Applied Machine Learning

Course Overview

Instructor:

Yi Fang (yfang@nyu.edu)

Lab Instructors:

Prince Steven Annor (prince.annor@nyu.edu)

Start with two examples…

- Google Assistant: <u>Link</u>

- Deep Fake: <u>Link</u>

Disclaimer: The views and opinions expressed in YouTube videos linked above are those of the authors and do not reflect the official policy or position of the class instructors and NYUAD.

Why machine learning?

Critical importance of addressing urban challenges: disease, crime, terrorism, poverty, environment…

Increasing size and complexity of available data, thanks to the rapid growth of new and transformative technologies.


Much more computing power, and scalable data analysis methods, enable us to extract actionable information from all of this data.

Machine learning techniques have become increasingly essential for data analysis, and for the development of new, practical information technologies that can be directly applied to address critical challenges in many areas (Engineering, Science, Social Science, Economics and many more)

More Applications of ML

- Inferring urban dynamics from heterogeneous data
- Computer vision: pedestrian/traffic counts, security/law enforcement (face recognition), traffic accident detection, remote sensing (air content, IR, etc.)
- Street noise (decomposition, localization, classification)
- Economic patterns detection and prediction
- Health pattern detection and prediction
- Energy usage prediction
- Traffic modeling and prediction
- Land use classification
- 3-D landscape recognition
- Event detection from urban activity
- Detecting trends from social media
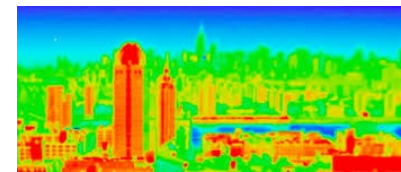
Infrared data from CUSP's Urban Observatory

Some motivating examples

Early detection of emerging disease outbreaks

Discovering new "best practices" for patient care

Substance abuse and overdose surveillance

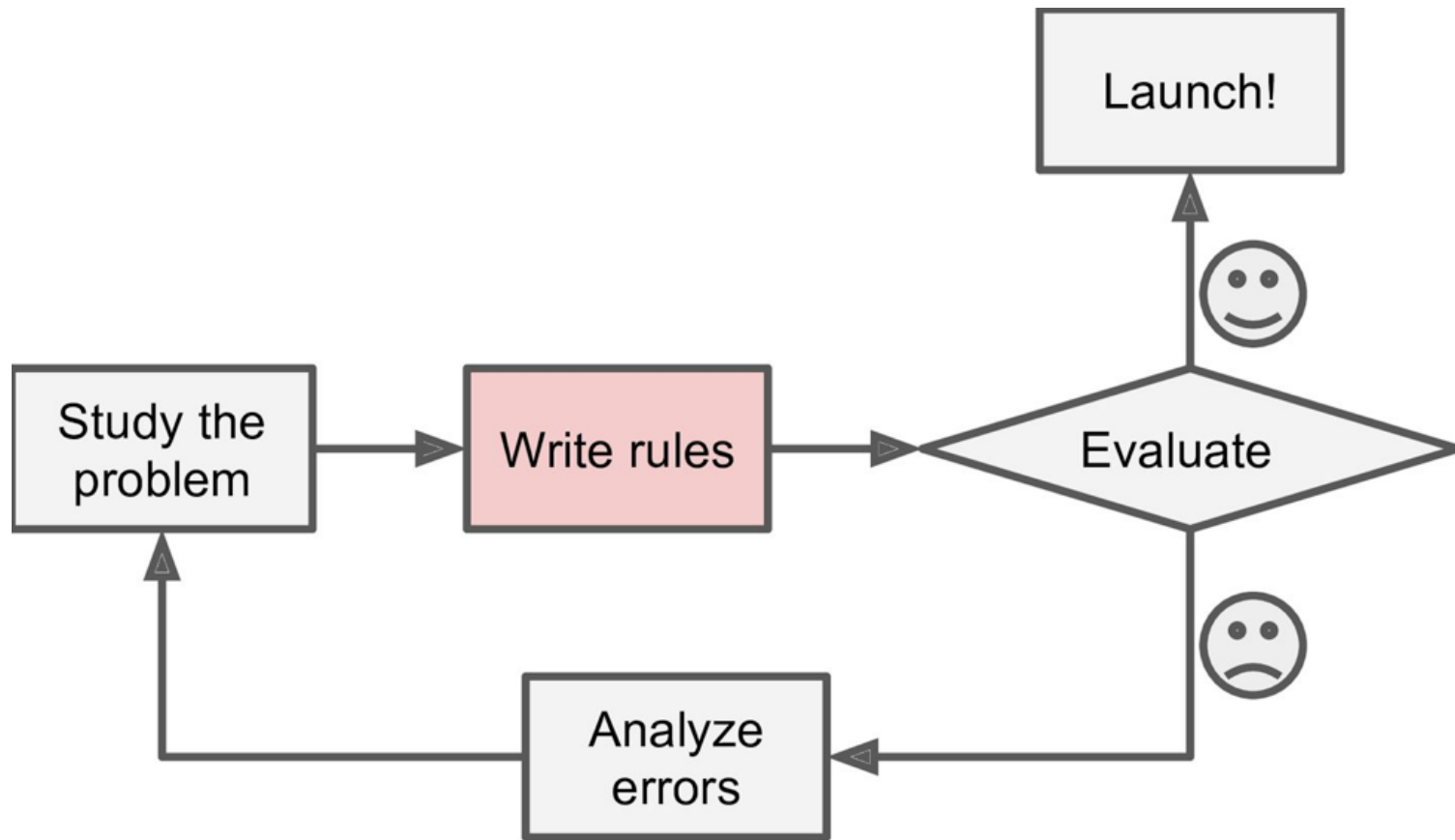Preventing rat infestations (using "311" service calls)

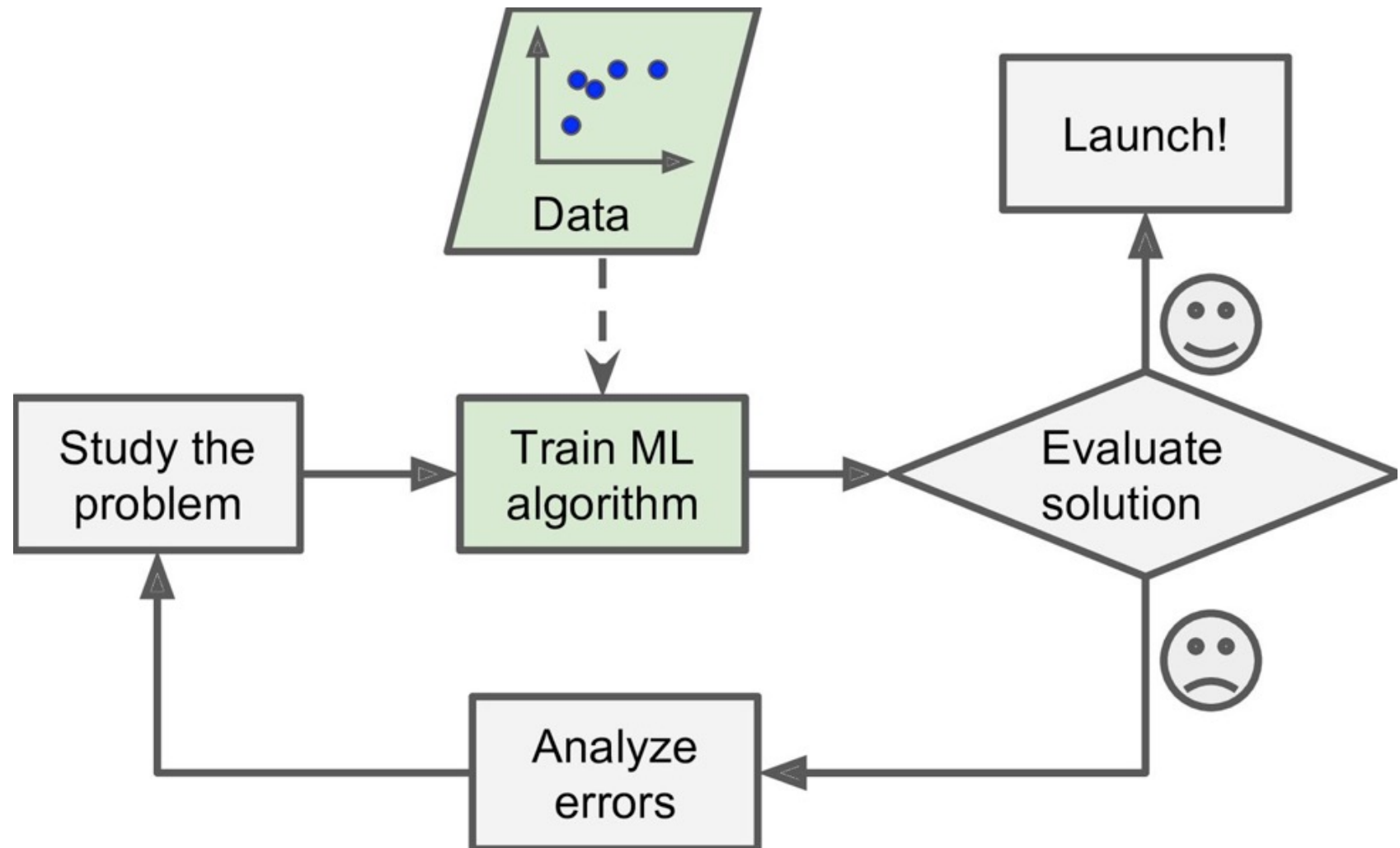Predicting civil unrest (using Twitter data)

Preventing violent crime (in Chicago & Pittsburgh)

Machine Learning Methods V.S.

