Supervised Learning, Python Tutorial

อ. ปรัชญ์ ปิยะวงศ์วิศาล

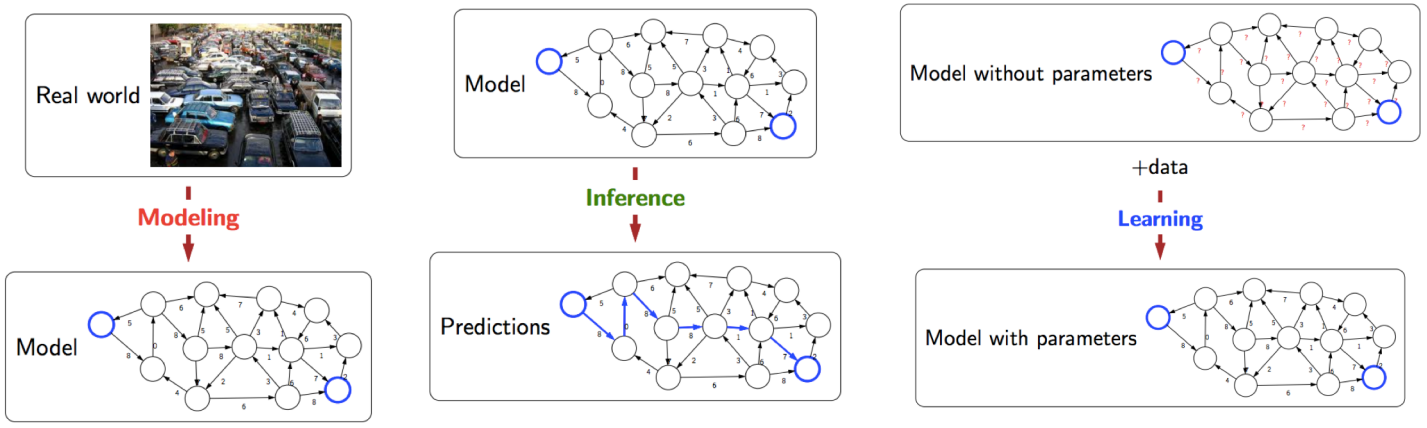
Pratch Piyawongwisal

Today

- Recap
- Machine Learning
  - Supervised vs Unsupervised Learning
  - Classification vs Regression
- Python Tutorial

Recap: Artificial Intelligence (AI)

- Study of how to make “Intelligent Agents”
- Strong AI vs Weak AI
- Two sources of complexity that makes AI hard:
  - Computational complexity
  - Information complexity
- Model-Inference-Learning

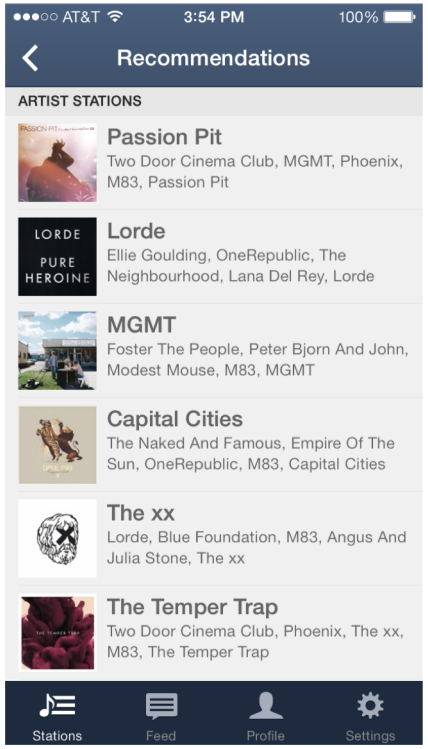


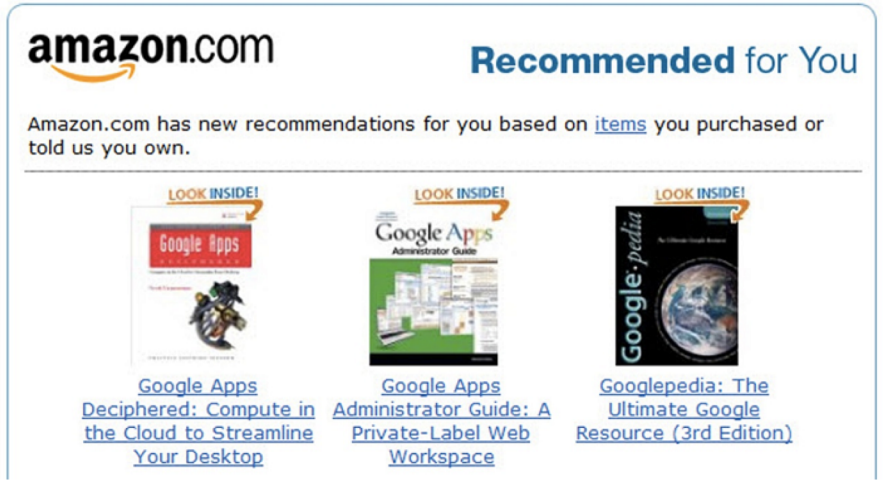
Machine Learning (ML) – Informal Definition

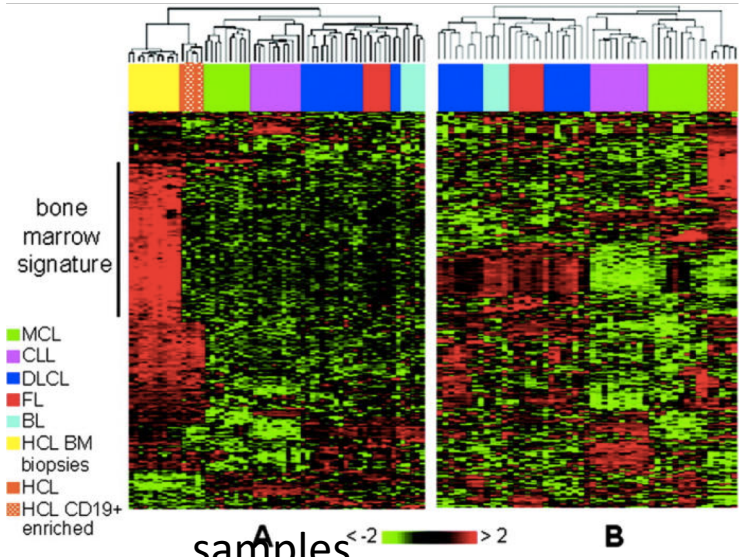
- The science of getting machines to **“learn”** from data and make predictions **without being explicitly programmed**

