



FACULTY OF ENGINEERING
AT SRI RACHA
.....
DEPARTMENT OF COMPUTER ENGINEERING

03603351 วิทยาศาสตร์ข้อมูลเบื้องต้น
Introduction to Data Science

Machine Learning

ผศ.ดร. กุลวดี สมบูรณ์วิวัฒน์

kulwadee@eng.src.ku.ac.th

Machine Learning คืออะไร

- The science (and art) of programming computers so they can *learn from data*.

[Machine Learning is the] field of study that gives computers the ability to learn without being explicitly programmed.

Arthur Samuel, 1959

*A computer program is said to learn from **experience E** with respect to some task **T** and some **performance measure P**, if its performance on T, as measured by P, improves with experience E.*

Tom Mitchell, 1997

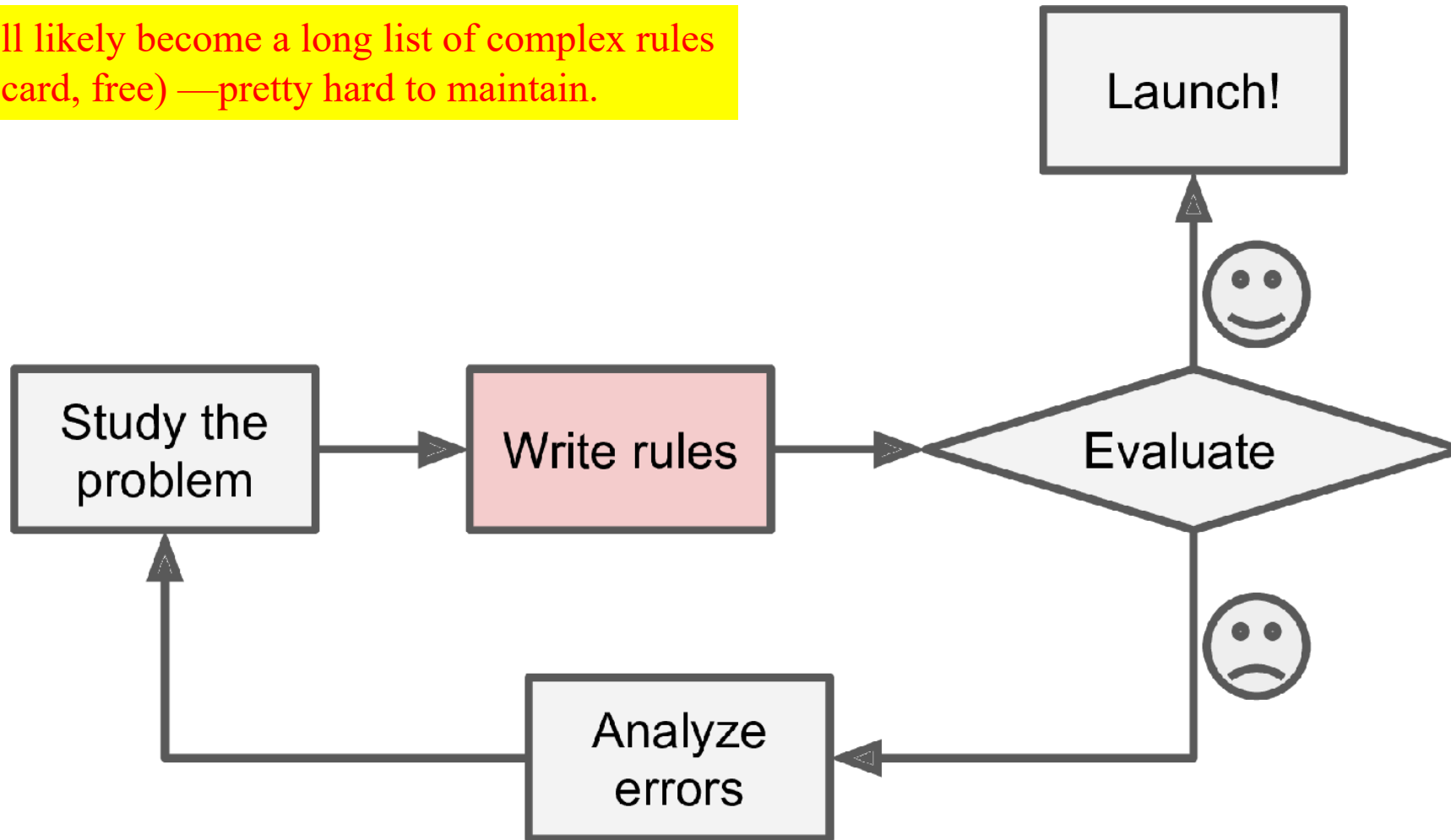
Spam Filter



- A machine learning program, that given examples of spam emails and regular emails, can learn to flag spam emails.
 - Task T = flag a new email as spam or regular
 - Experience E = training set (spam and regular emails)
 - Performance P = Accuracy (% of correctly classified emails)

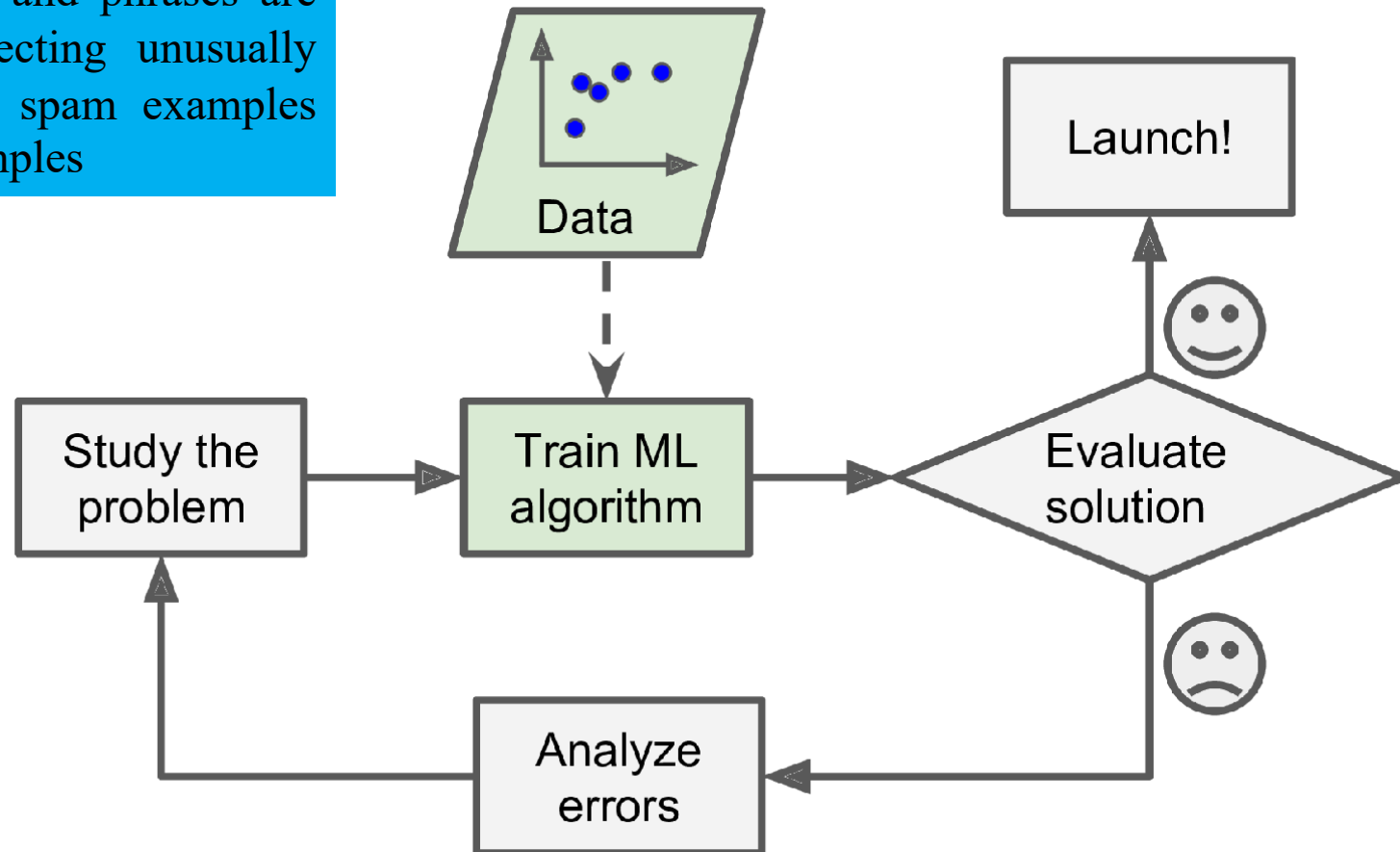
Spam Filtering Program (traditional approach)