Traditional Ones
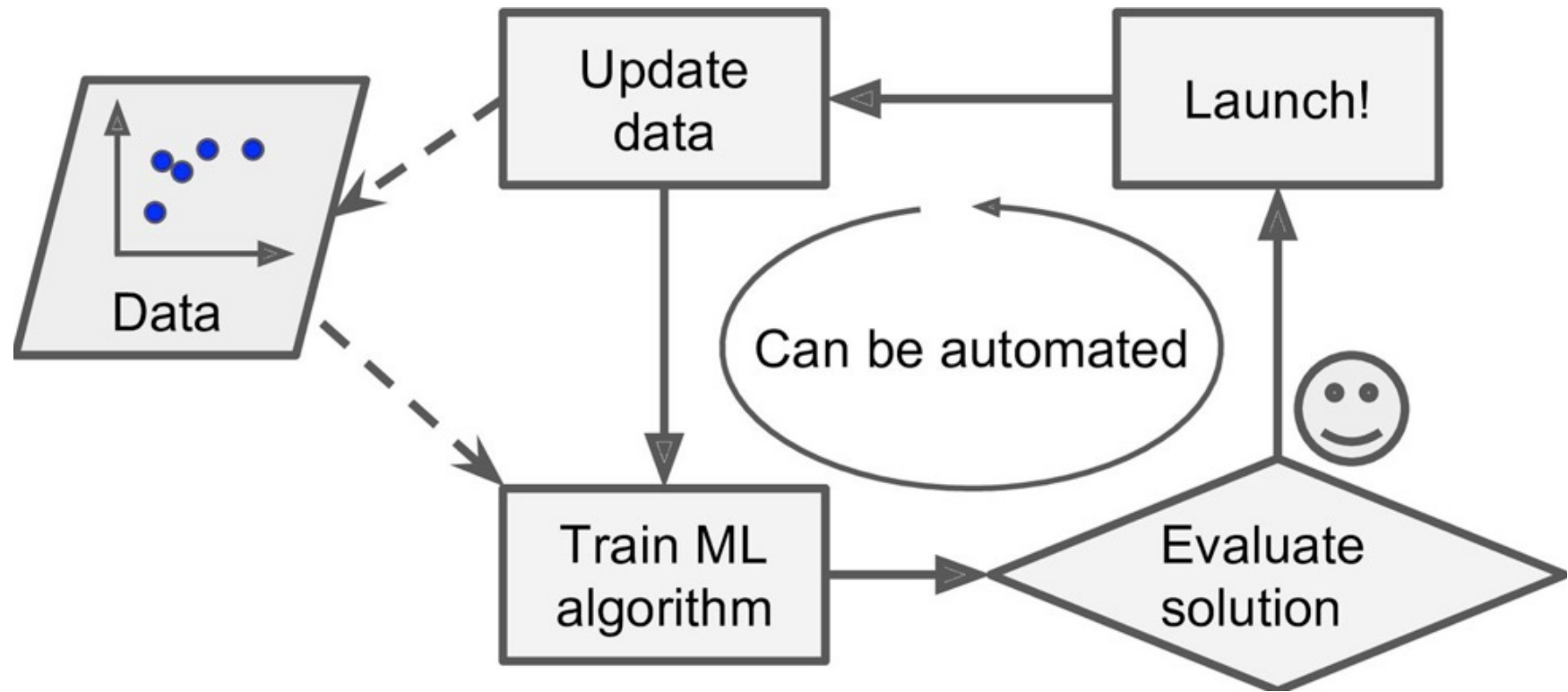
Prof. Yi Fang (yfang@nyu.edu)

Traditional Method
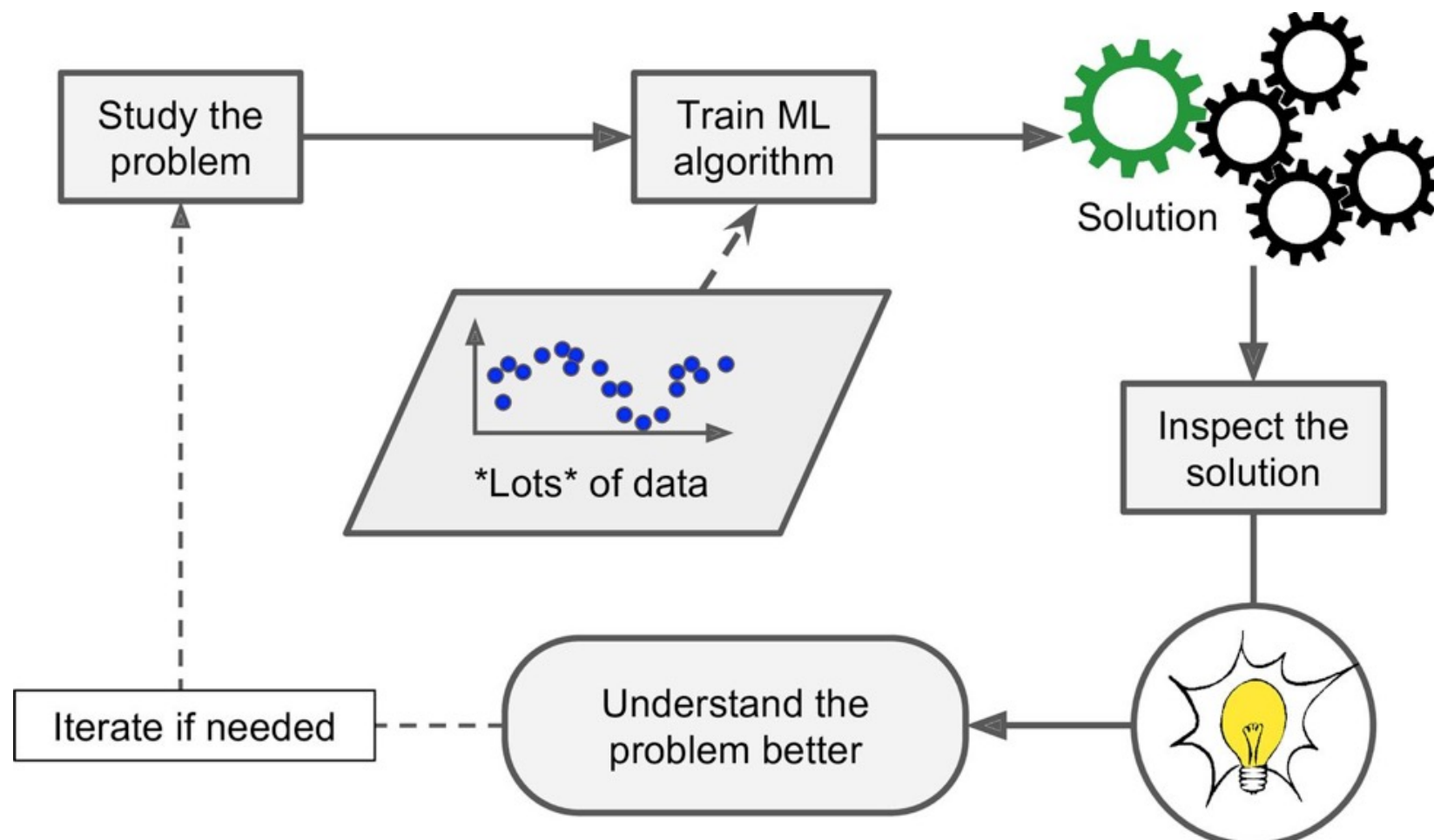
Machine Learning Method

New data coming in …

Help human learns

What is machine learning?

Machine Learning (ML) is the study of systems that improve their performance with experience (typically by **learning** from data).

"A computer program is said to learn from experience E wrt. some class of tasks T and performance measure P, if its performance at tasks in T as measured by P improves with experience." (T. Mitchell)

"Learning denotes changes in the system that are adaptive in the sense that they enable it to do a task, or tasks drawn from the same population, more efficiently and effectively next time." (H. Simon)

ML vs. Humans

Machine Learning (ML) is the study of systems that improve their performance with experience (typically by **learning** from data).

"A computer program is said to learn from experience E wrt. some class of tasks T and performance measure P, if its performance at tasks in T as measured by P improves with experience." (T. Mitchell)

"Learning denotes changes in the system that are adaptive in the sense that they enable it to do a task, or tasks drawn from the same population, more efficiently and effectively next time." (H. Simon)

Computers now consistently beat the top players in the world at checkers (1995), chess (1997), and Go (2016), as well as the game show Jeopardy (2011).

ML vs. Humans

Machine Learning (ML) is the study of systems that improve their performance with experience (typically by **learning** from data).

"A computer program is said to learn from experience E wrt. some class of tasks T and performance measure P, if its performance at tasks in T as measured by P improves with experience." (T. Mitchell)

"Learning denotes changes in the system that are adaptive in the sense that they enable it to do a task, or tasks drawn from the same population, more efficiently and effectively next time." (H. Simon)

ML methods have had tremendous success on tasks including control (self-driving cars), image recognition, speech recognition, recommender systems, machine translation, etc.

CMU's "Boss", winner of the DARPA Urban Challenge

Examples of performance metrics

Machine Learning (ML) is the study of systems that improve their performance with experience (typically by **learning** from data).

"A computer program is said to learn from experience E wrt. some class of tasks T and performance measure P, if its performance at tasks in T as measured by P improves with experience." (T. Mitchell)

Task	Performance metric	Experience
Play checkers	Percentage of wins vs. given opponent	Games previously played w/ outcomes
Recognize handwritten digits	Percentage of correct recognitions	Set of digit writing w/ labels
Control a self-driving car	Average speed in given conditions provided that safety standards are met	Previous driving record w/ evaluation
Predict stock prices	Average prediction accuracy	History of stock prices

Table credit: Stanislav Sobolevsky

ML as Optimization

Machine Learning (ML) is the study of systems that improve their performance with experience (typically by **learning** from data).

"A computer program is said to learn from experience E wrt. some class of tasks T and performance measure P, if its performance at tasks in T as measured by P improves with experience." (T. Mitchell)

1. Select performance metric and dataset to evaluate it

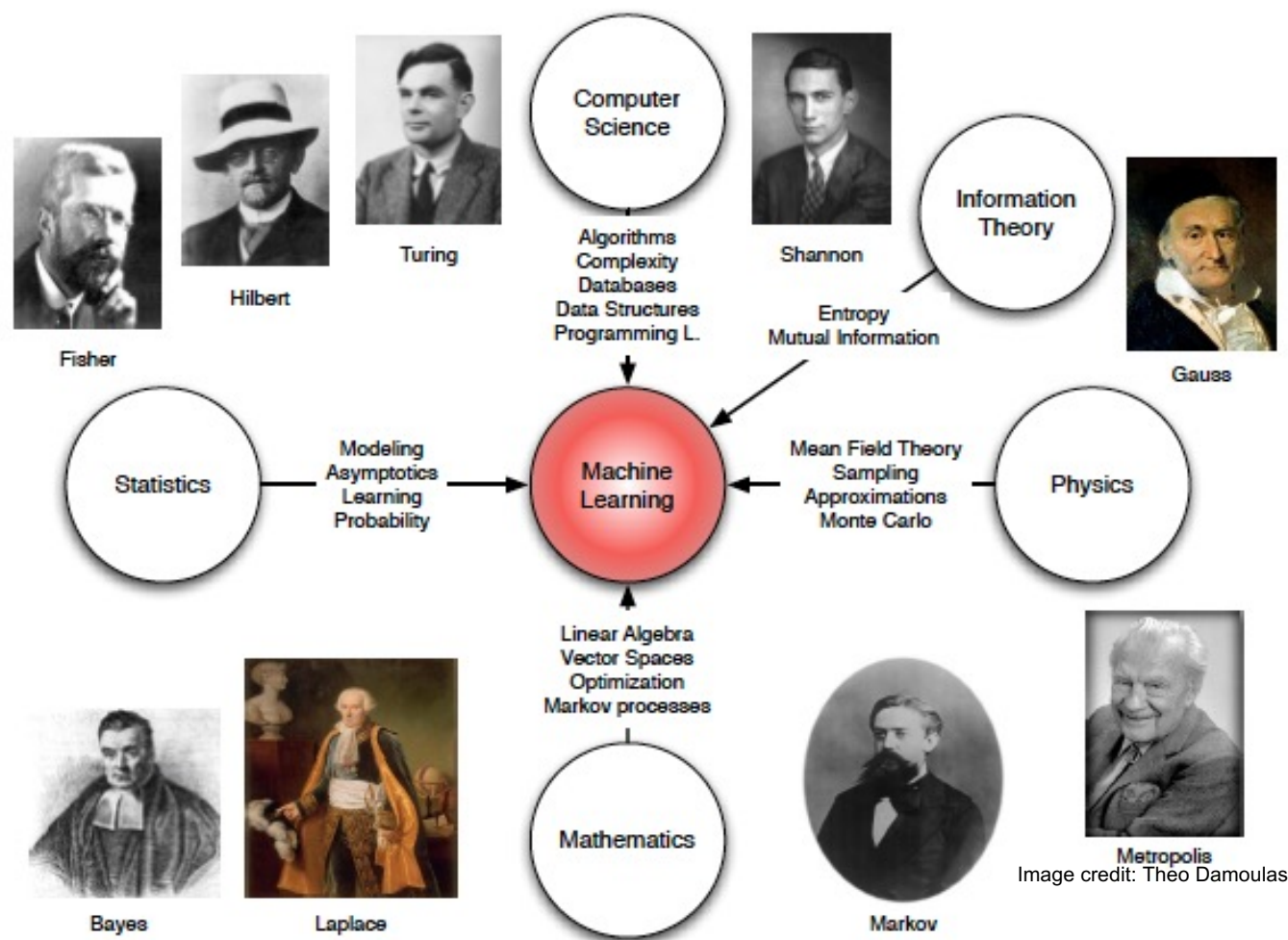2. Pick a machine learning model depending on the unknown parameters to learn