- การเรียนรู้ของเครื่อง คือการสอนให้เครื่องแก้ปัญหา โดยให้เครื่อง**เรียนรู้**จากข้อมูลเอง แทนการเขียนโปรแกรมแก้ปัญหาตรงๆ
- เครื่องจะหา pattern ในข้อมูลเพื่อนำไปใช้**ทำนาย**คำตอบที่ถูกต้องแม่นยำ

Machine Learning เป็นสิ่งใกล้ตัว







Gene expression

Example 1

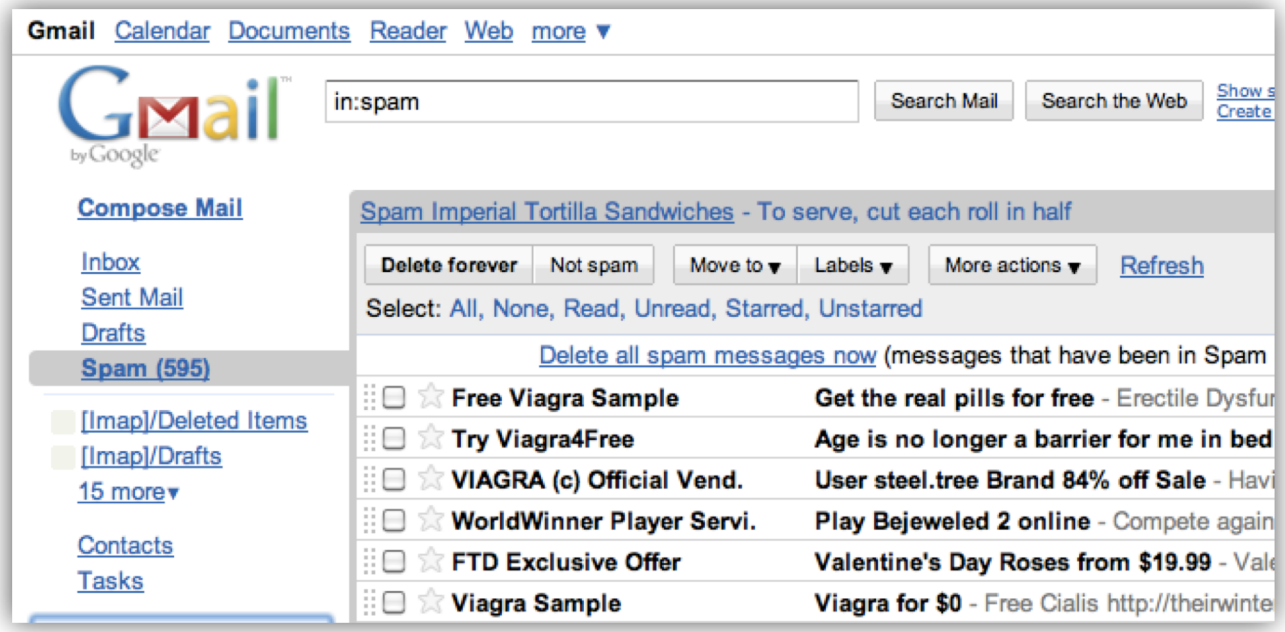
- สมมติว่าเราจะเขียนเกม rock-paper-scissors โดยใช้ภาพถ่ายจากมือถือ



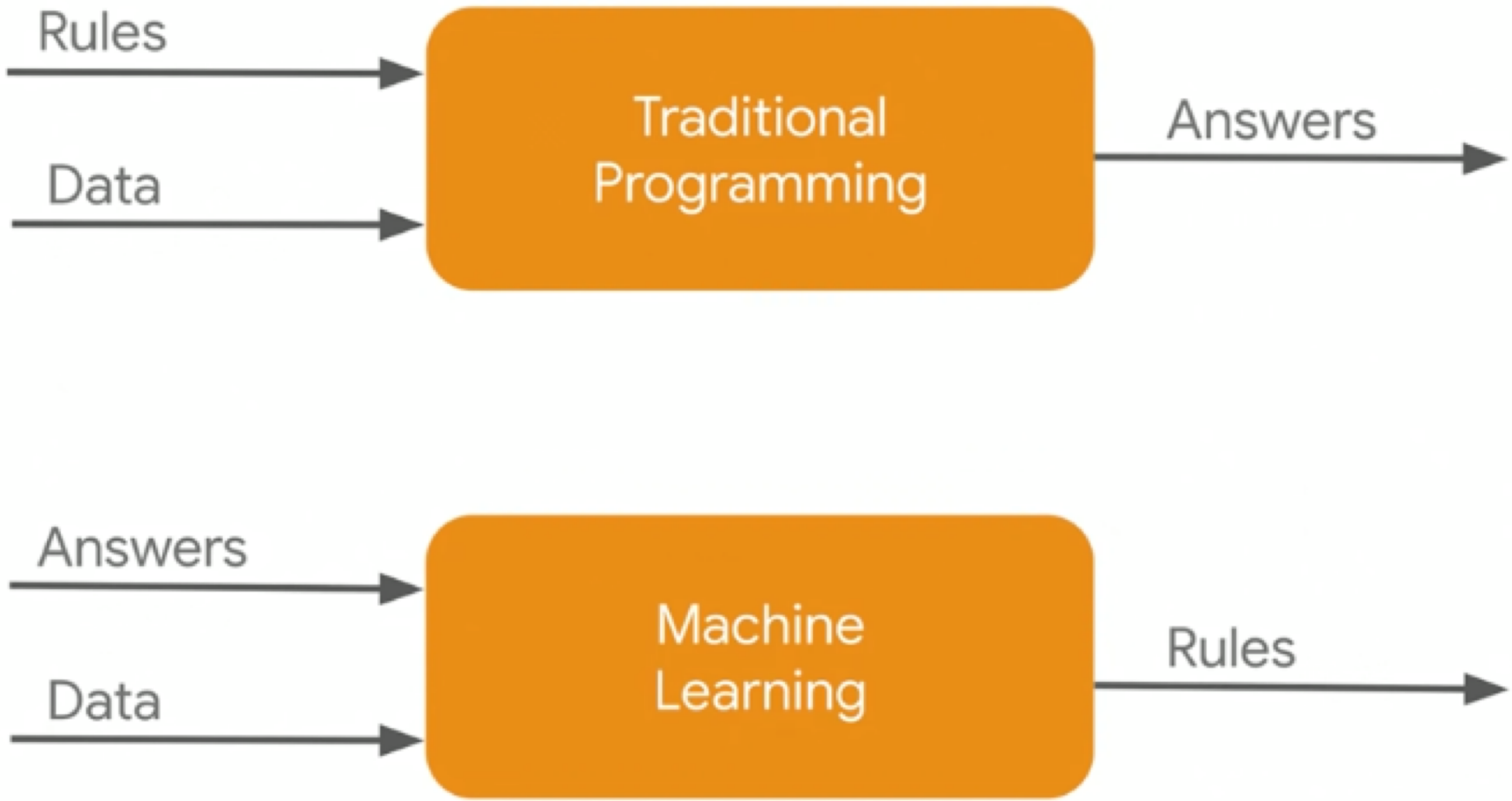
Source: Machine Learning Zero to Hero (Google I/O'19)

