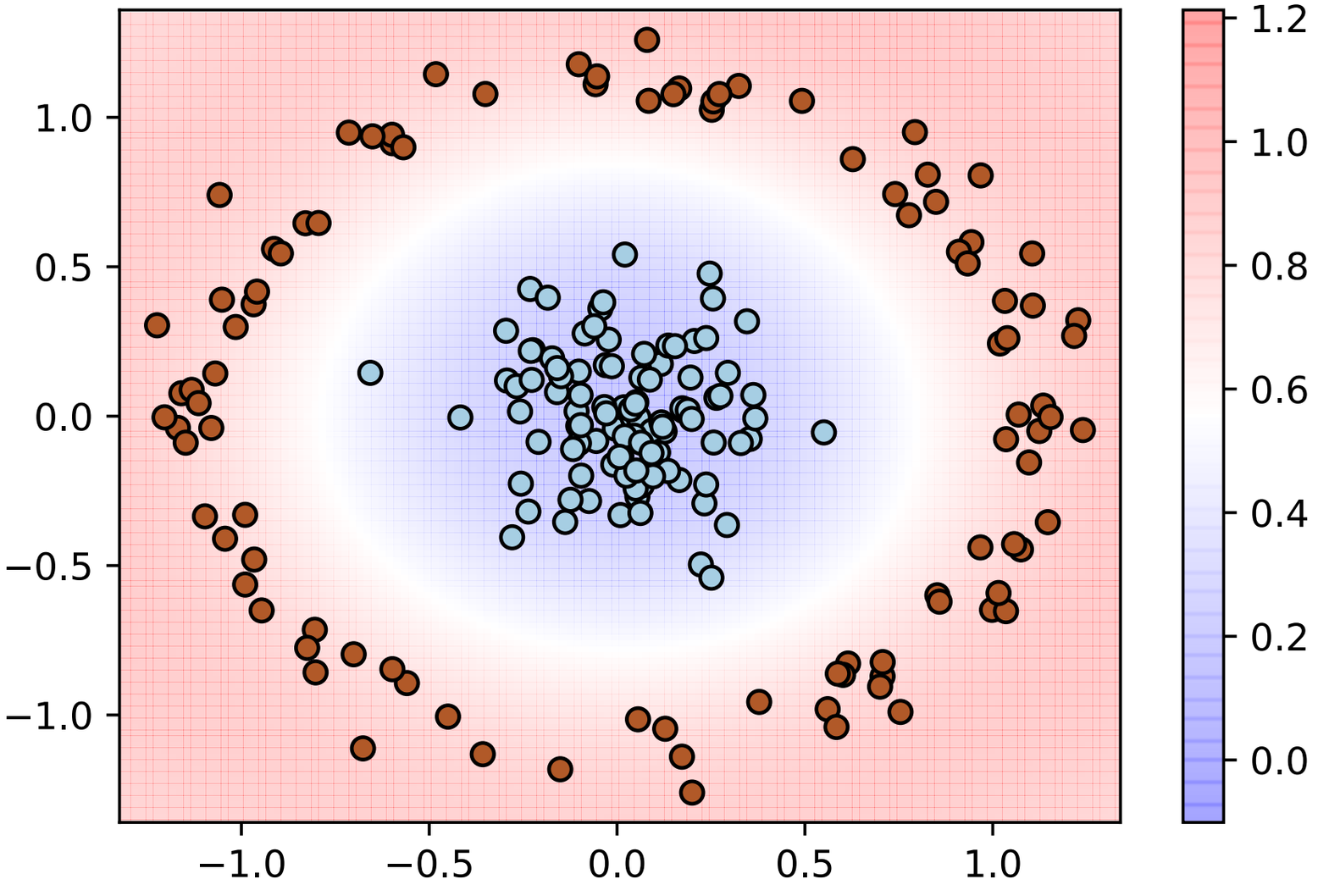
Finding a function that best fits the curve

Logistic Regression



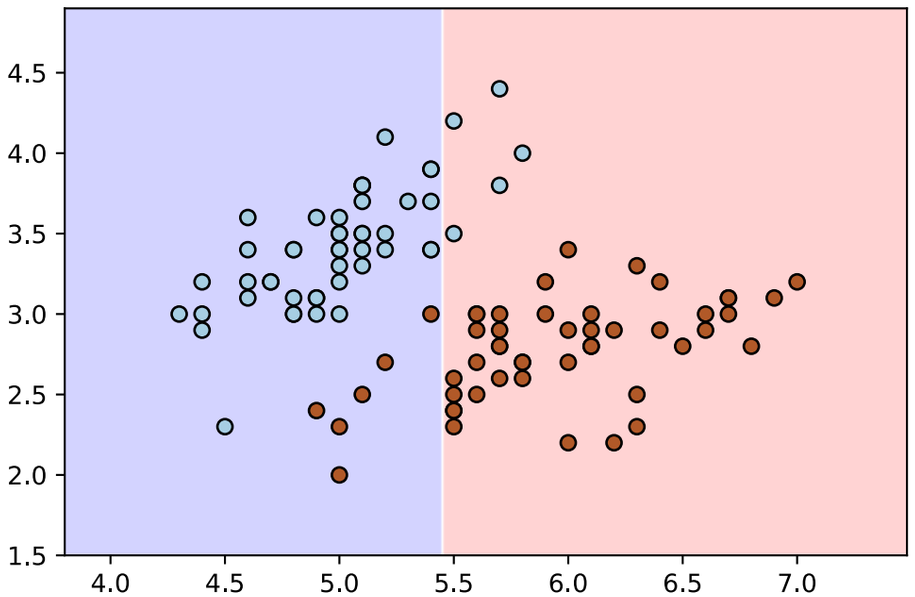
A function is used to define the boundary line

