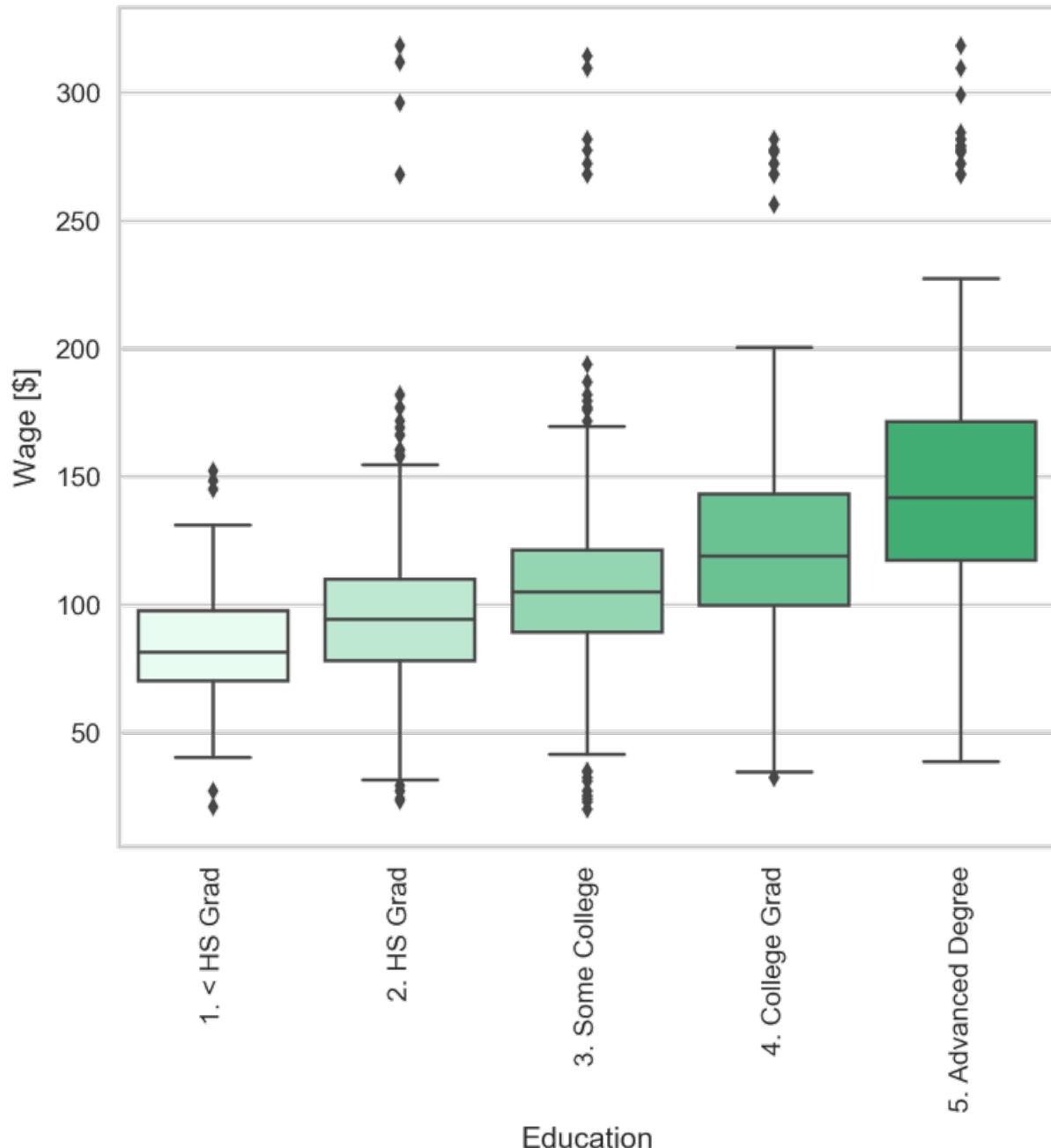