Example 2

- เขียนโปรแกรมเพื่อคัดกรองอีเมล **spam** ออกจากอีเมลปกติ **(ham)**



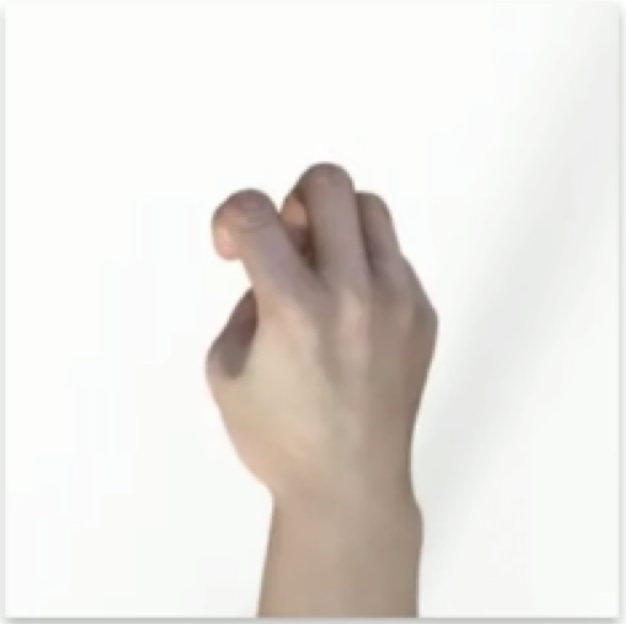
Source: https://towardsdatascience.com/applied-text-classification-on-email-spam-filtering-part-1-1861e1a83246



Source: Machine Learning Zero to Hero (Google I/O'19)

Data and Labels

- ในการสอนเครื่อง เราจะให้มนุษย์กำหนด **label** เฉลย สำหรับข้อมูลแต่ละชิ้น

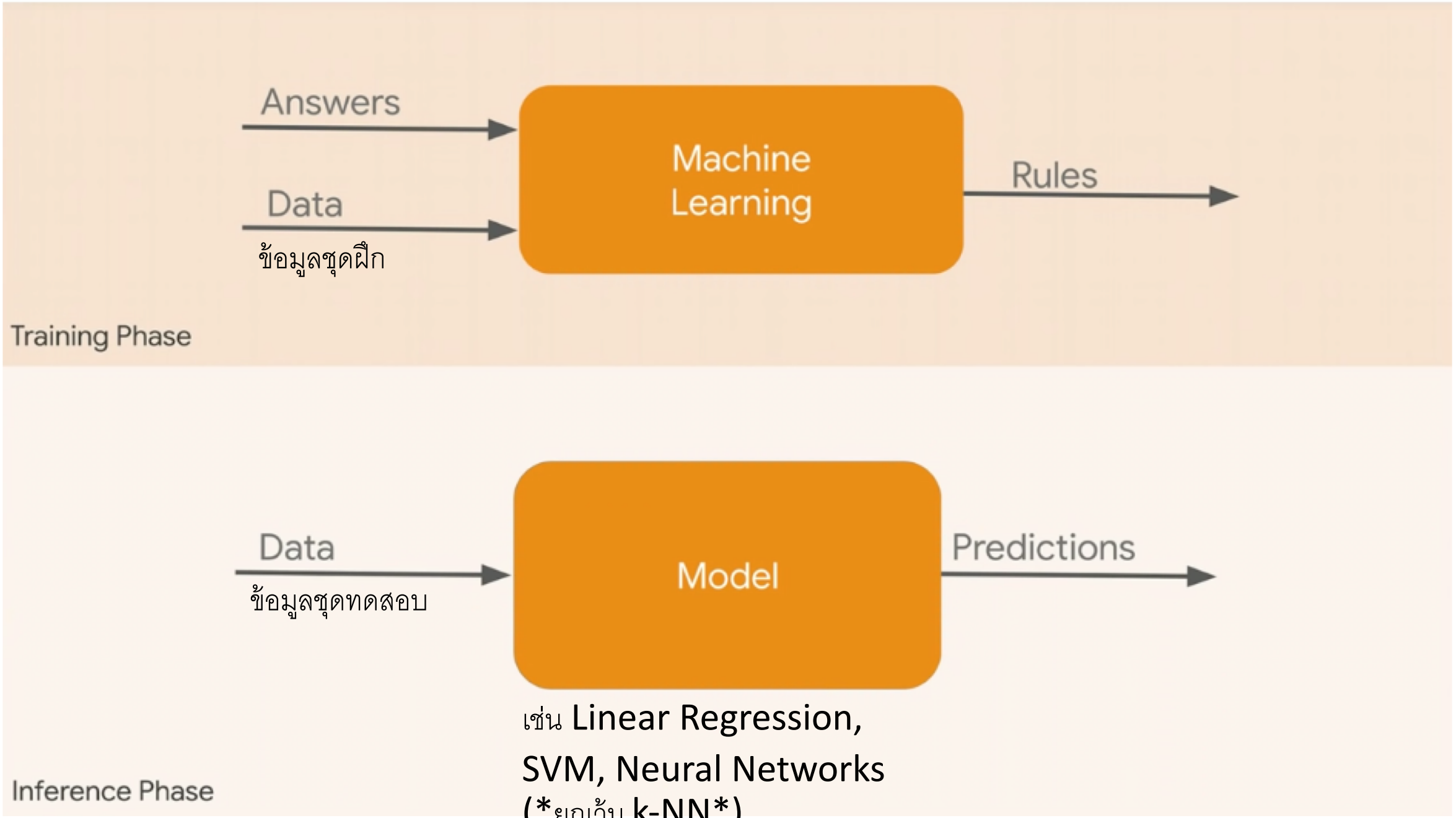


Label = ROCK

Label = PAPER

Label = SCISSORS

Source: Machine Learning Zero to Hero (Google I/O'19)