Supported Vector Machine (SVM)

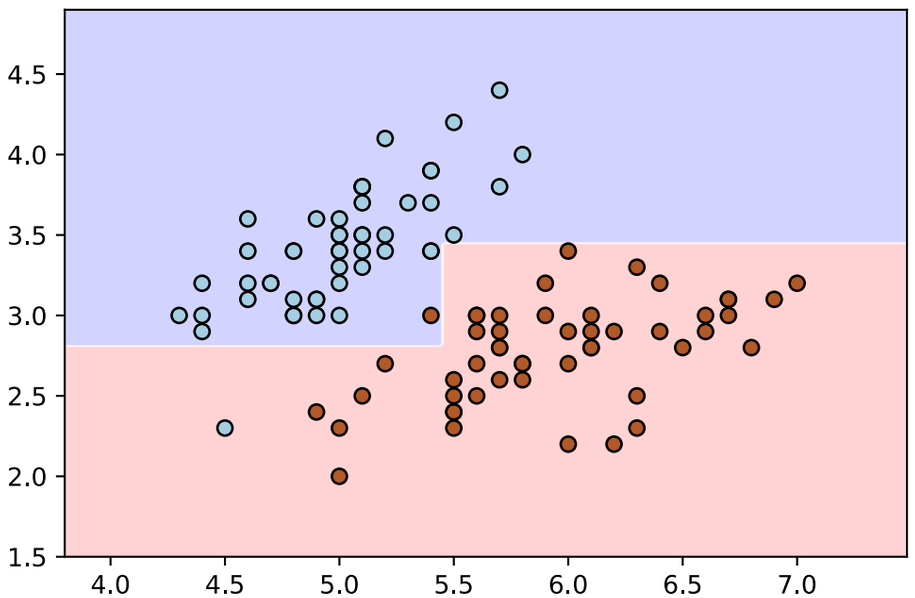


The boundary curves are non-linear.