3. Look for the set of model parameters that optimize the given performance metric

4. Evaluate different models and finally pick the best one

Figure credit: Stanislav Sobolevsky

ML draws from many disciplines

Image credit: Theo Damoulas

Also cognitive psychology, evolution, economics, neuroscience, and many more!

ML and related fields

Machine Learning (ML) is the study of systems that improve their performance with experience (typically by **learning** from data).

Artificial Intelligence (AI) is the science of automating complex behaviors that normally require human intelligence: vision, language understanding, learning, problem solving, decision making, etc.

Data Mining (DM) is the process of extracting useful information from massive quantities of complex data.

I would argue that these are not three distinct fields of study! While each has a slightly different emphasis, there is a tremendous amount of overlap in the problems they are trying to solve and the techniques used to solve them.

Many of the techniques we will learn are **statistical** in nature, but are very different from classical statistics.

ML/AI/DM systems and methods:

Scale up to large, complex data
Learn and **improve** from experience
Perceive and **change** the environment
Interact with humans or other agents
Explain inferences and decisions
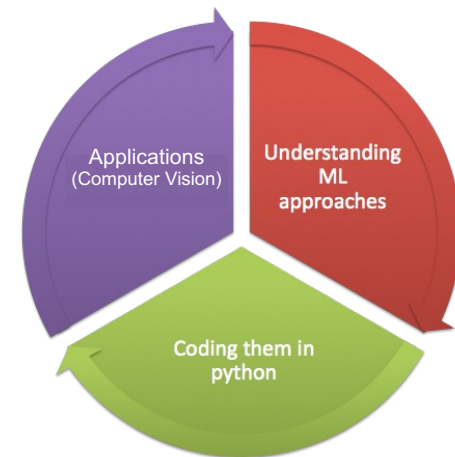**Discover** new and useful patterns

Course objectives

Goals: familiarize students with modern machine learning methods and show how they can be effectively applied to visual data.

More precisely, this course is intended to enable students to:

1) Understand the **motivation** behind different machine learning methods and their **applicability** in a given practical context.

2) Be able to **implement** methods adapted to the problem at hand using existing software libraries.*

3) Know how to **interpret** the results appropriately.

* This course includes a significant programming component, with **Python** as the primary programming language.

Please bring your laptop (Mac or Ubuntu) and follow along.

Applications (Computer Vision)

Understanding ML approaches

Coding them in python

Image credit: Stanislav Sobolevsky

Structure of the course

- Lectures: Core ML methods with visual data analysis.

- Seven Projects: Five non-deep learning projects, Two deep learning projects.

- First Part of Course: Focus on standard ML techniques.

- Second Half of Course: Emphasis on deep learning in computer vision.

- Homework: Ungraded assignments for practice.

- Final Exam: Comprehensive assessment.

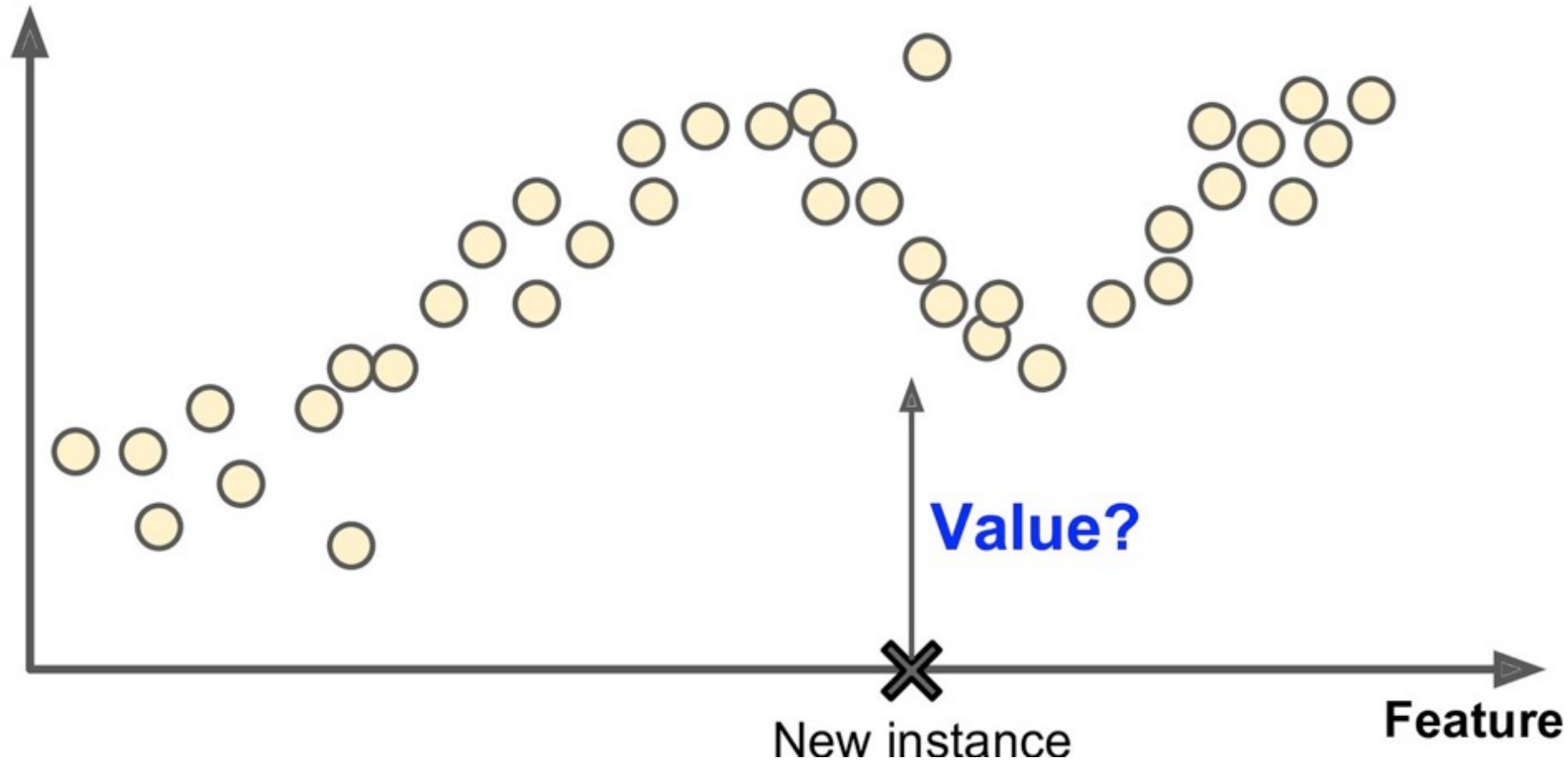# Machine Learning

Overview of ML Methods

Supervised Learning

Labeled training set for supervised learning (classification)
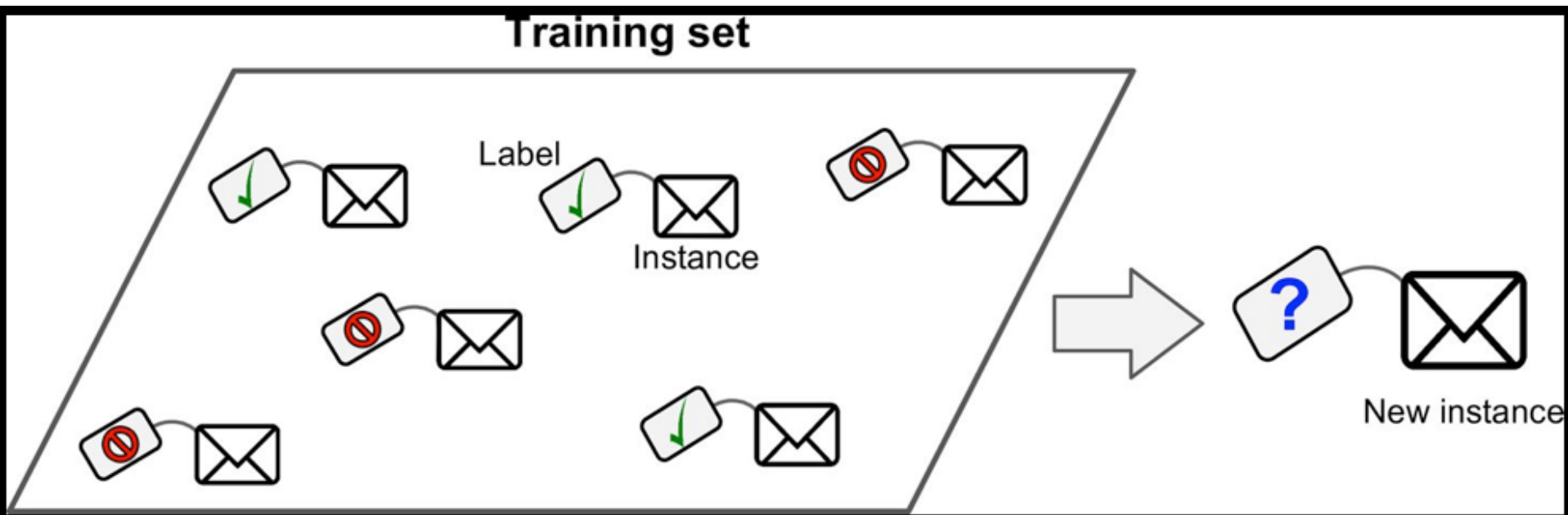
Labeled training set for supervised learning (regression)