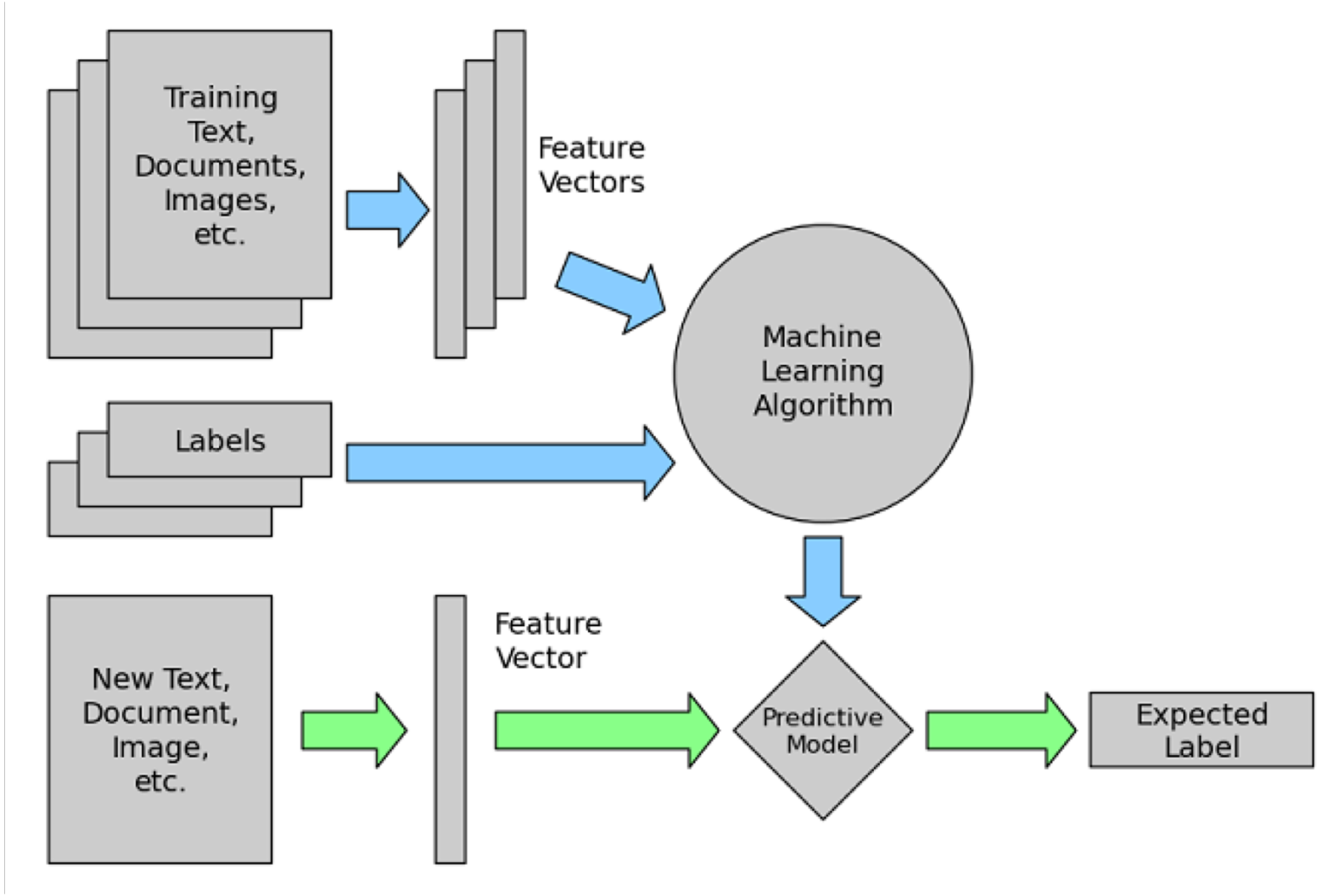


เช่น Linear Regression,
SVM, Neural Networks
(*ยกเว้น k-NN*)

* inference = การอนุมาน

Source: Machine Learning Zero to Hero (Google I/O'19)

How ML works?



Source: https://ogrisel.github.io/scikit-learn.org/sklearn-tutorial/tutorial/text_analytics/general_concepts.html