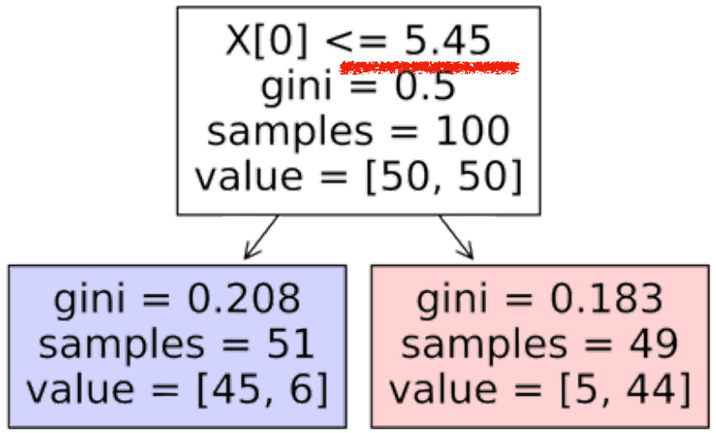
Decision Tree



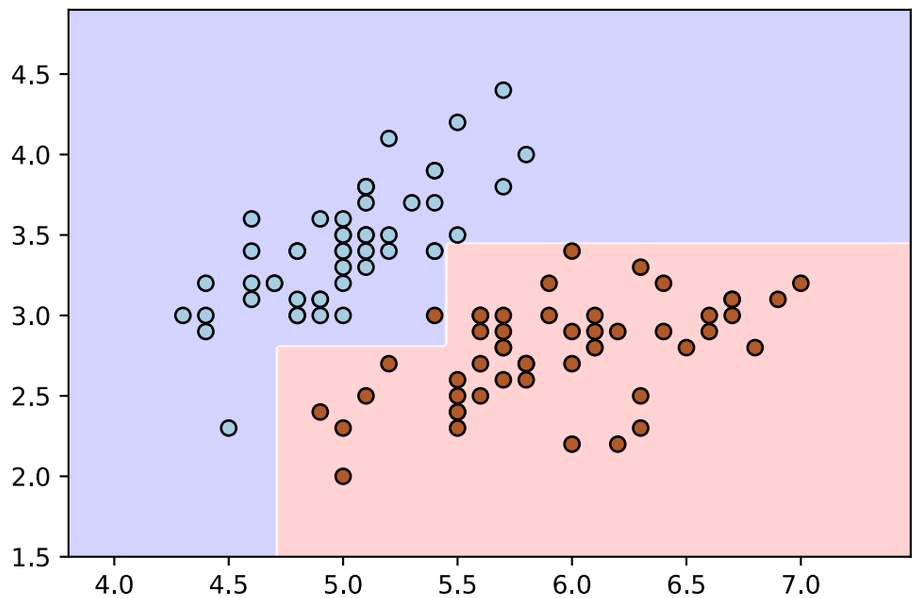
tree height = 1



tree height = 2

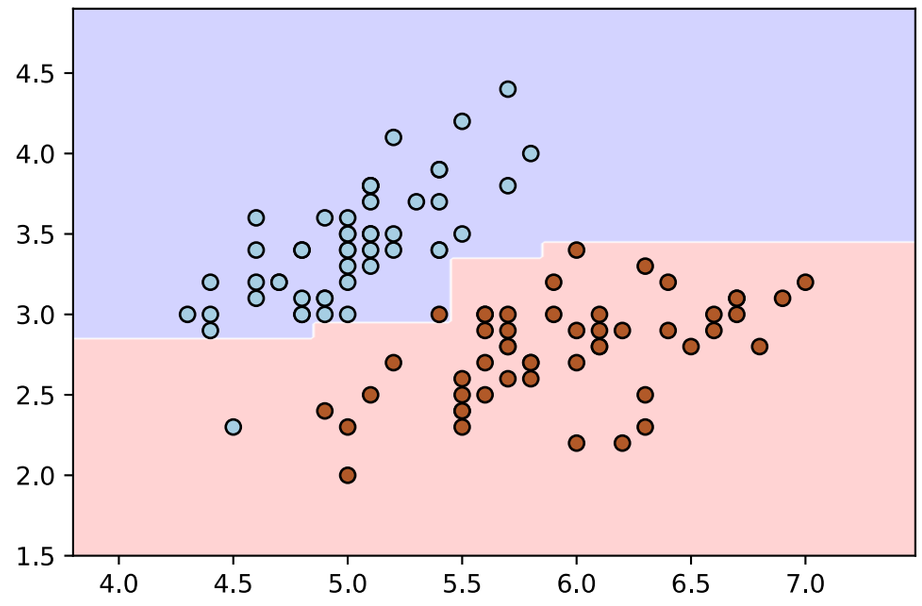


Decision tree with height 1

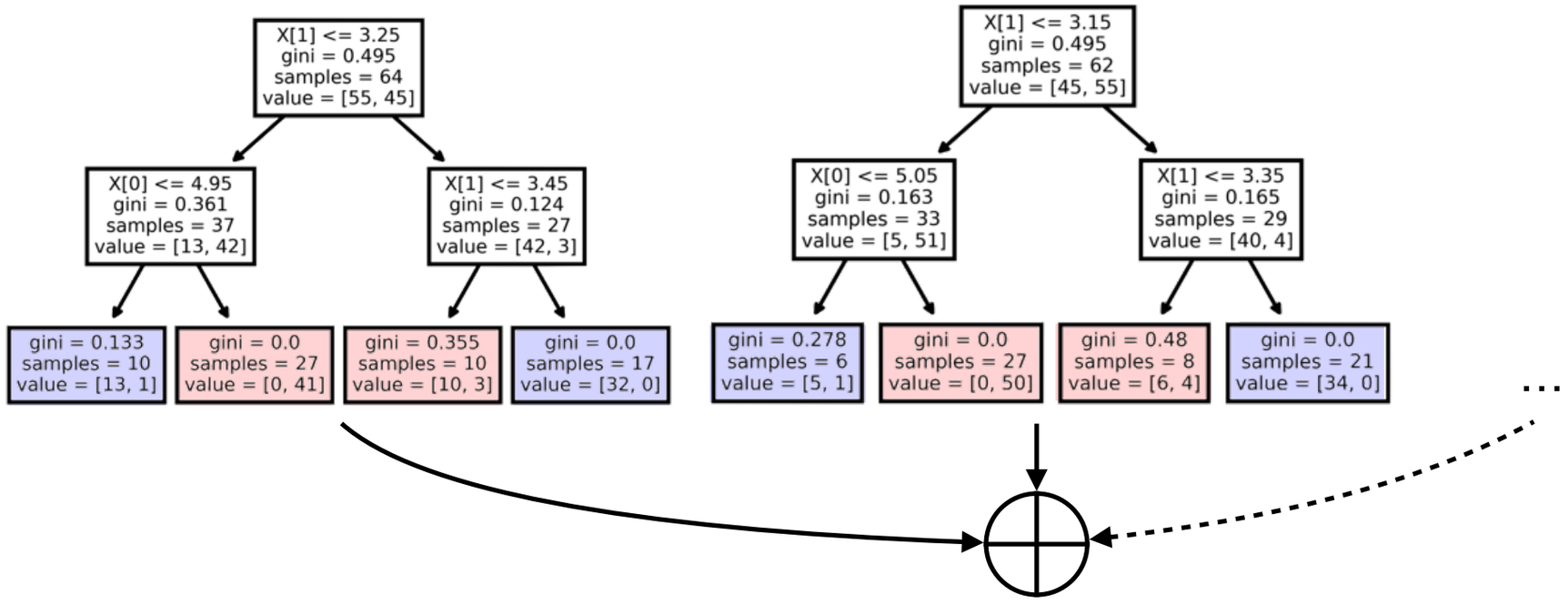


tree height = 3

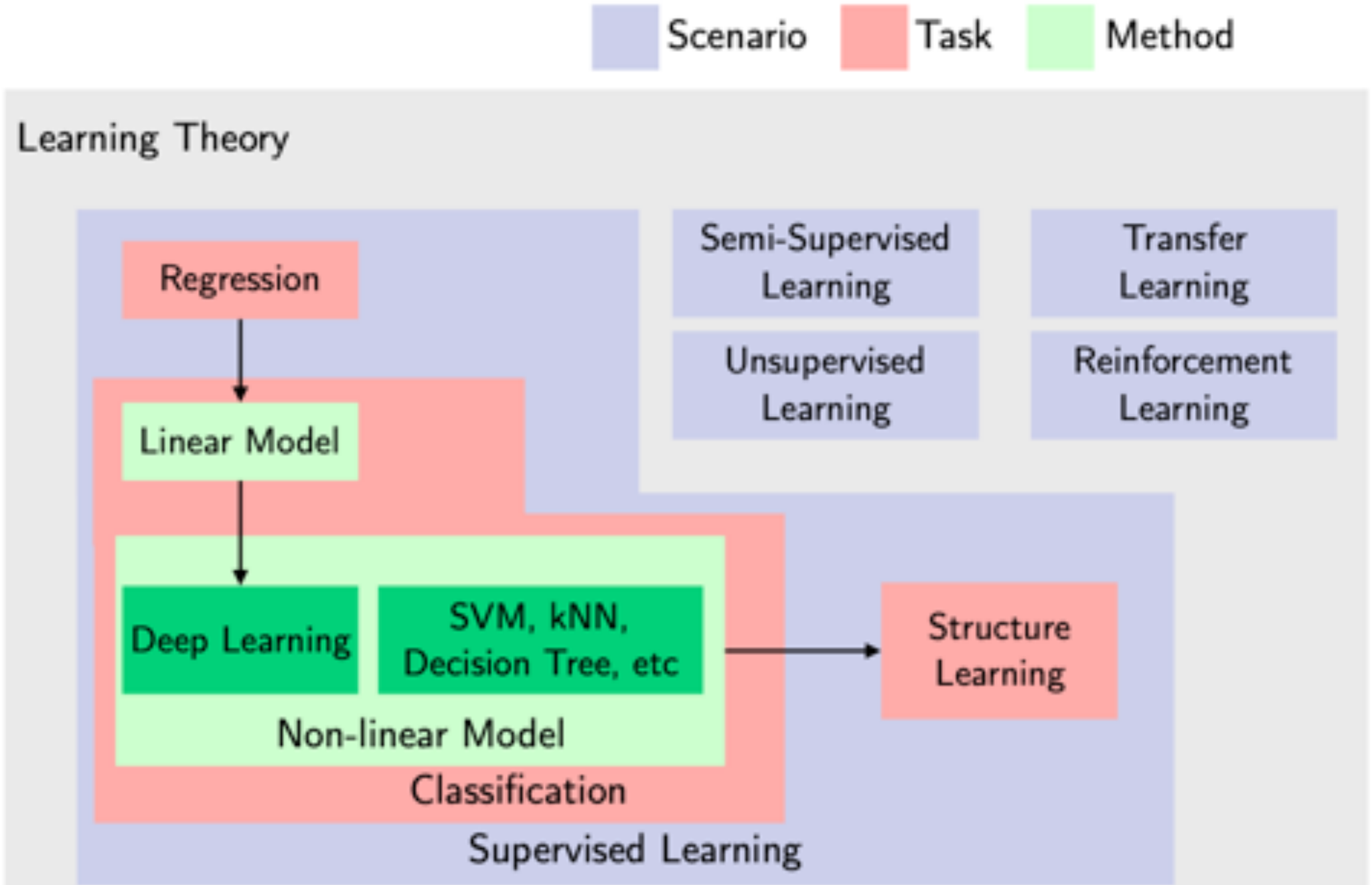
Random Forest



number of trees = 10, tree height = 2