Unsupervised Learning

Unlabeled training set for unsupervised learning

Training set
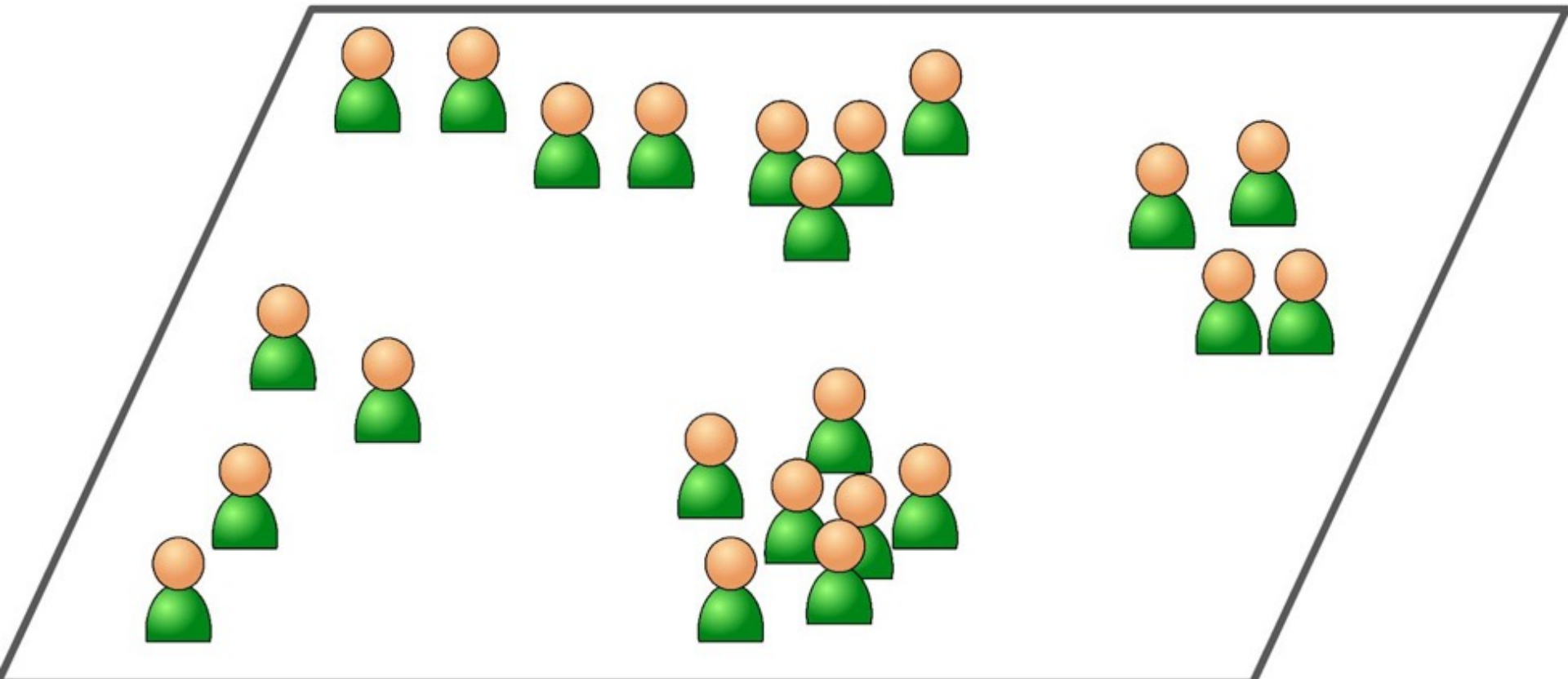
Unlabeled training set for unsupervised learning (clustering)
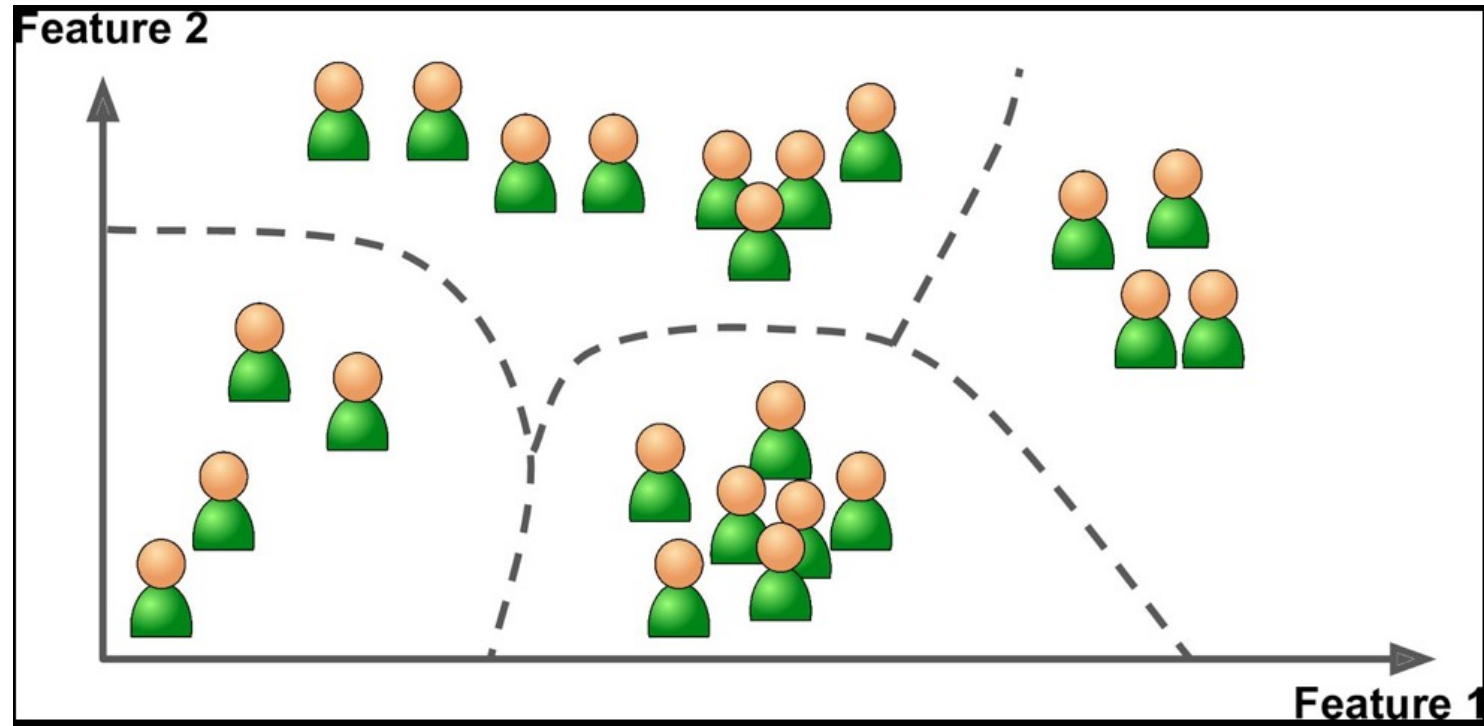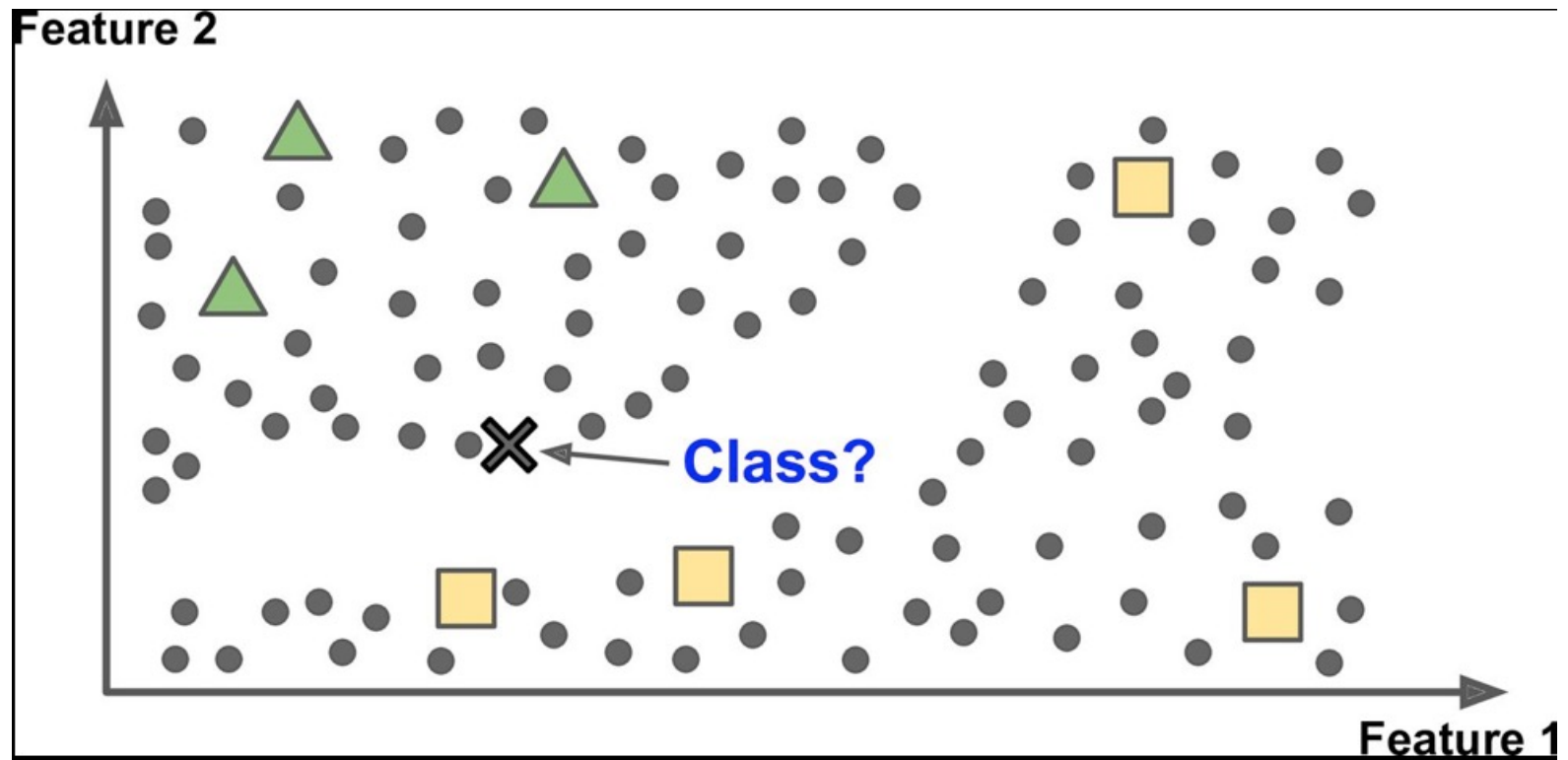
Unsupervised learning (Visualization)