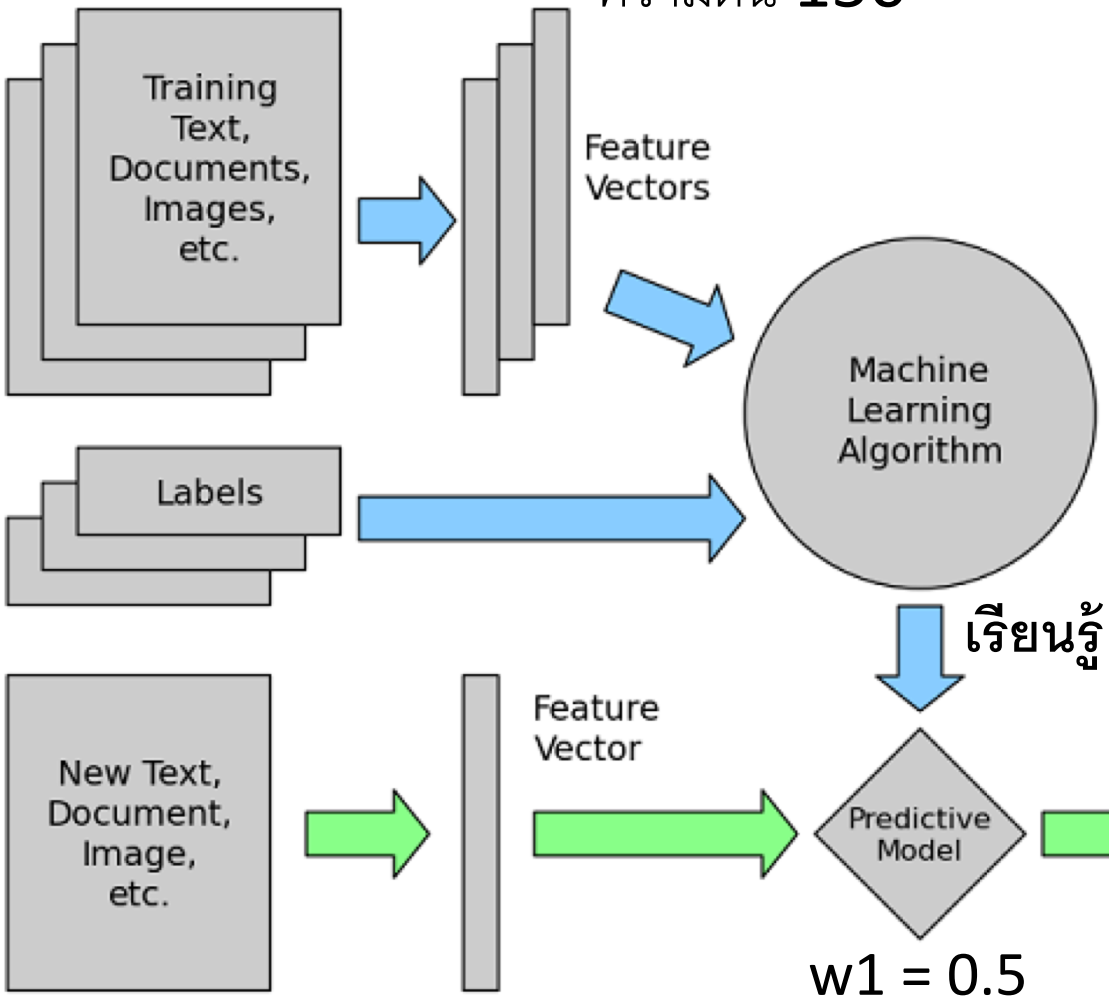
Example

นาย A
น้ำหนัก 60
สูบบุหรี่ 1
ความดัน 150

ข้อมูลคนไข้

นาย A
เป็นมะเร็งในปอด

นาย Z
น้ำหนัก 75
สูบบุหรี่ 0
ความดัน 88



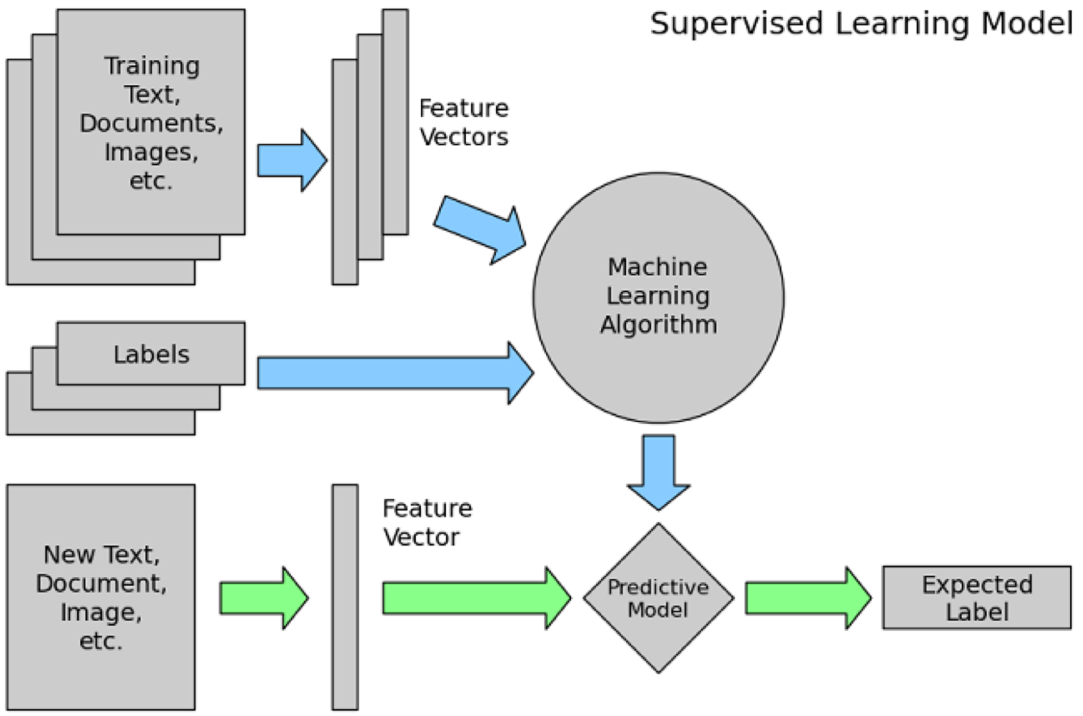
โมเดล:
Y = w1*น้ำหนัก+w2*สูบบุหรี่
เป็นมะเร็งในปอดถ้า Y > 100

w1 = 0.5
w2 = 50

นาย Z
Y = 37.5
ไม่เป็นมะเร็งในปอด !!