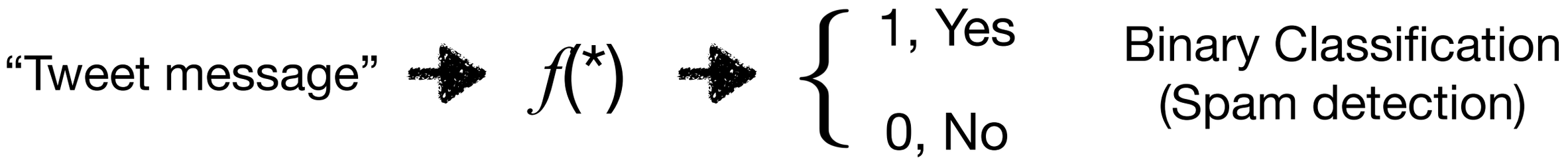


Learning Map

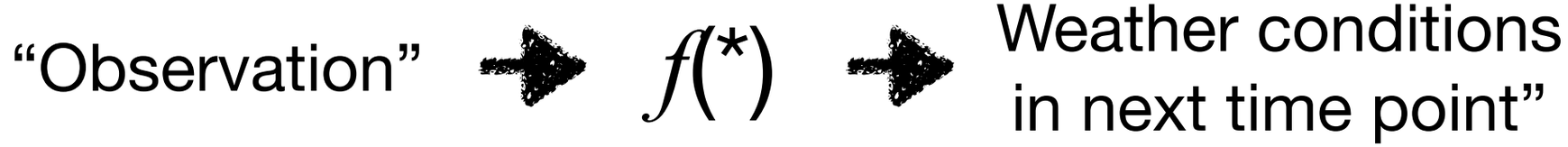


Supervised Learning

- **Classification**
  - Each element in the sample is labeled as belonging to some class. No order among classes.



- **Prediction**
  - Elements in the sample have the inherent relationships to weather condition at some time point.