Legend:
- + cat
- ○ automobile
- · truck
- · frog
- × ship
- □ airplane
- ◇ horse
- △ bird
- ▽ dog
- ▷ deer
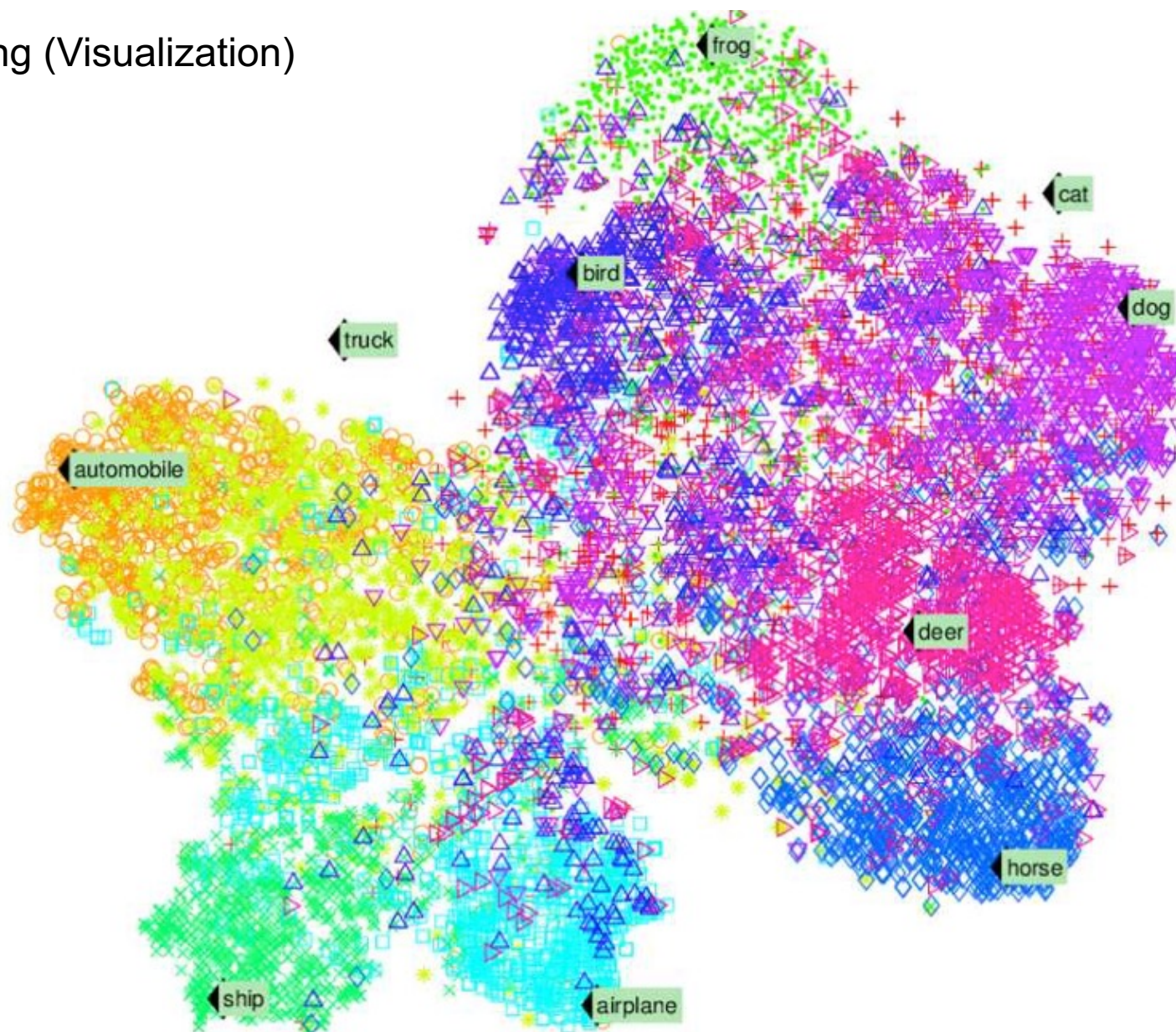
Unsupervised learning (Anomaly detection)
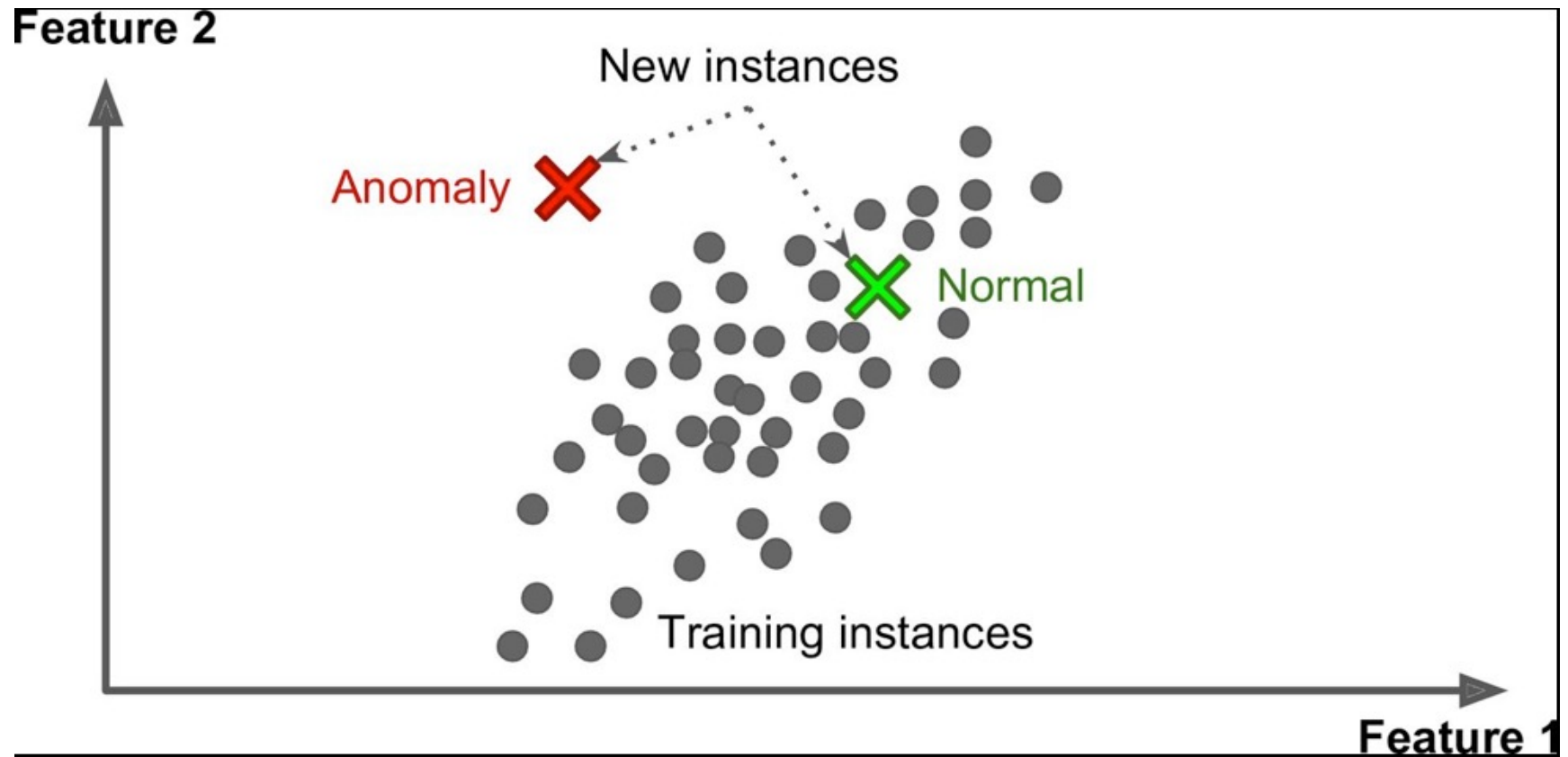
Semi-supervised Learning

Semi-supervised learning
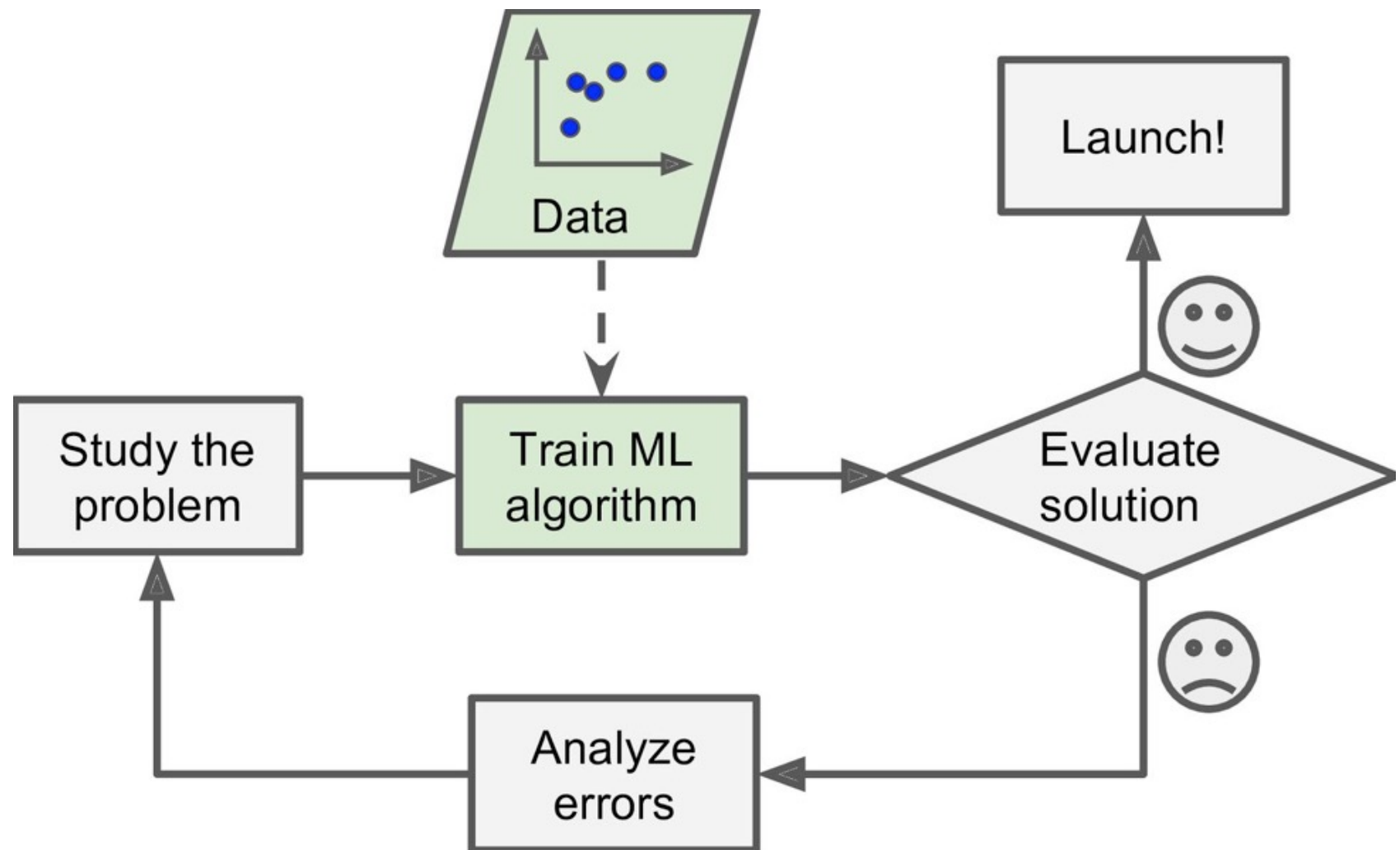
Reinforcement Learning