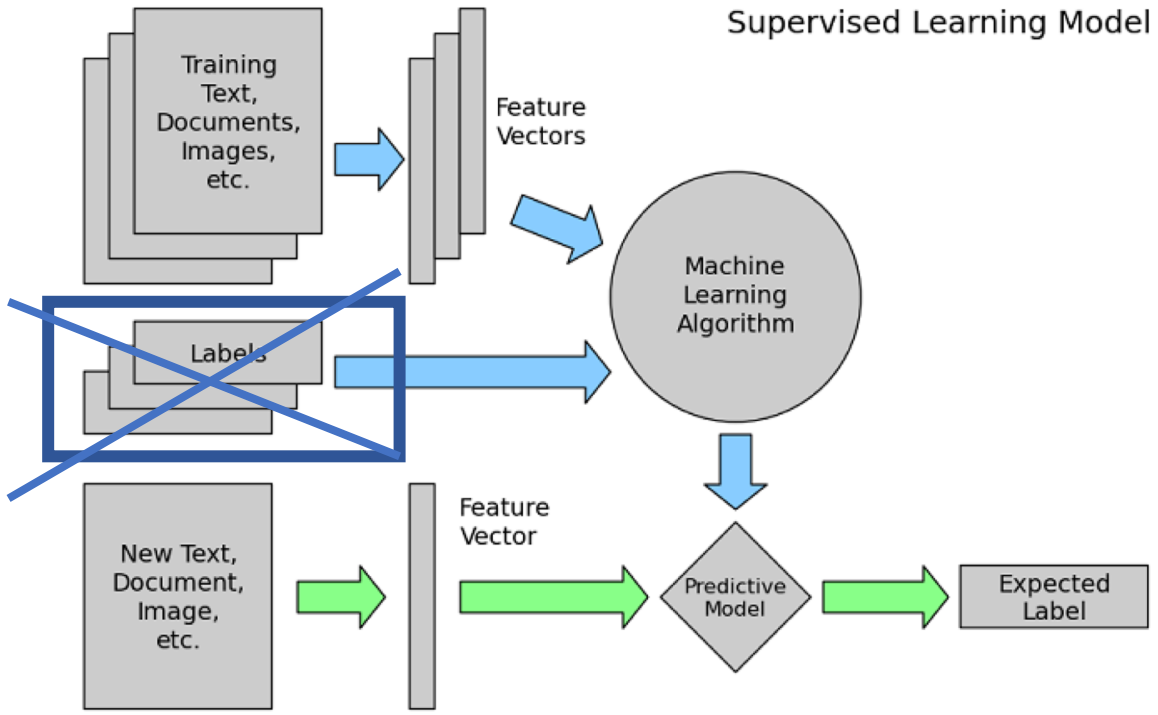
Must-know Terminology

- Training set ข้อมูลชุดสอน
- Test set ข้อมูลชุดทดสอบ
- Feature คุณลักษณะเด่น
- Class Label ชนิดที่จำแนก
- Model แบบจำลองคณิตศาสตร์
- Predictor ตัวทำนาย
- Classifier ตัวจำแนกชนิด
- Training Error ค่าคาดเคลื่อนในการฝึก
- Testing Error ค่าคาดเคลื่อนในการทดสอบ



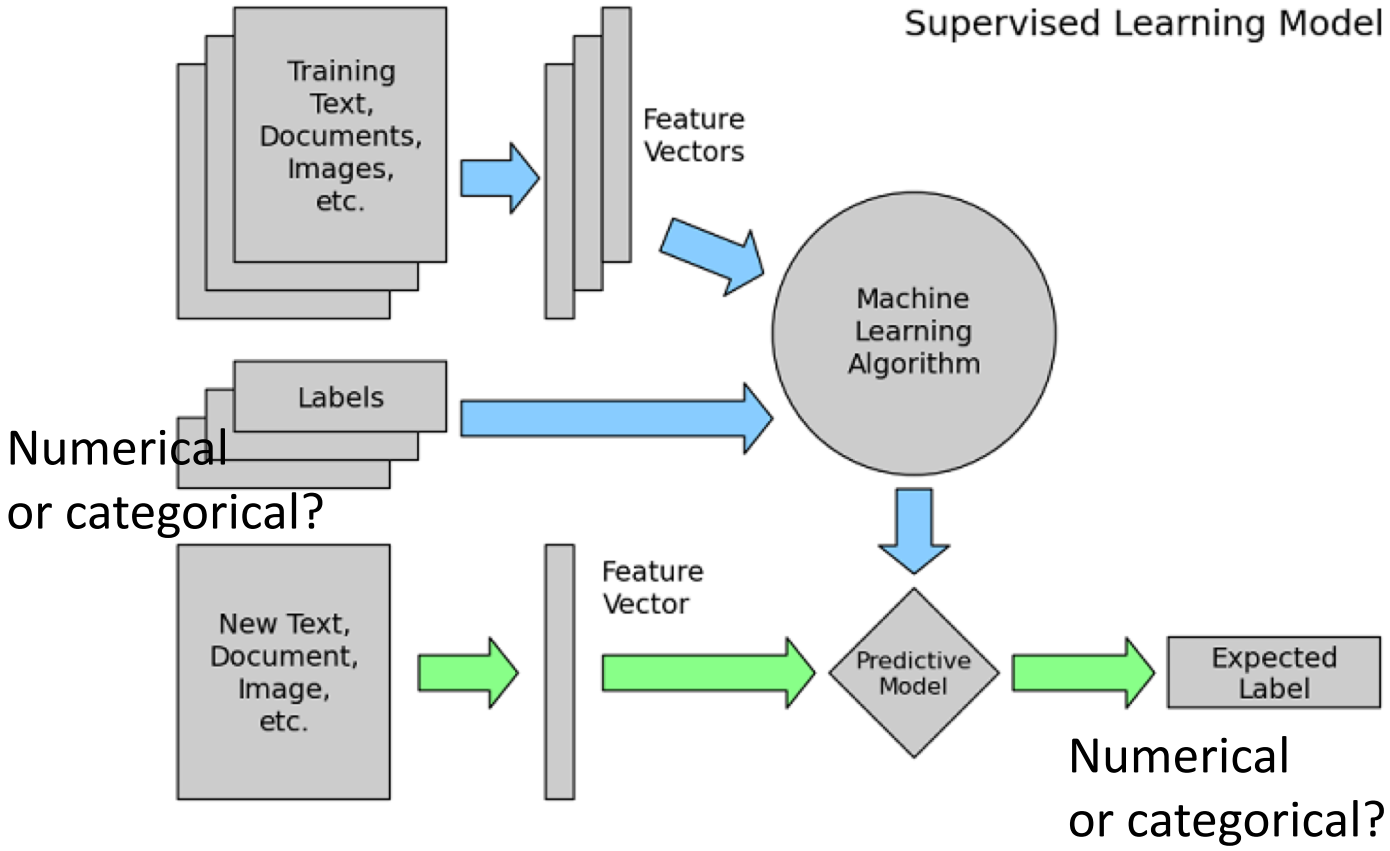
2 Types of ML Systems

- Supervised Learning
  - Trained **with** human supervision
  - Training set has **labels**
- Unsupervised Learning
  - Trained **without** human supervision
  - No class labels