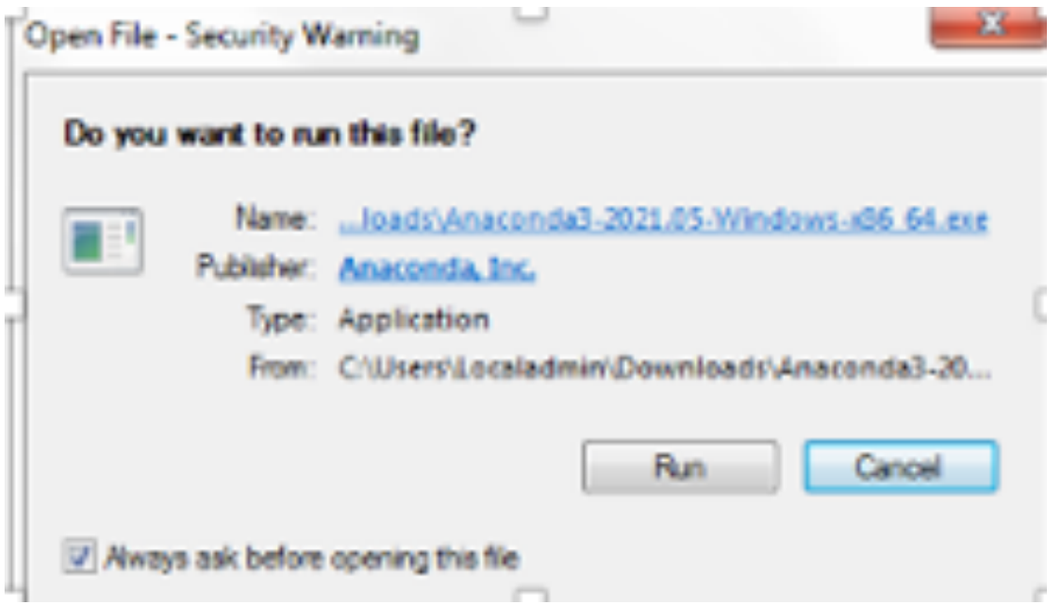


Before starting, we need to know Python

- Python provides a set of libraries including different ML packages
- Standard libraries provide the ready-to-use implementation of algorithms
- The scikit-learn is the one we will use in this course