Our program will likely become a long list of complex rules (E.g. 4U, credit card, free) —pretty hard to maintain.



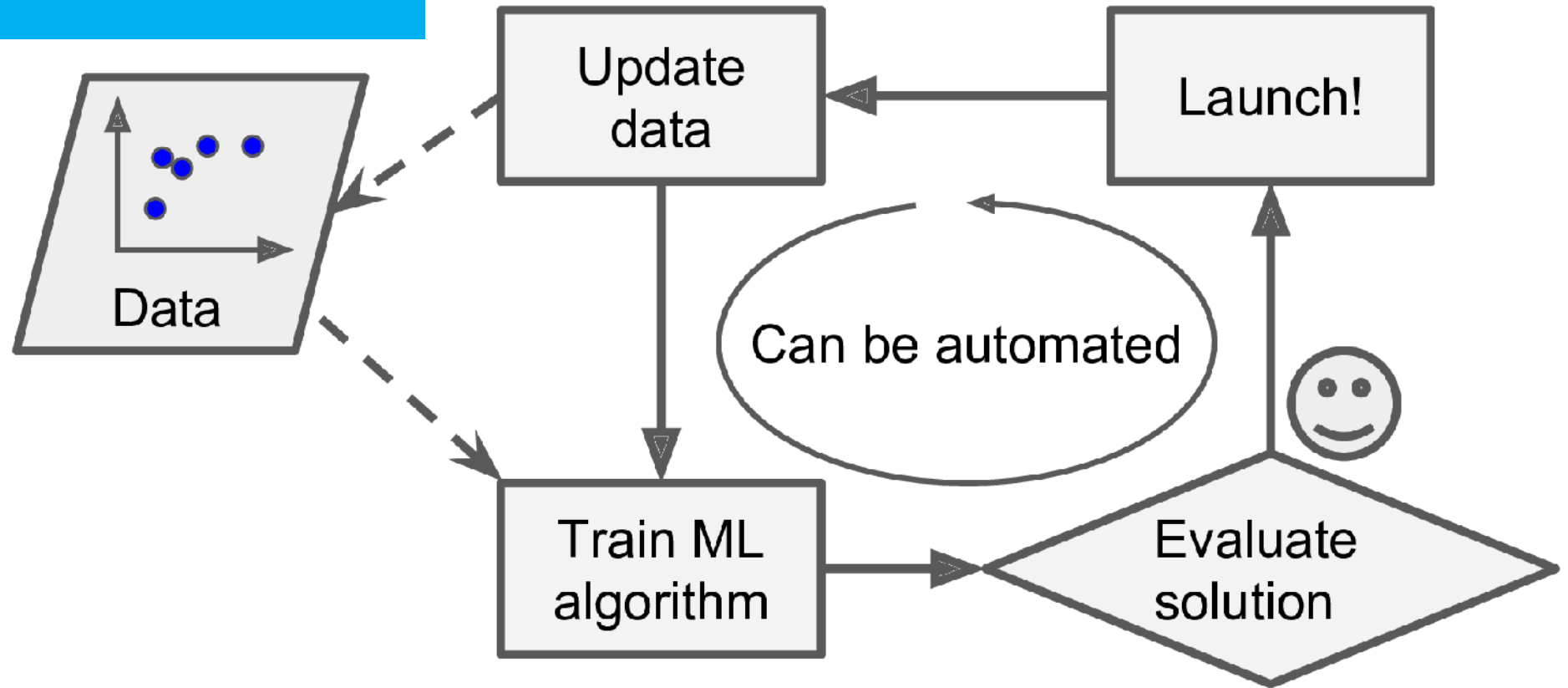
Spam Filtering Program (ML approach)