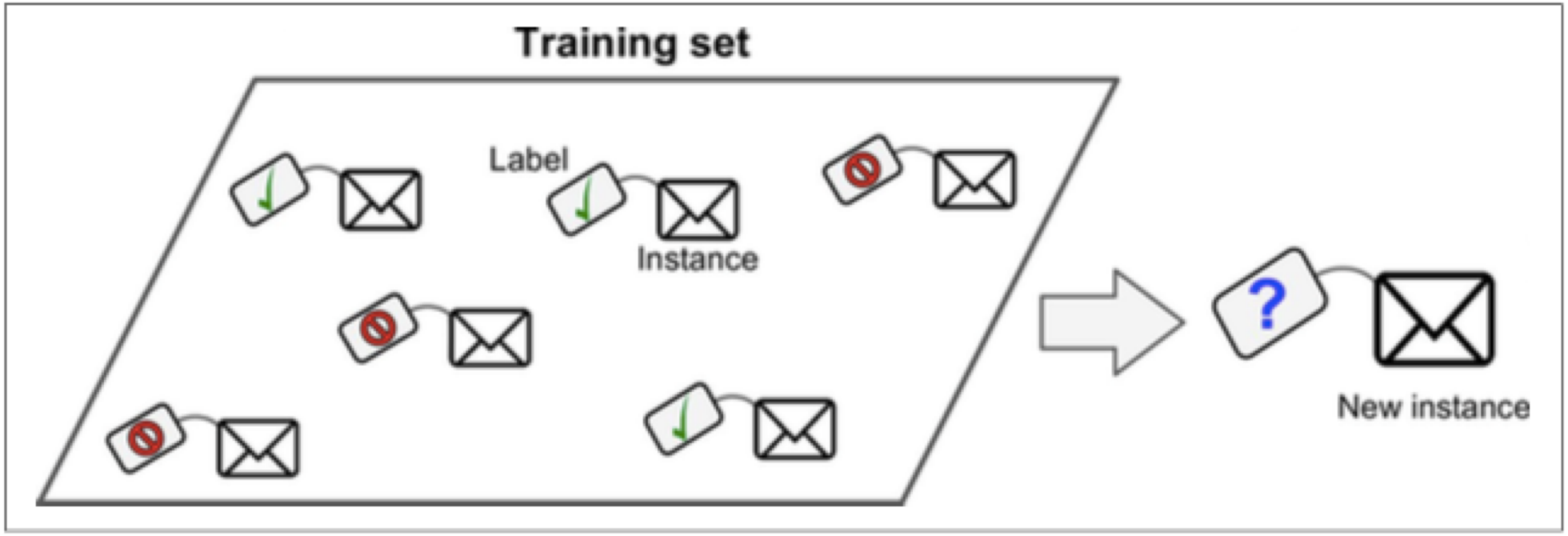


Common Supervised Learning Tasks

- Classification
  - Predicts class labels/categories
  - Example:
    - cancer/no cancer
    - husky/malamute/chiba/akita
    - 1/2/3/4/5/6/7/8/9/0
- Regression
  - Predicts continuous values
  - Example:
    - House pricing
    - Temperature



Example – Spam Filtering



Is this Classification or Regression?

Source: Hands-On Machine Learning with Scikit-Learn and TensorFlow, Aurelien Geron

Supervised Learning Algorithms

- k-Nearest Neighbors (k-NN)
- Linear Regression
- Logistic Regression
- Support Vector Machines (SVM)
- Decision Trees
- Neural Network

ML เหมาะกับงานประเภทใด

- ปัญหาที่ต้องแก้ด้วยโปรแกรมที่มีกฎและเงื่อนไขจำนวนมาก
  - e.g. rock paper scissors, handwritten digit recognition
- ปัญหาที่มีความซับซ้อนสูงและไม่สามารถหาคำตอบได้ด้วยวิธีการเดิมๆ
  - e.g. stock market prices, medical diagnosis
- ปัญหาที่ขึ้นกับสภาพแวดล้อมที่มีการเปลี่ยนแปลงอยู่ตลอดเวลา
  - e.g. spam filtering
- ปัญหาการสังเคราะห์สารสนเทศที่เป็นประโยชน์จากข้อมูลจำนวนมหาศาล
  - e.g. gene expression data

Python Tutorial

- Download Python (3.7.x)
  - https://www.python.org/getit/
- Anaconda Package
  - https://www.anaconda.com/download/
  - Spyder, Jupyter notebook, Numpy, Pandas

Python

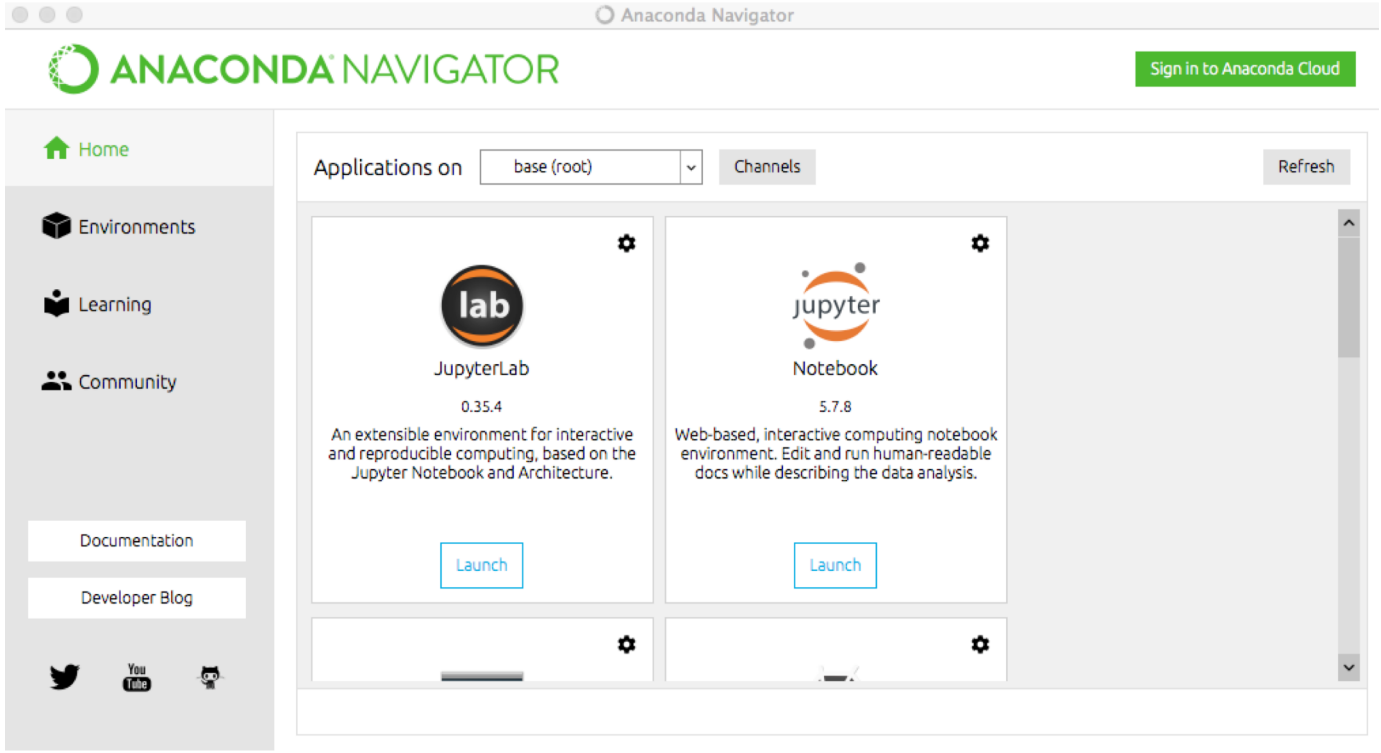
- Python is an interpreted language
- Focuses on simplicity => make programmers' lives easier
- Lots of useful libraries