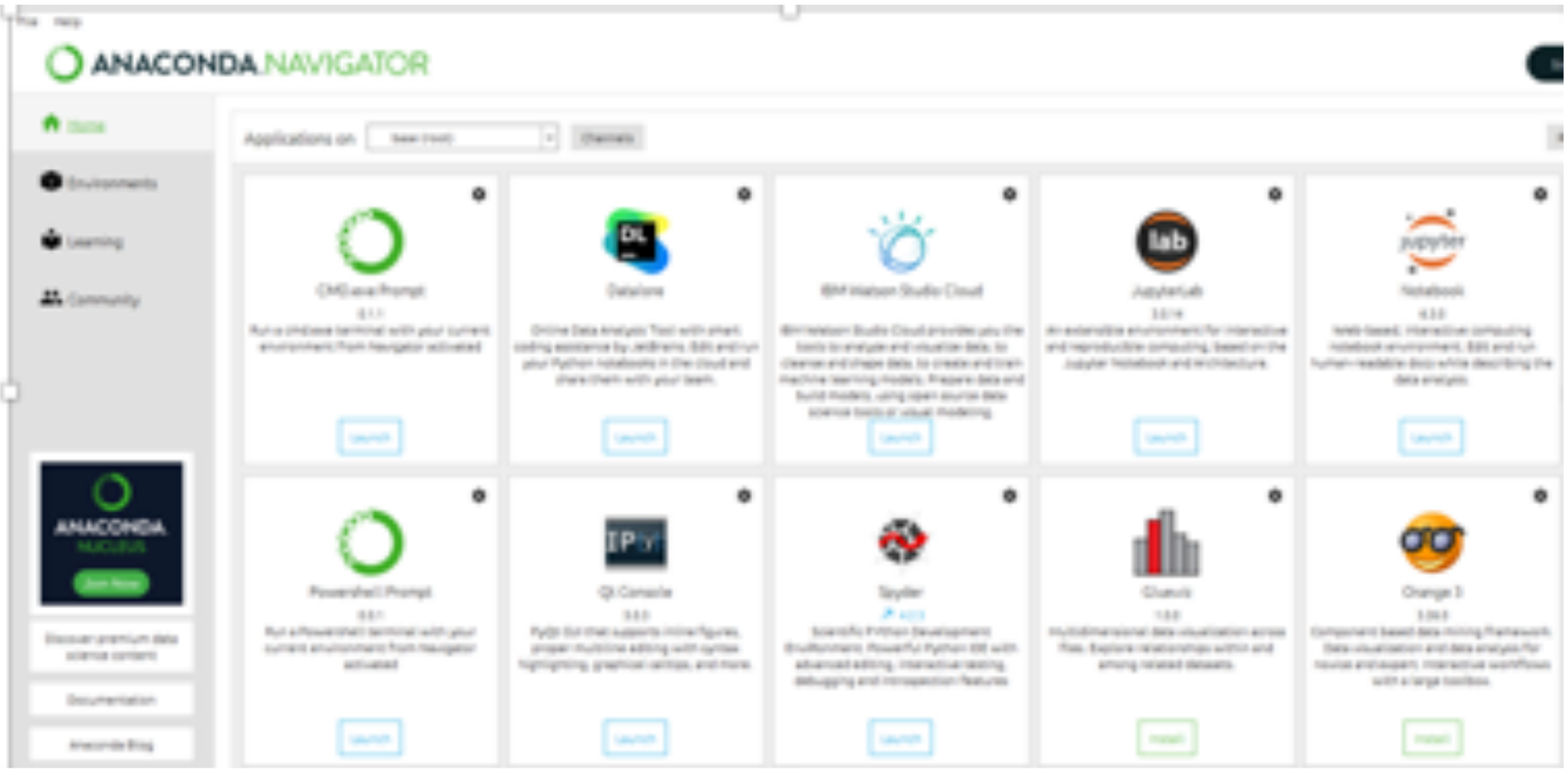
Installing Anaconda Navigator

1. Browse https://docs.anaconda.com/anaconda/install/windows/
2. Click on Download the Anaconda installer
   - Check your OS bit version: Start button->Settings->System->About: Device specification System Type
   - Click on (your_OS_bit_version)-Bit Graphical Installer, e.g., 64-Bit Graphical Installer, and click on save (will take a while for downloading)
3. Double click the installer to launch and click on Run for installation
4. Click on Next -> I Agree -> Next ->Next->Install (for default settings)



Installing Anaconda Navigator (Continuing…)

5. Click Next->Next->Finish to complete the installation (registration is not essential for operation).
6. Open Anaconda Navigator: It will pop up an icon in the status bar.
7. Click on the icon and click on the launch button of Jupyter Notebook.

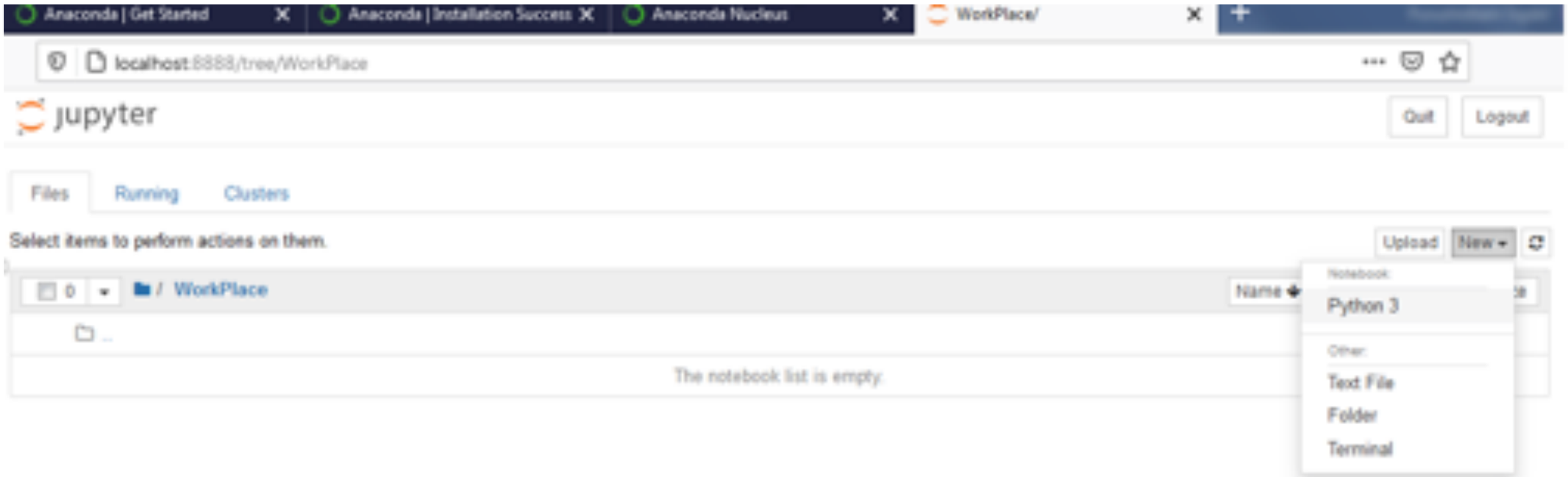


Installing Anaconda Navigator (Continuing…)

8. It will open the browser and show your files and directory (folders) from C:\Users\Your_user_account.
9. For the time being, create a working directory at C: \Users\Your_user_account\[yourWorkingDirectory]