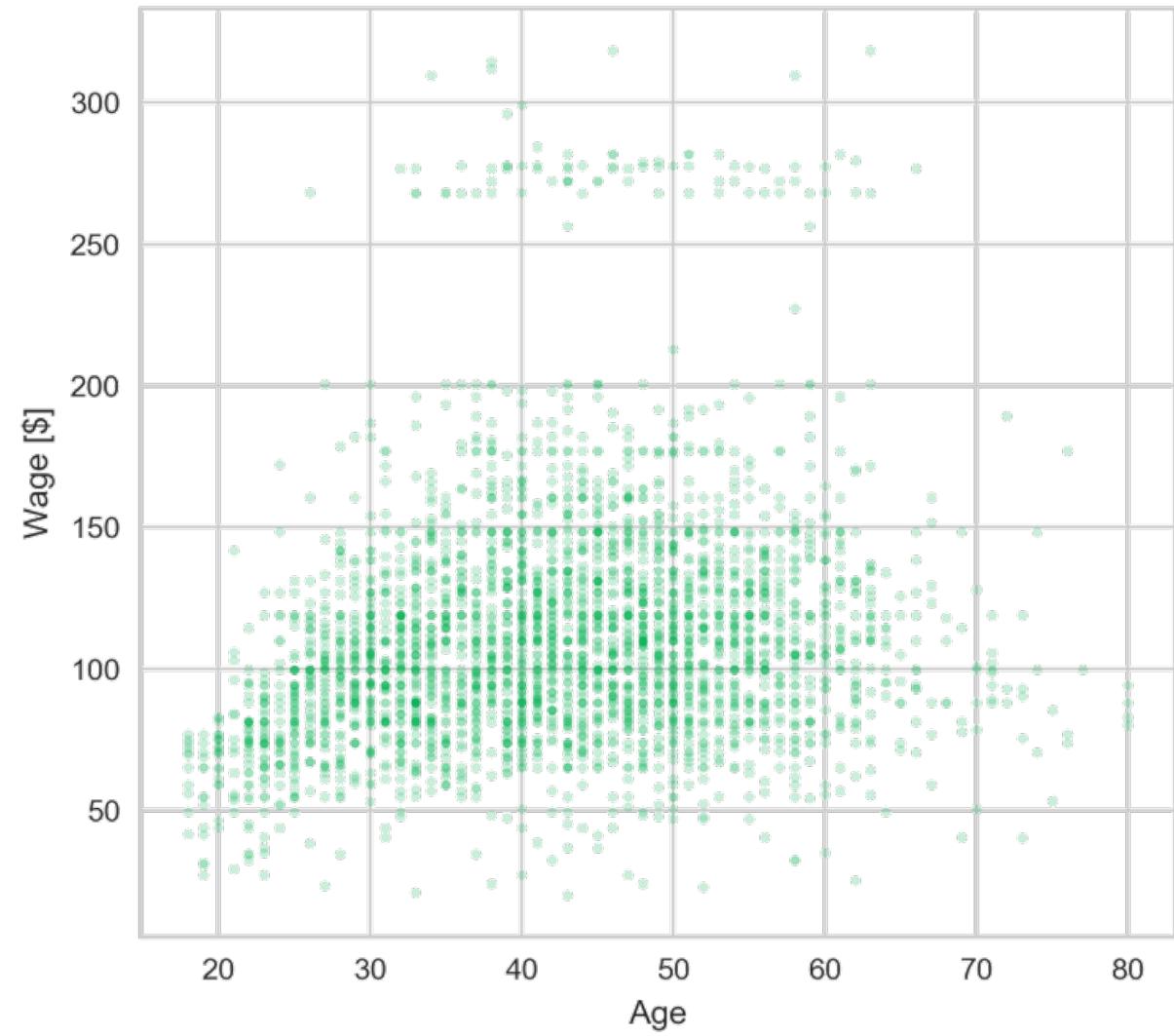
What is machine learning?