Reinforcement Learning
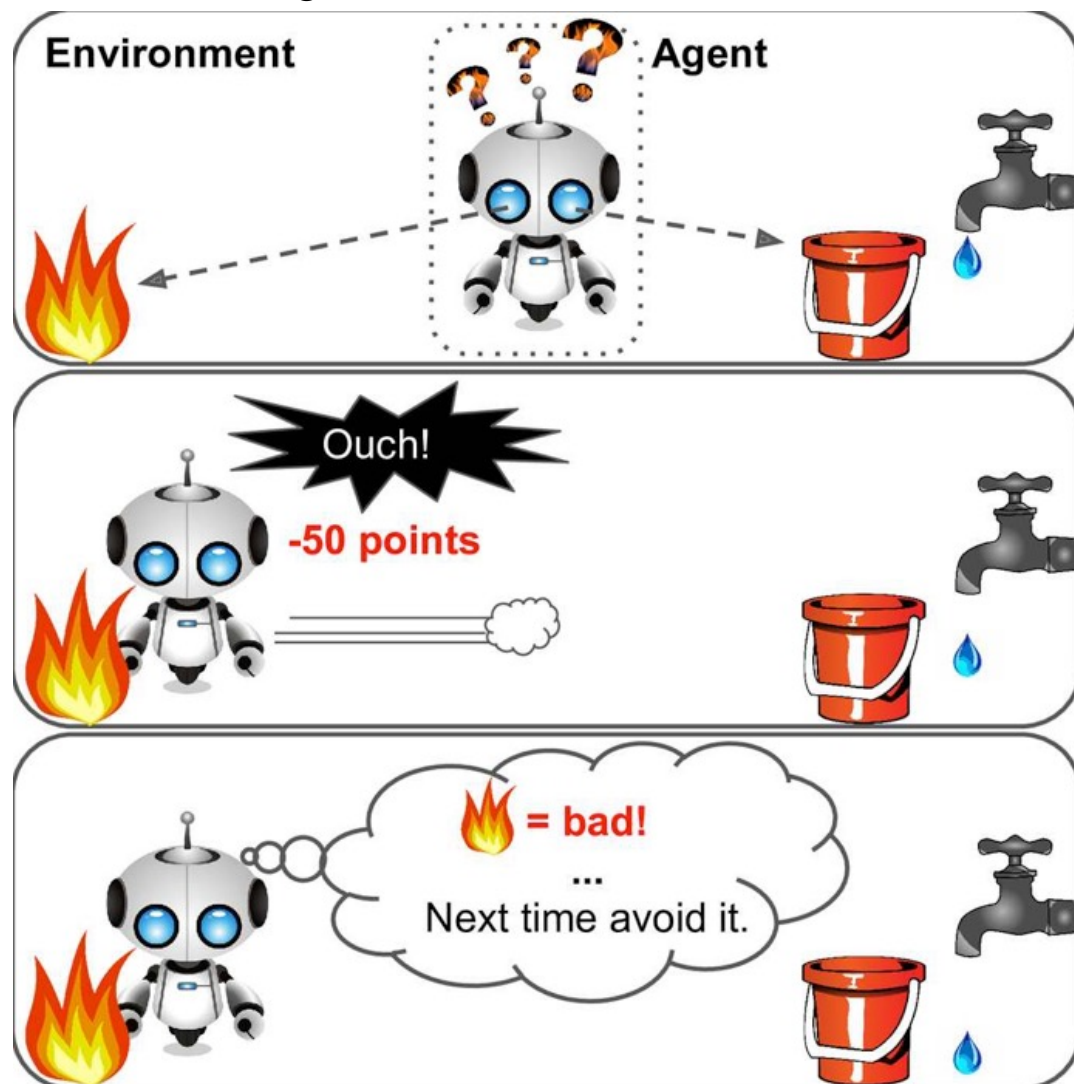
Demo

Disclaimer: The views and opinions expressed in YouTube videos linked above are those of the authors and do not reflect the official policy or position of the class instructors and NYUAD.
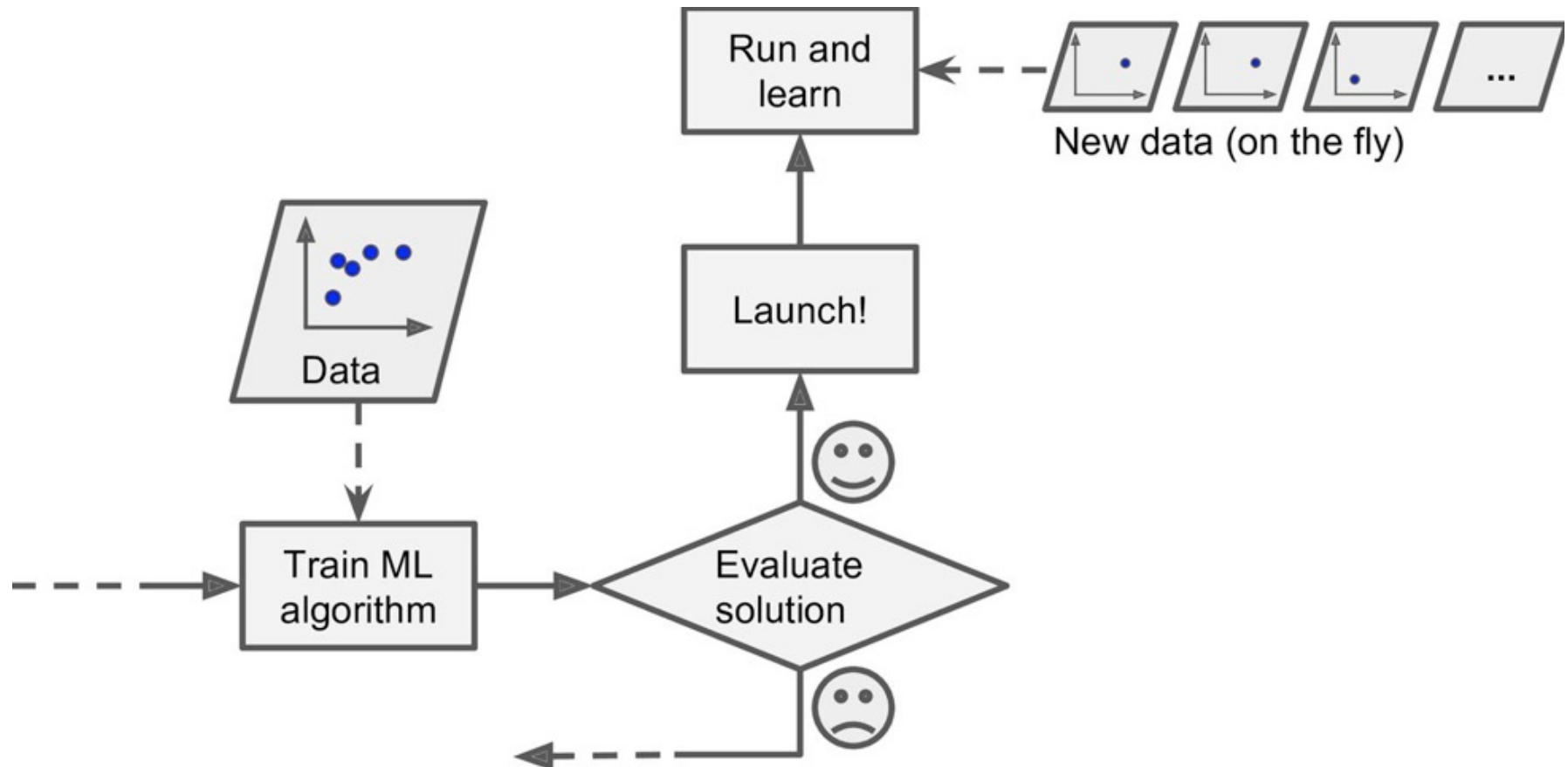
Batch Learning
VS
Online Learning