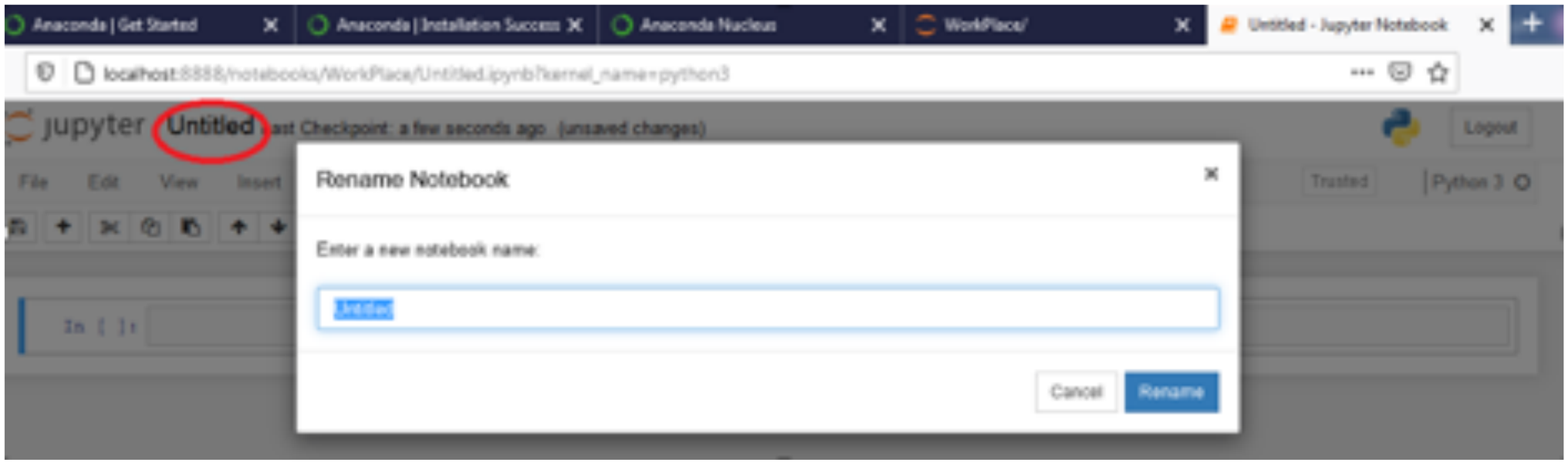
Installing Anaconda Navigator (Continuing…)

10. Click on your working directory (in my case, it is ‘workPlace’). It will take you to a new window.
11. Click on the New dropdown button (on the right side) and click on the Python 3.

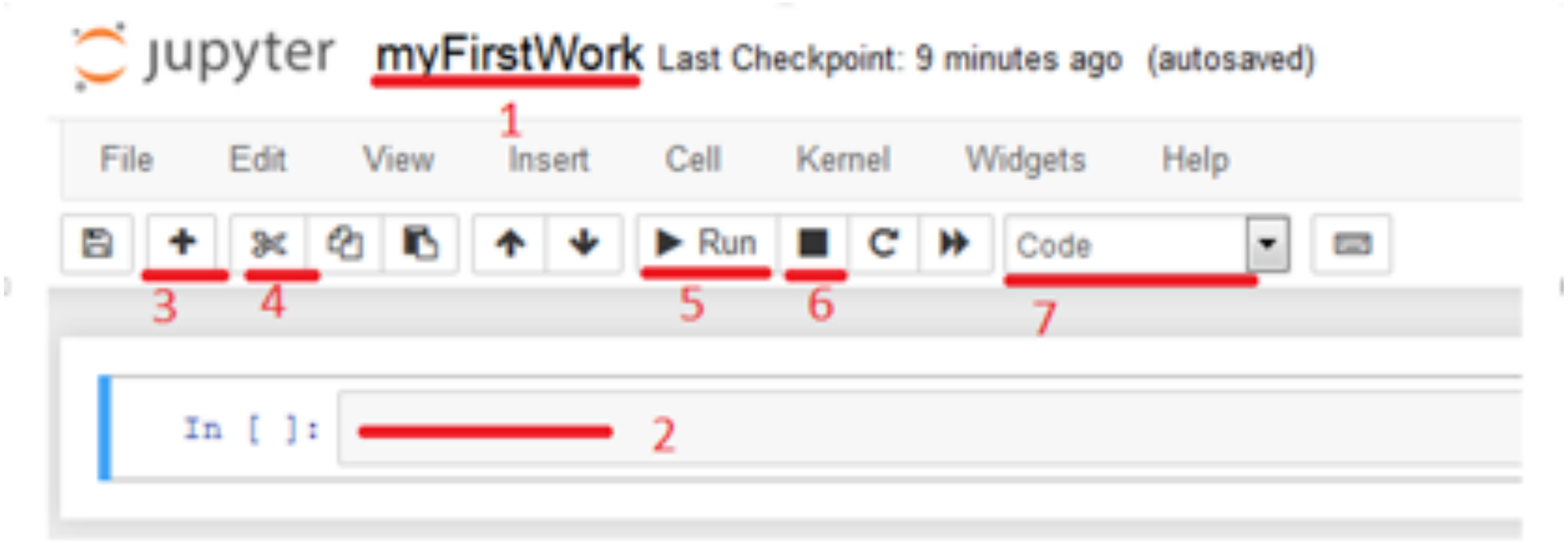


Installing Anaconda Navigator (Continuing…)

12. It will open a new page in the browser with the Untitled – Jupyter Notebook page. To change the name, click on the Untitled label (on the top left) and rename your file.