Anaconda Package for Python

- Bundle of packages for scientific computing
- Numpy, Scikit-learn, Jupyter Notebook, Spyder
- Spyder is an IDE

Basic Python

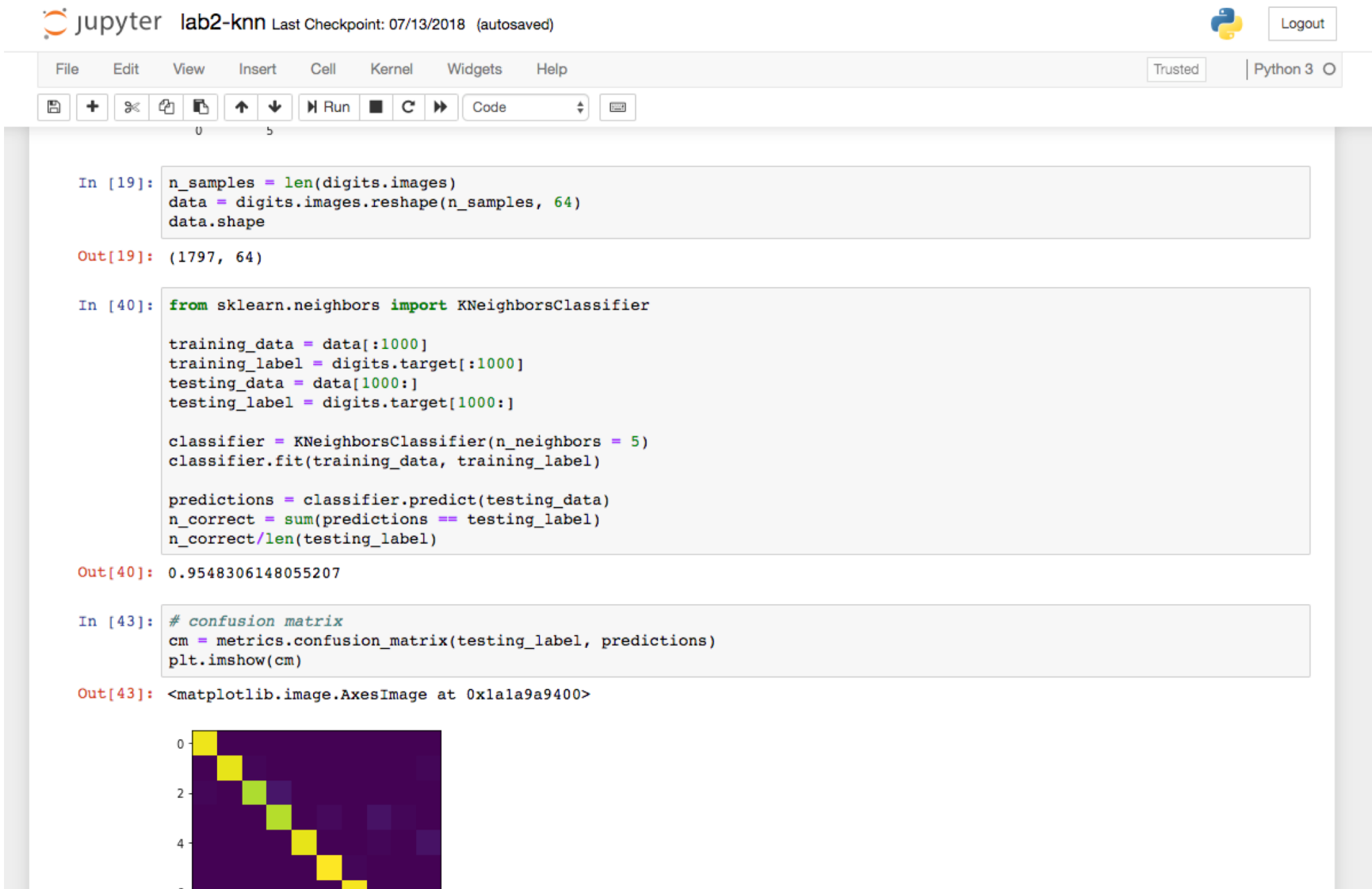
- Open up **Jupyter Notebook** from Anaconda Navigator
- Create New Python3 Notebook



Jupyter Notebook Environment

Most used shortcuts

- H show shortcuts
- Shift + Enter run cell
- A insert cell above
- B insert cell below
- D,D delete cell
- Tab code autocomplete
- Shift + Tab function docstring



Basic commands

- Try executing these commands in Jupyter NB cell:
  - 3+4
  - 1+2*3+5/2
  - 3/4
  - 3./4
  - 2**3

Variables

- Try executing these commands in Jupyter NB cell:
  - a = 3
  - a
  - print(a)
  - print(“a=”, a)
  - float(a)/7
  - nums = [1, 2, 3]
  - print(nums[0])

Math Functions

- Try executing these commands in Jupyter NB cell:
  - exp(2.0)
  - sin(2.0)

  - Error… why ?

Importing Math Library

- import math
- math.exp(2.0)
- math.pi
- math.sin(math.pi)

Different ways of importing

- import math
- import math as m
- from math import pi, exp
- from math import *

- Useful:
  - dir(math)

Data Structure

- List x = [”apple”, “banana”]
- Set x = {"apple", "banana", "cherry"}
- Tuple x = ("apple", "banana", "cherry")
- Dictionary
  - X = {
    "brand": "Ford",
    "model": "Mustang",
    "year": 1964
    }

Writing a script

```
from math import *

a = [0, pi/2, pi]
x = a[2]
if sin(x) == 0:
    print("zero")
elif sin(x) == 1:
    print("one")
else:
    print("huh")
```

Writing a script

```
from math import *

a = [0, pi/2, pi]
x = a[2]
if sin(x) == 0:
    print("zero")
elif sin(x) == 1:
    print("one")
else:
    print("huh")
```

Numpy Basics – Matrix/array

```
import numpy as np

a = np.array([[0,-1],[1,0]])
b = np.array([3,4])

print(a)
print(b)
print(a.dot(b))
```

Numpy Basics – random sampling

```
import numpy as np

a = np.array([[0,-1],[1,0]])
b = np.array([3,4])

print(a)
print(b)
print(a.dot(b))
```

Plotting with matplotlib

```
import numpy as np
import matplotlib.pyplot as plt

n = 1000
x = np.arange(n)
y = np.random.rand(n)
plt.scatter(x,y)
```

Next week

- Supervised Learning Algorithm
  - K-NN