Batch Learning (offline)
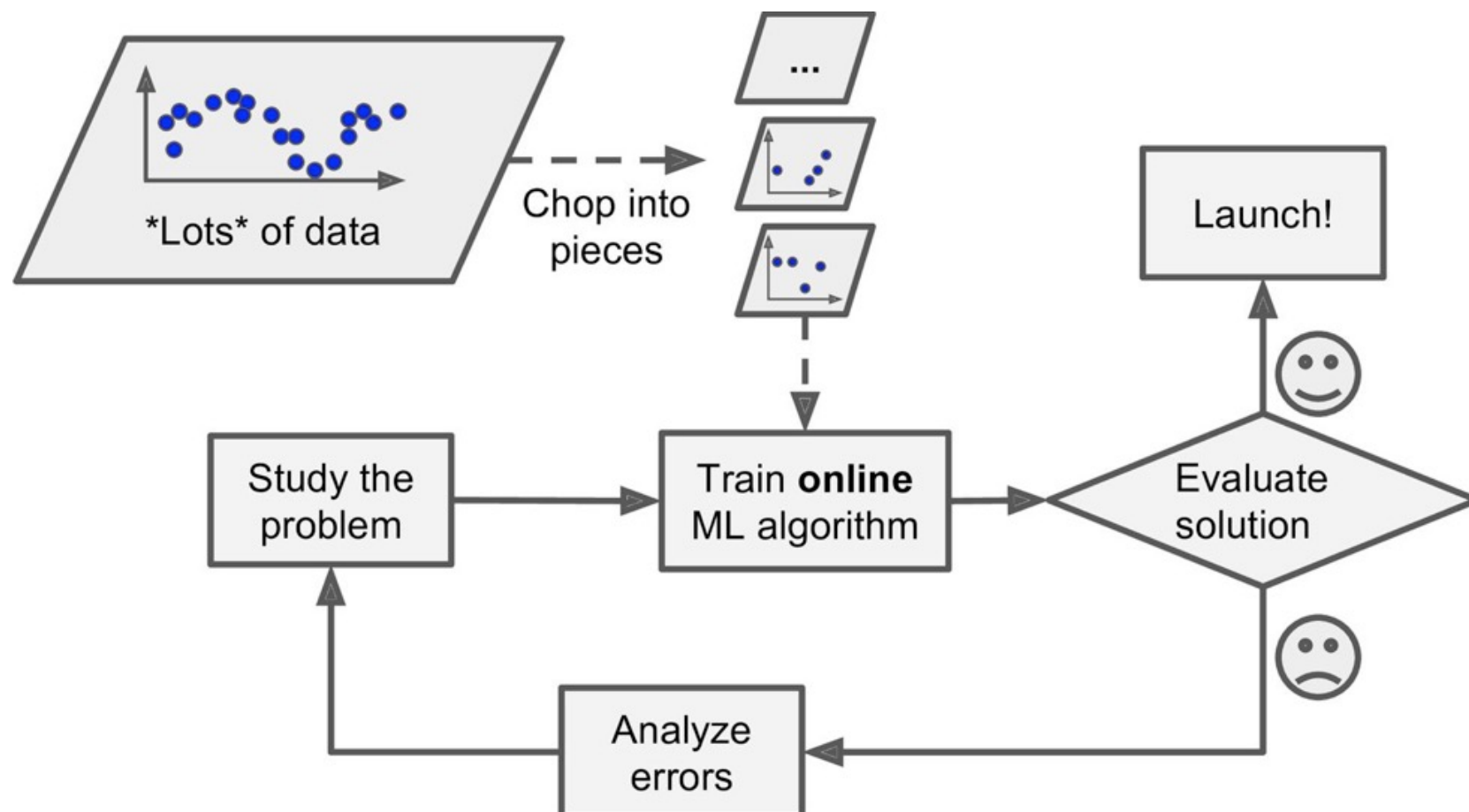
Online Learning
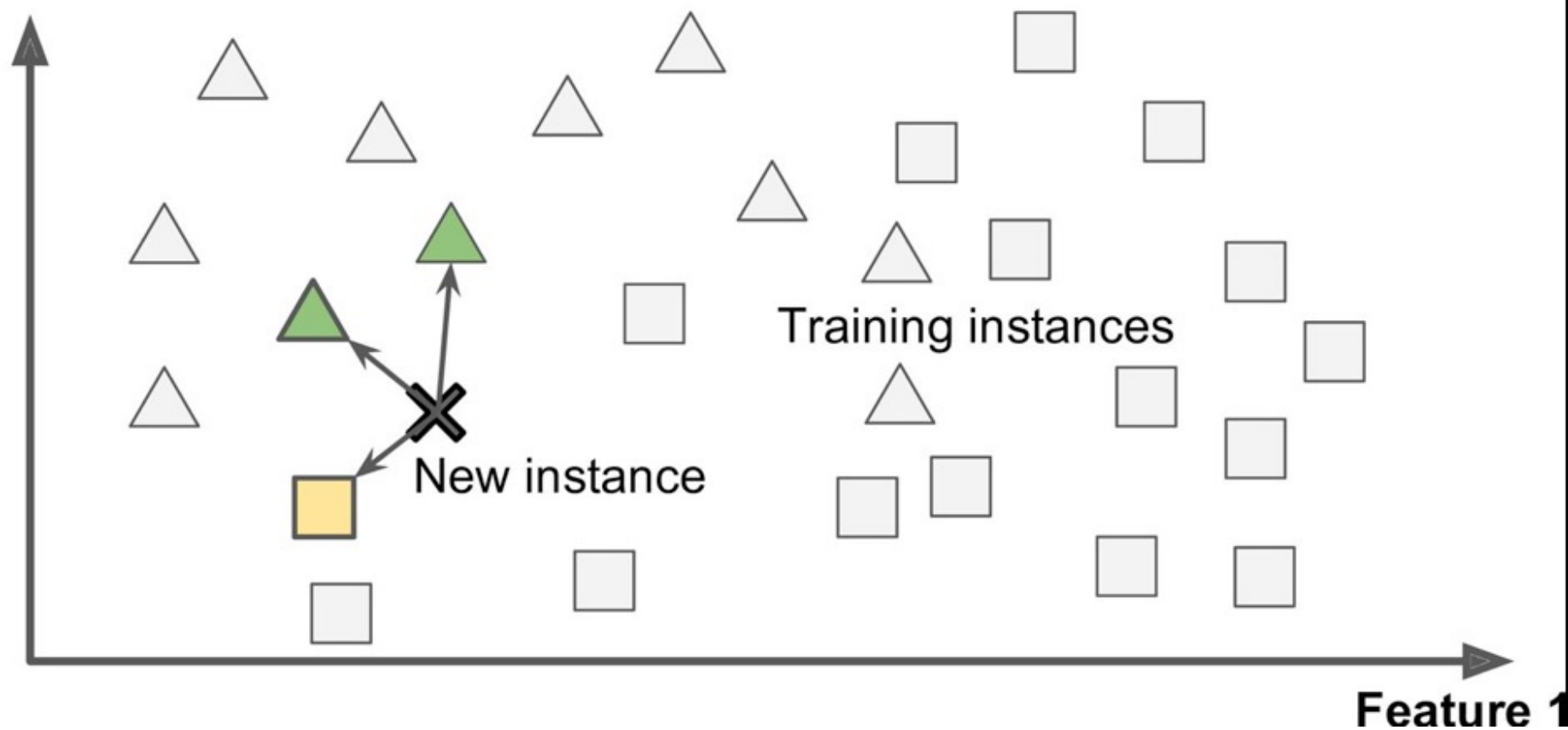
Online Learning (incremental learning)

Online Learning (handle huge datasets)
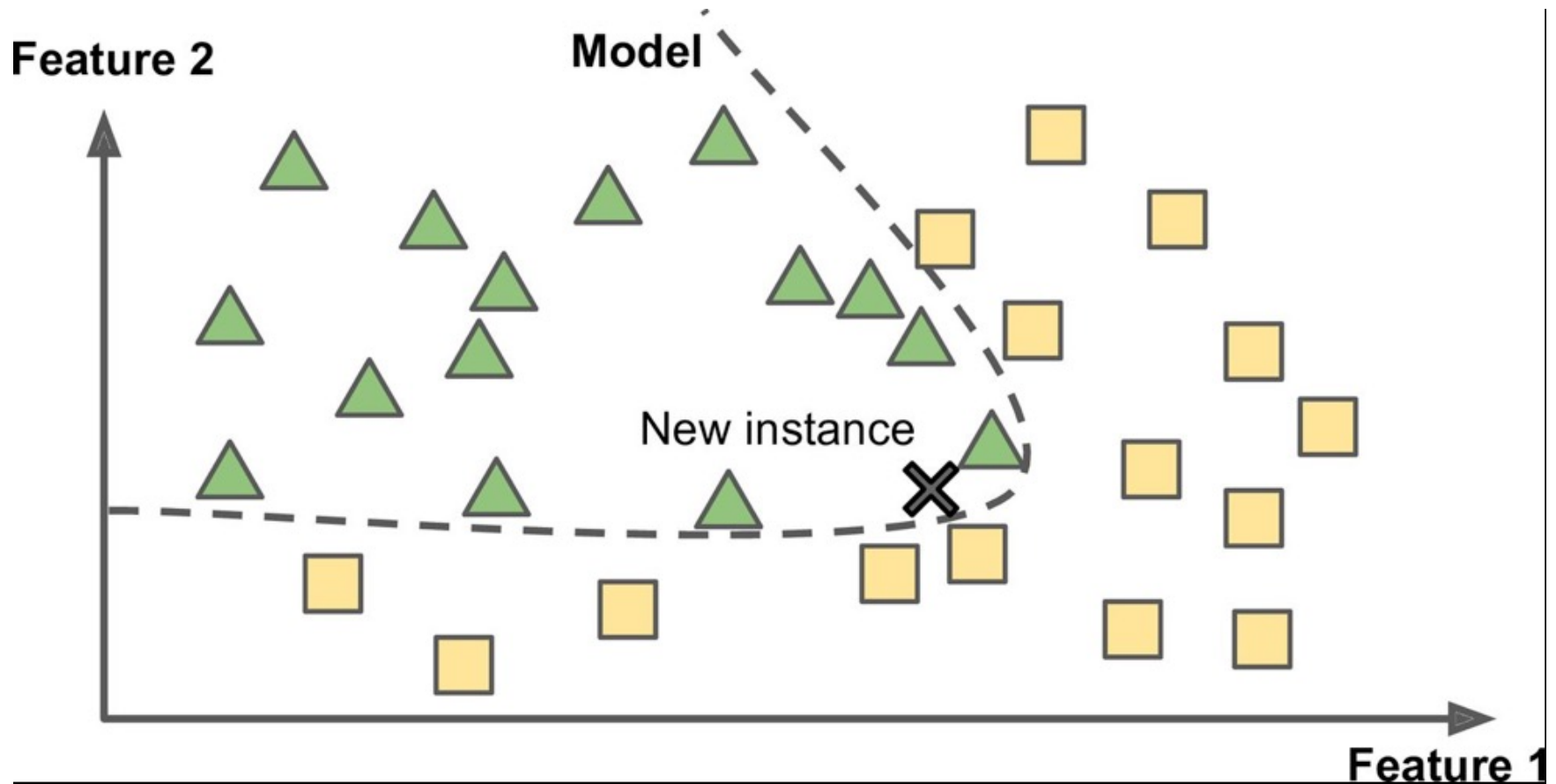
Instance-based Learning
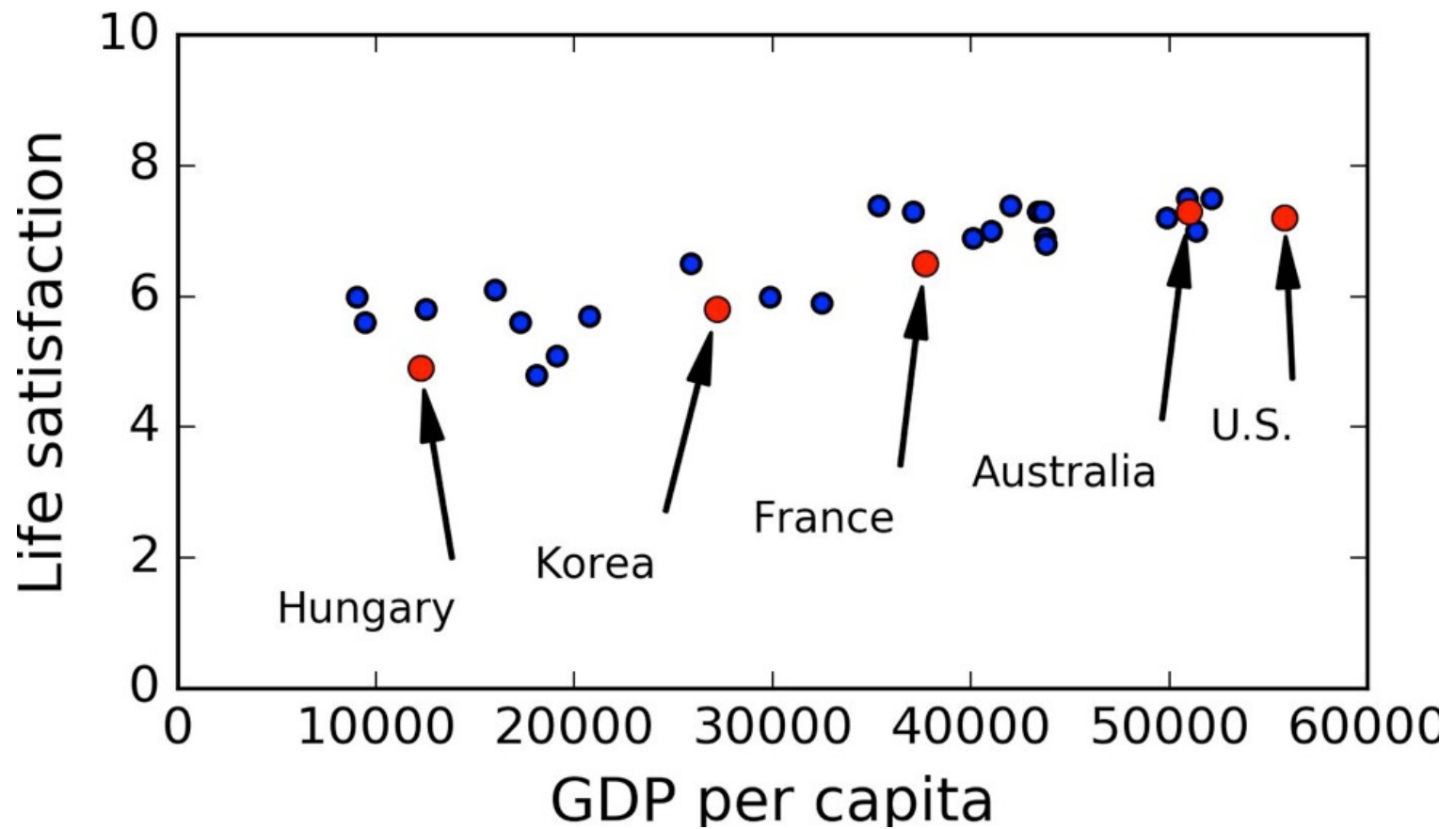
Instance-based learning

Feature 2

Training instances

New instance

Feature 1

Model-based Learning

Model-based learning

Trend?