Automatically learns which words and phrases are good predictors of spam by detecting unusually frequent patterns of words in the spam examples compared to the regular emails examples



Spam Filtering Program (ML approach)

Automatically adapt to changes – automatically notices that “For U” has become unusually frequent in spam flagged by users, and it starts flagging them without your intervention.



สรุป: ML เป็นวิธีการแก้ปัญหาที่เหมาะสมกับสถานการณ์ดังนี้

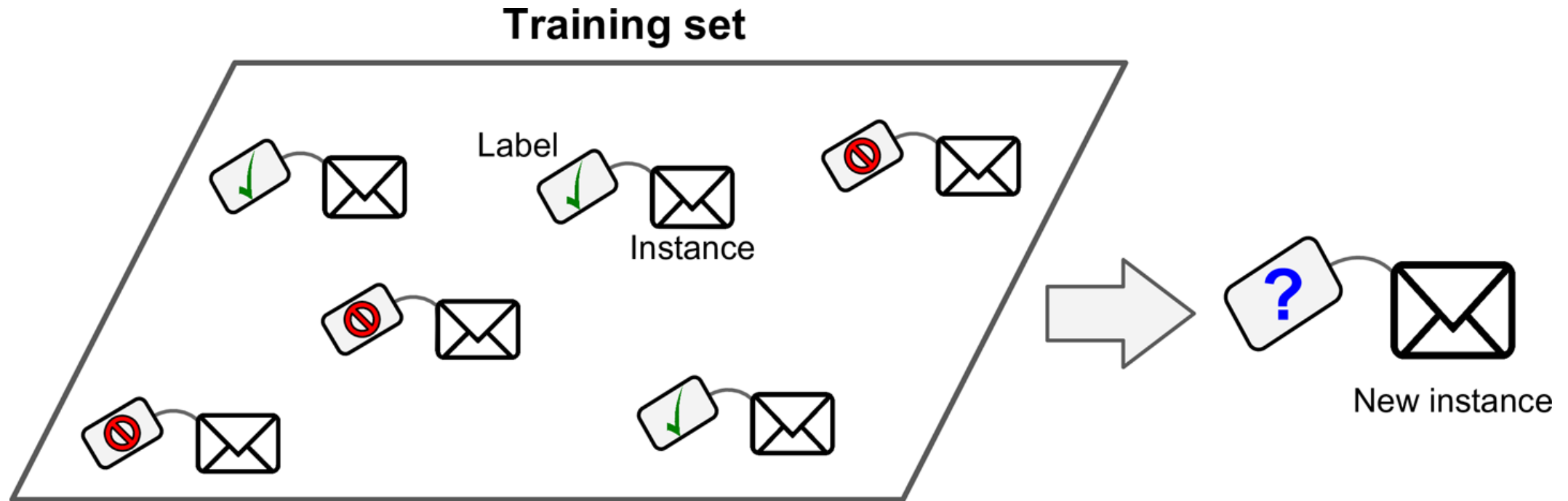
- ปัญหาที่ประกอบด้วยกฎเกณฑ์ในการแก้ปัญหาที่ซับซ้อนจำนวนมาก
- ปัญหาที่ไม่สามารถแก้ได้โดยใช้อัลกอริทึมที่มีอยู่ เช่น speech recognition, face recognition
- สภาพแวดล้อมมีการเปลี่ยนแปลงสูง

ประเภทของ ML แบ่งตามวิธีการเรียนรู้

- Supervised Learning
- Unsupervised Learning
- Semi-supervised Learning
- Reinforcement Learning