Lecture 01

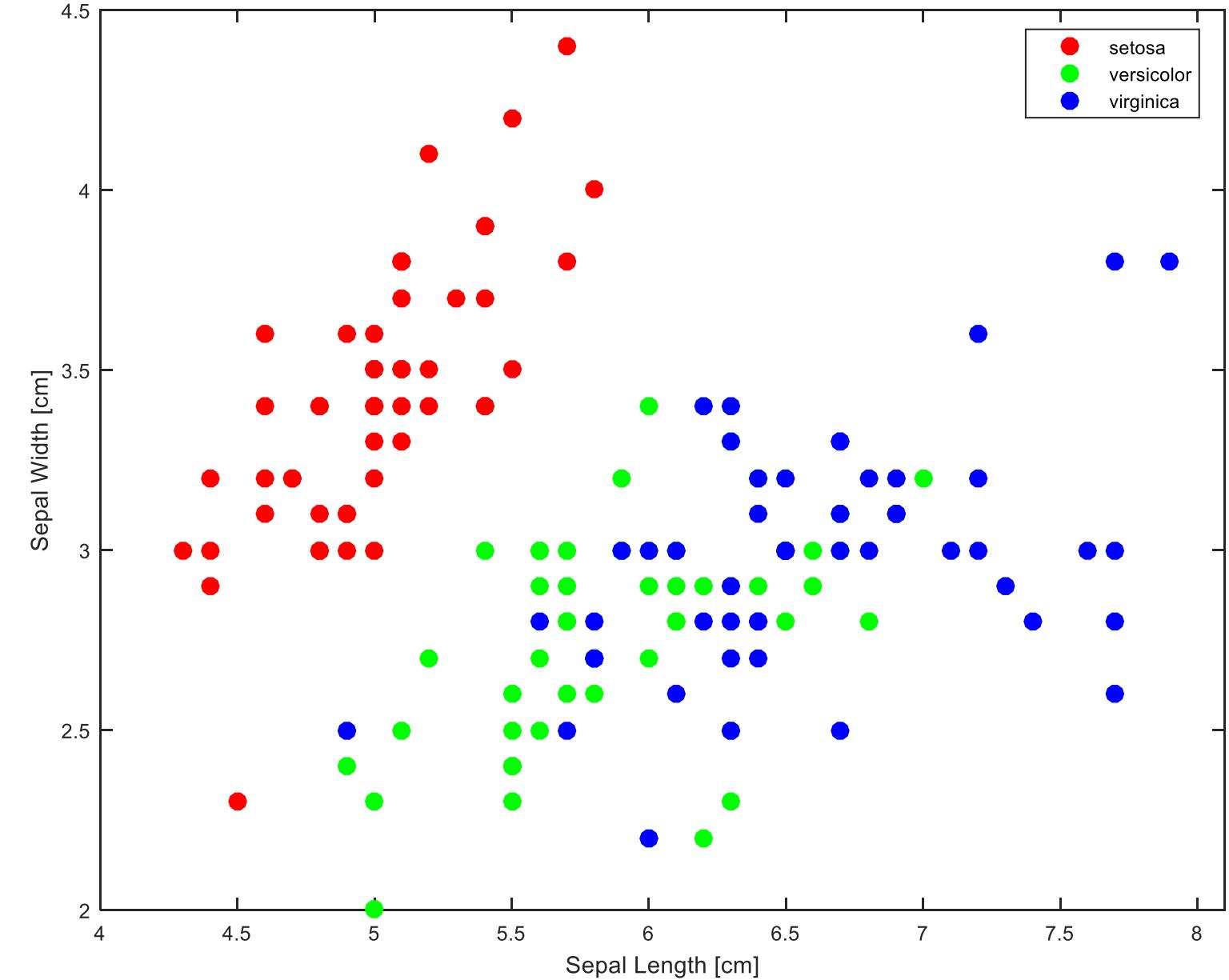
Introductions



Data source: James et al., 2013



Wage data from workers in the mid-Atlantic region