Frequently Used buttons



Examples

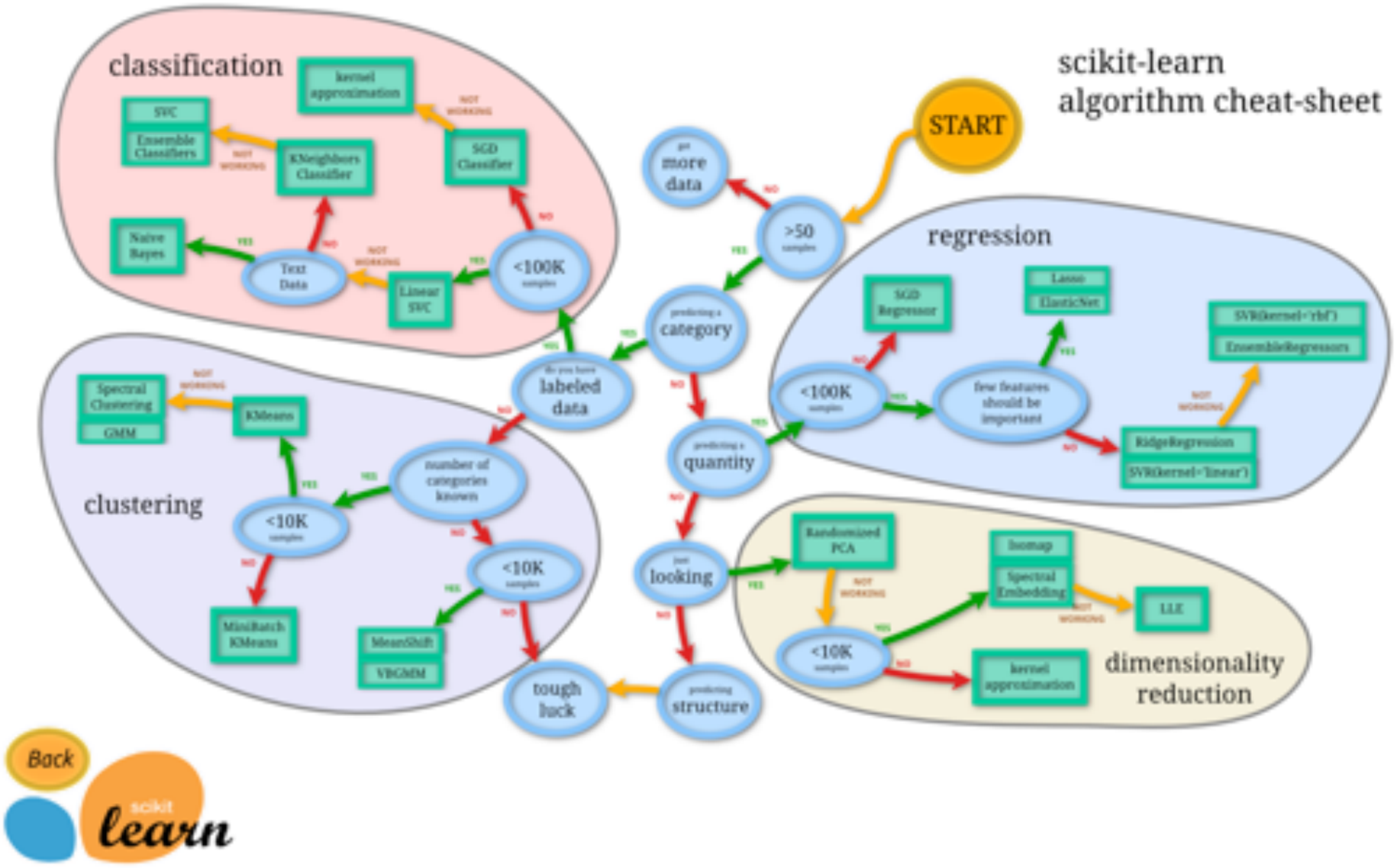
jupyter myFirstWork Last Checkpoint: 40 minutes ago (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

Run Code

```
In [1]: print("Hello World!")
Hello World!

In [2]: 7+5
Out[2]: 12

In [3]: import numpy as np
        x = np.random.rand(1)

In [4]: x
Out[4]: array([0.95841561])

In [5]: print(x)
        [0.95841561]

In [ ]:
```

Scikit-learn



Source: https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html

Example 1

```
from sklearn import svm

X = [[0, 1],[1, 2],[2, 1],[2, 3],[1, 3],[2, 2]]

y = ['a', 'a', 'b', 'b', 'a', 'b']

clf = svm.SVC()

clf.fit(X, y)

result1 = clf.predict([[3, 1]])

print(result1)

result2 = clf.predict([[0, 2]])

print(result2)
```

```
['b']
['a']
```

Example 2

```
from sklearn import svm

from sklearn.datasets import load_iris

#iris dataset contains 150 samples, each has 4 features
X, y = load_iris(return_X_y = True)

'''
Parameter 'return_X_y = True' is required in
load_iris()function to get the sample and label data in
seperate variables.
'''

print("The size of the sample:", X.shape)

print("First 5 samples:\n", X[0:5])
print("First 5 labels:\n", y[0:5])

clf = svm.SVC()

clf.fit(X, y)

result = clf.predict(X[45:55])

print("Predicted labels\n",result)

print("Actual labels\n",y[45:55])
```

```
The size of the sample: (150, 4)
First 5 samples:
 [[5.1 3.5 1.4 0.2]
 [4.9 3.  1.4 0.2]
 [4.7 3.2 1.3 0.2]
 [4.6 3.1 1.5 0.2]
 [5.  3.6 1.4 0.2]]
First 5 labels:
 [0 0 0 0 0]
Predicted labels
 [0 0 0 0 0 1 1 1 1 1]
Actual labels
 [0 0 0 0 0 1 1 1 1 1]
```

More Resources

Please check
https://people.cmix.louisiana.edu/yuan/2022_Summer_Tutorial_Courses.html
for more examples

Q&A

Thank You!