Supervised Learning

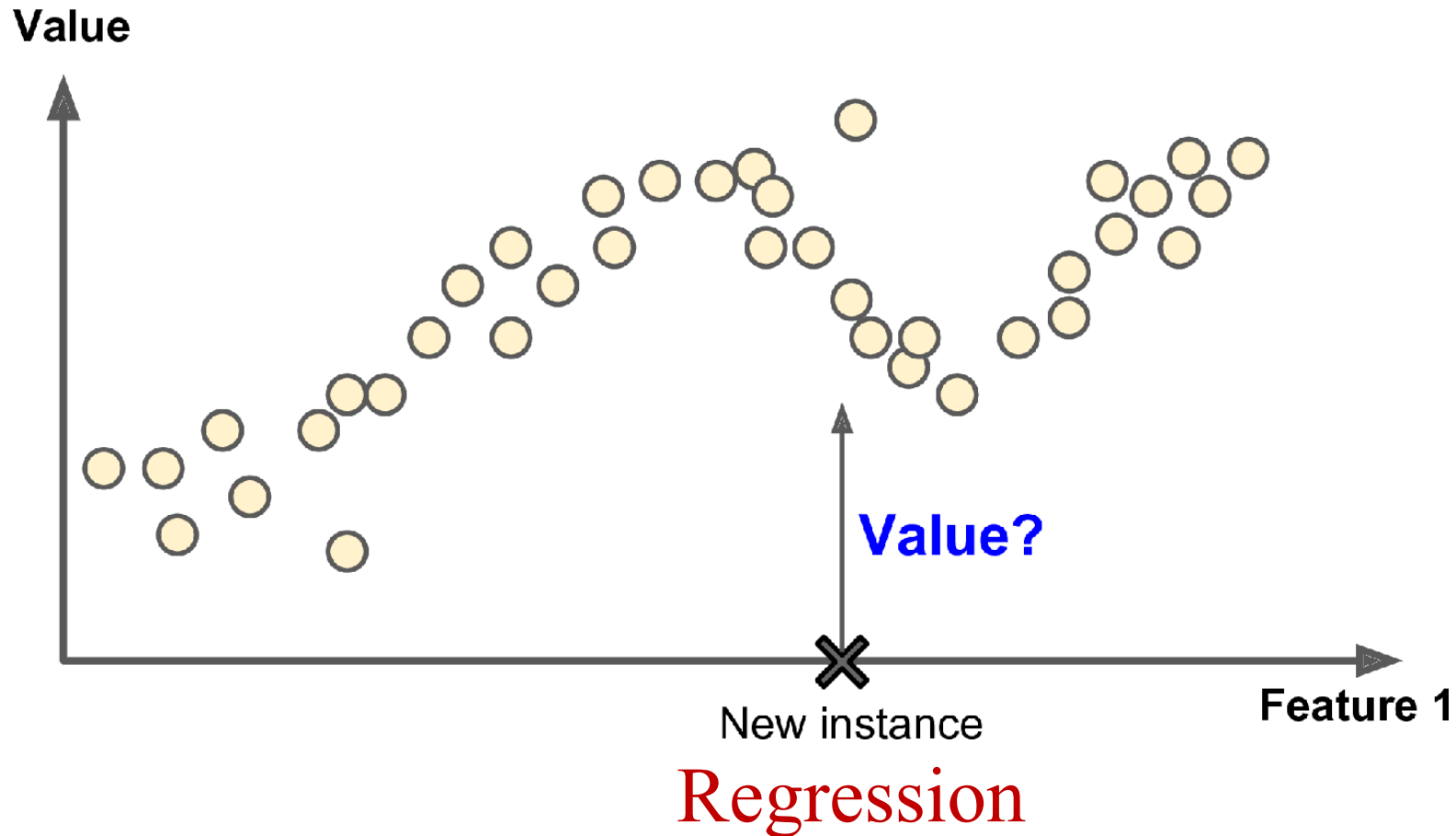
the training set you feed to the algorithm includes the desired solutions, called *labels*