Possible linear models
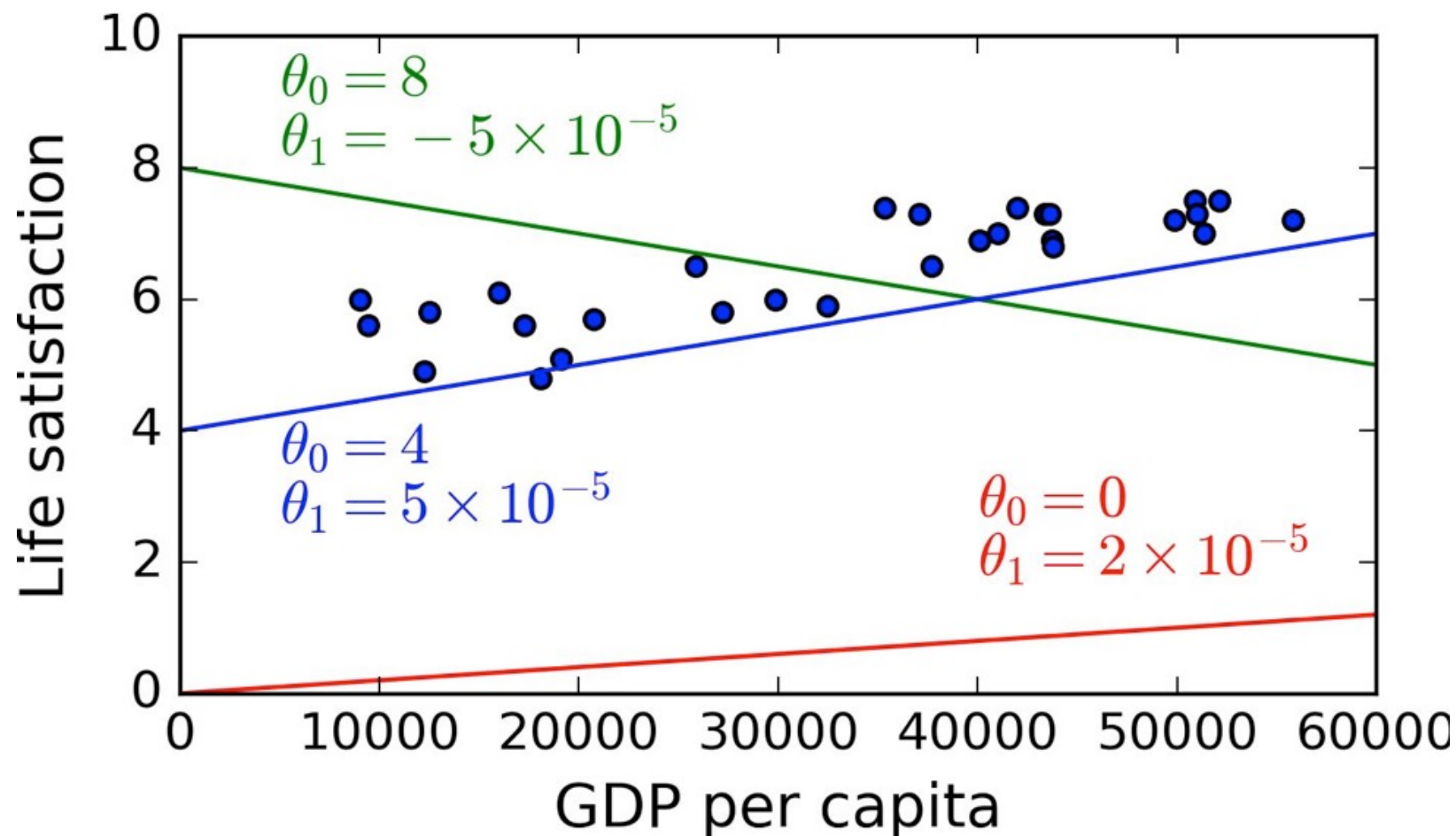
Optimal linear model

$\theta_0 = 4.85$
$\theta_1 = 4.91 \times 10^{-5}$
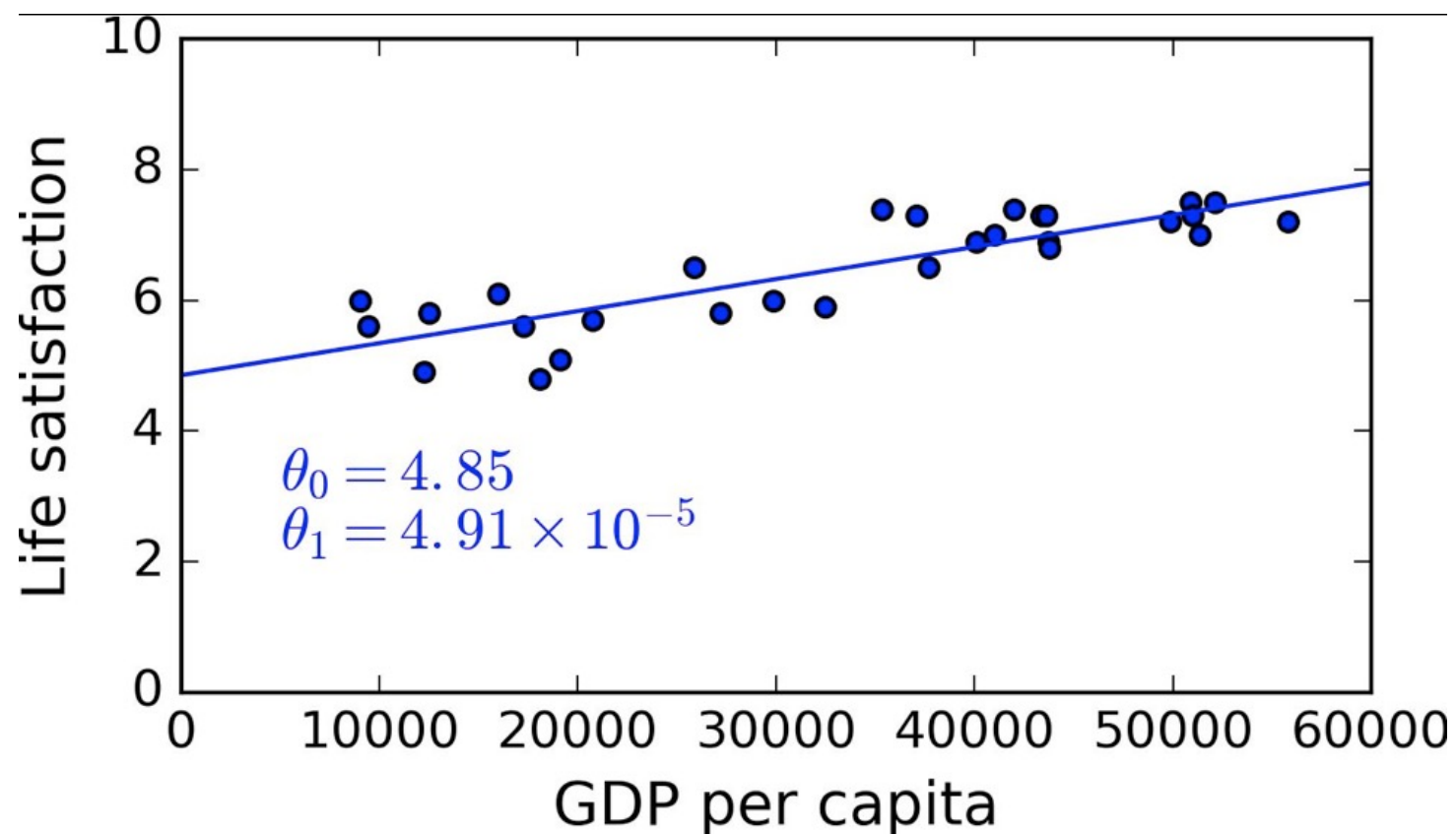
Example

Example 1-1. Training and running a linear model using Scikit-Learn

```python
import matplotlib
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import sklearn.linear_model

# Load the data
oecd_bli = pd.read_csv("oecd_bli_2015.csv", thousands=',')
gdp_per_capita = pd.read_csv("gdp_per_capita.csv",thousands=',',delimiter='\t',
                             encoding='latin1', na_values="n/a")

# Prepare the data
country_stats = prepare_country_stats(oecd_bli, gdp_per_capita)
X = np.c_[country_stats["GDP per capita"]]
y = np.c_[country_stats["Life satisfaction"]]

# Visualize the data
country_stats.plot(kind='scatter', x="GDP per capita", y='Life satisfaction')
plt.show()

# Select a linear model
model = sklearn.linear_model.LinearRegression()

# Train the model
model.fit(X, y)

# Make a prediction for Cyprus
X_new = [[22587]]  # Cyprus' GDP per capita
print(model.predict(X_new)) # outputs [[ 5.96242338]]
```

NOTE

If you had used an instance-based learning algorithm instead, you would have found that Slovenia has the closest GDP per capita to that of Cyprus ($20,732), and since the OECD data tells us that Slovenians' life satisfaction is 5.7, you would have predicted a life satisfaction of 5.7 for Cyprus. If you zoom out a bit and look at the two next closest countries, you will find Portugal and Spain with life satisfactions of 5.1 and 6.5, respectively. Averaging these three values, you get 5.77, which is pretty close to your model-based prediction. This simple algorithm is called *k-Nearest Neighbors* regression (in this example, *k* = 3).

Replacing the Linear Regression model with k-Nearest Neighbors regression in the previous code is as simple as replacing this line:

```python
model = sklearn.linear_model.LinearRegression()
```

with this one:

```python
model = sklearn.neighbors.KNeighborsRegressor(n_neighbors=3)
```