Classification

Supervised Learning

the training set you feed to the algorithm includes the desired solutions, called *labels*



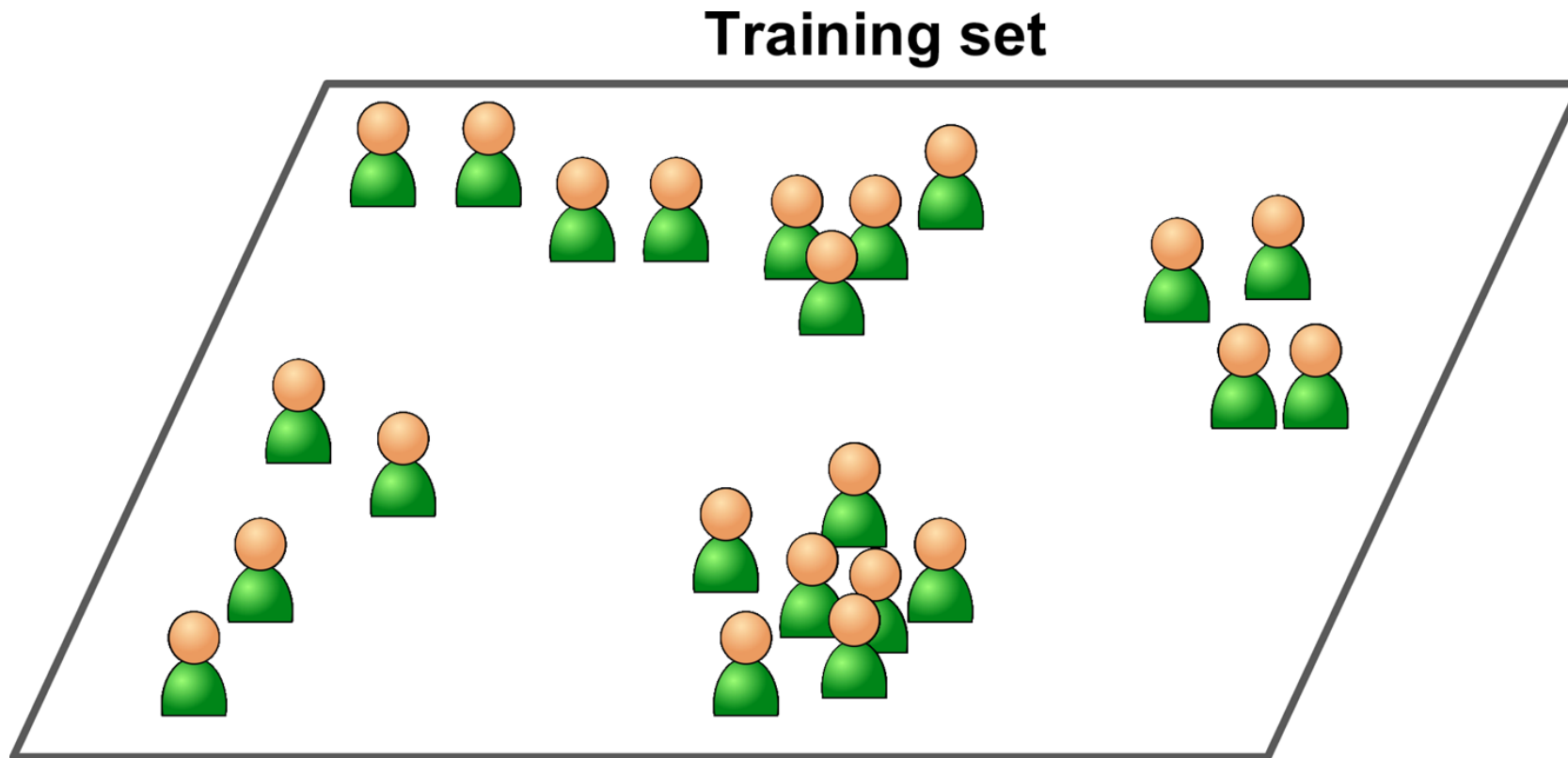
Supervised Learning

the training set you feed to the algorithm includes the desired solutions, called *labels*

- k-Nearest Neighbors
- Linear Regression
- Logistic Regression
- Support Vector Machines (SVMs)
- Decision Trees
- Random Forests
- Neural networks

Unsupervised Learning

the training set you feed to the algorithm
does not include the desired solutions,
called *labels*



Unsupervised Learning

the training set you feed to the algorithm **does not includes** the desired solutions, called *labels*

- Clustering

K-Means, Hierarchical Cluster Analysis (HCA)

- Anomaly detection and novelty detection

One-class SVM, Isolation Forest

- Visualization and dimensionality reduction

Principal Component Analysis (PCA),

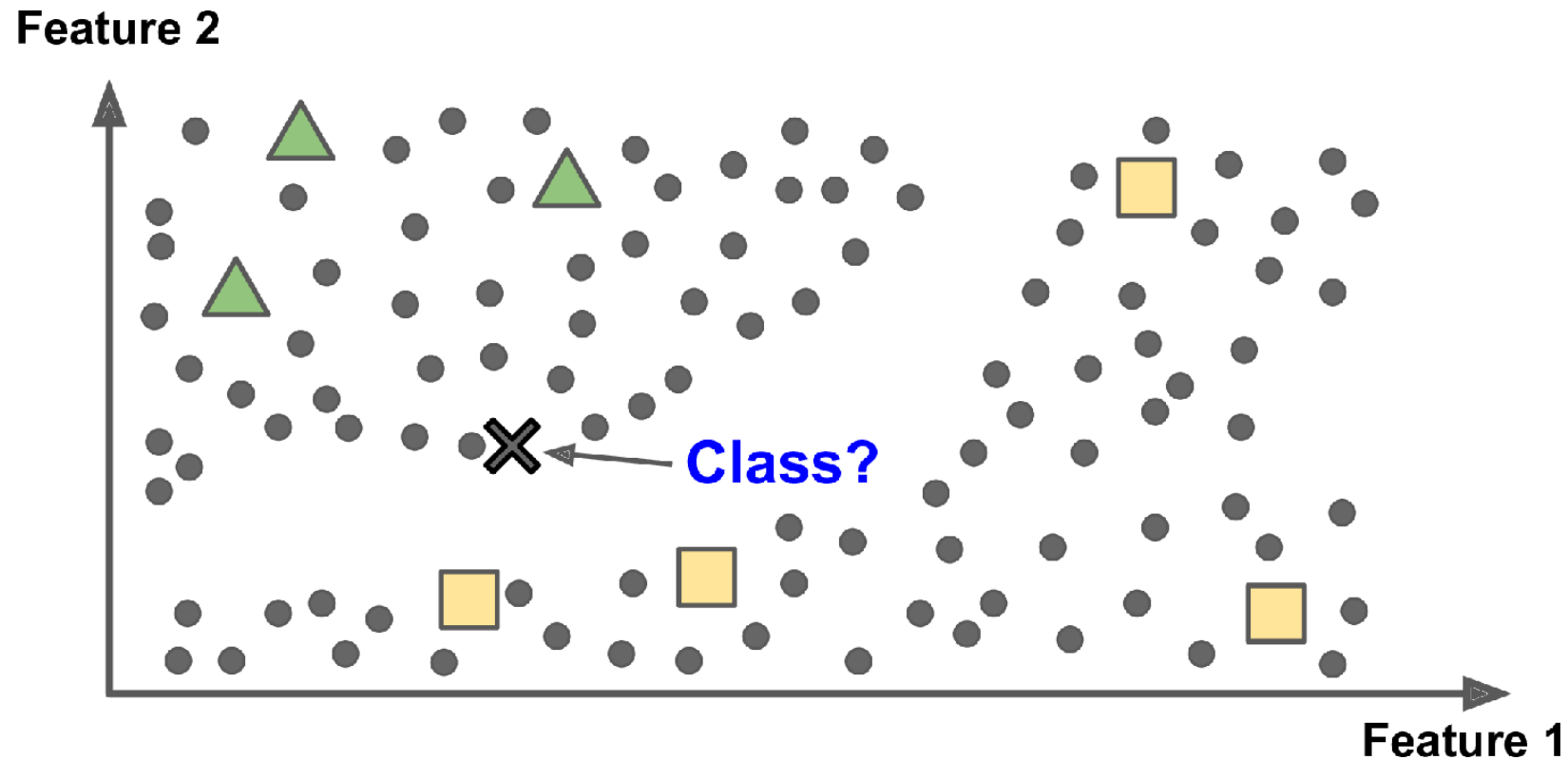
t-Distributed Stochastic Neighbor Embedding (t-SNE)

- Association rule learning

Apriori

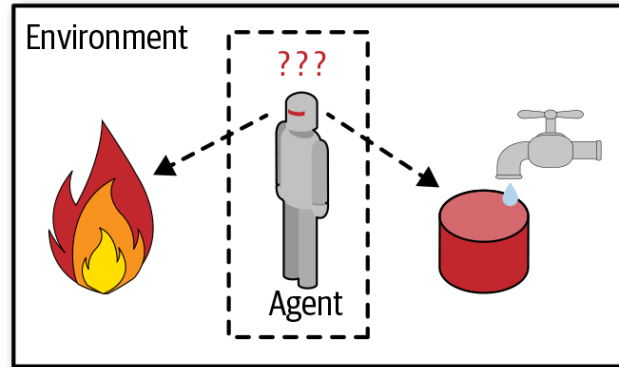
Semi-supervised Learning

the training set you feed to the algorithm
is **partially labelled**

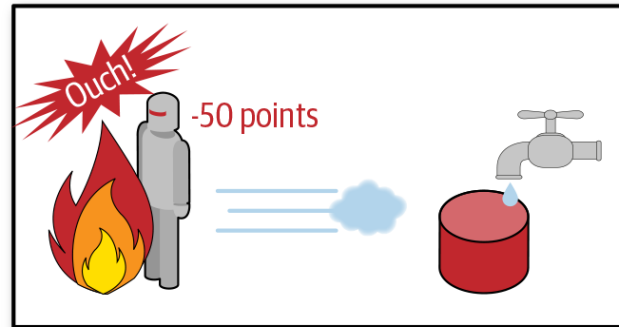


Reinforcement Learning

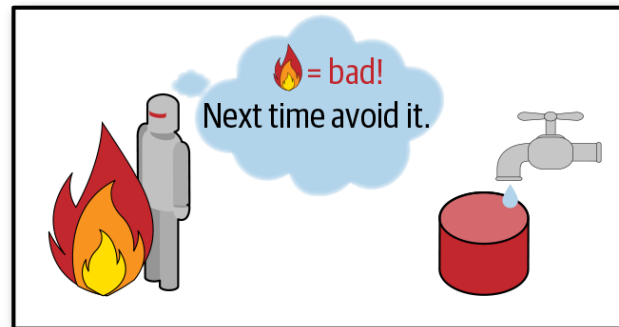
The learning system, called an **agent**, can observe the **environment**, select and perform **actions**, and get **rewards** in return (or penalties in the form of negative rewards). It must then **learn by itself what is the best strategy, called a policy**, to get the most reward over time.



- 1 Observe
- 2 Select action using policy



- 3 Action!
- 4 Get reward or penalty



- 5 Update policy (learning step)
- 6 Iterate until an optimal policy is found

ScikitLearn Project 1: MNIST Dataset

- Hand-written image of digits 0-9
- Each image : 8x8 pixel gray-scale
- MNIST_Data.ipynb
 - Load dataset
 - Explorte dataset
 - Dimensionality Reduction with PCA
 - Pre-process data
 - Scaling
 - Train/Test Split
 - Unsupervised Learning with K-Means
 - Supervised Learning with SVM