setosa



versicolor



virginica



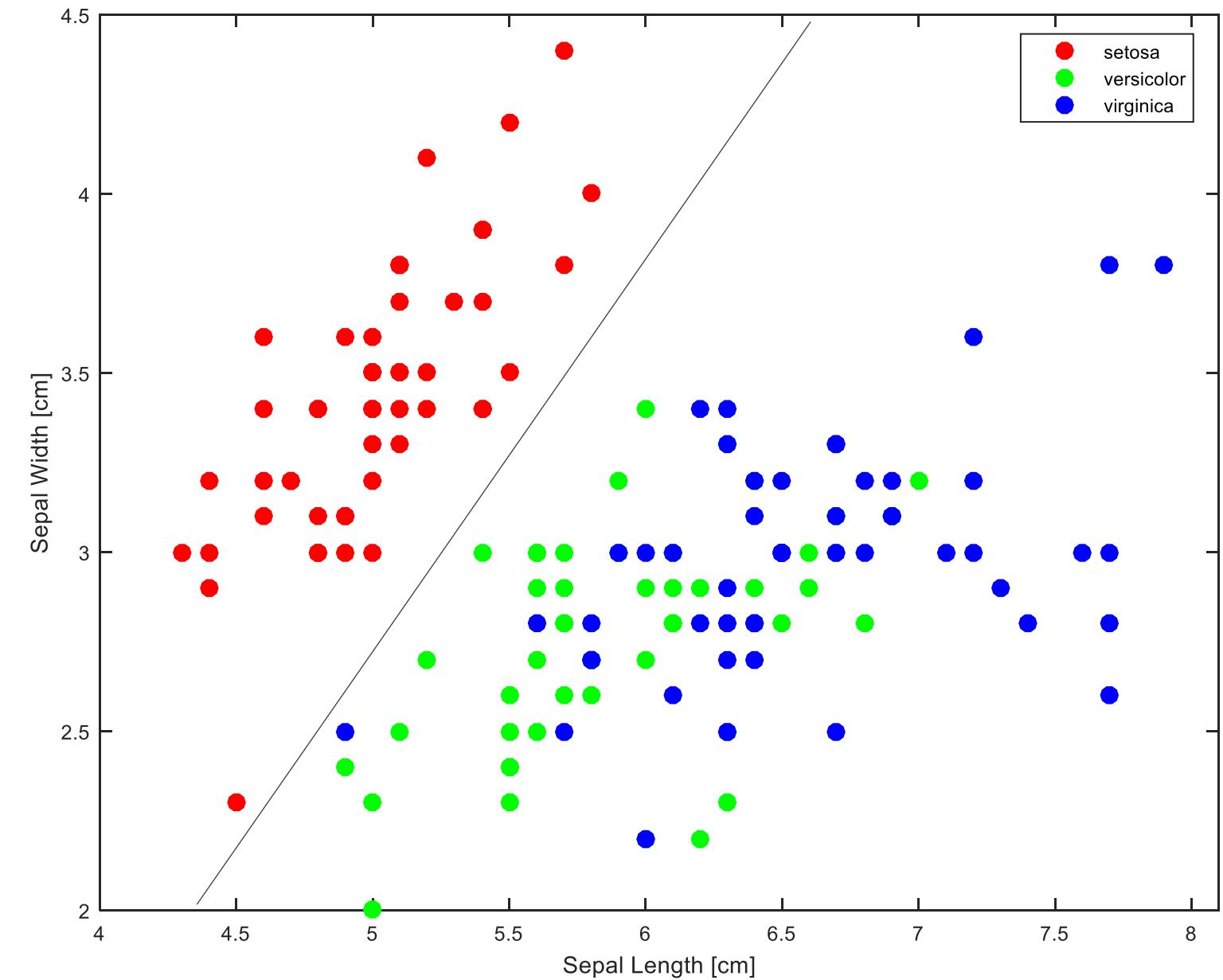
Data Source: Fisher Iris Data

K. Bradbury and L. Collins

What is machine learning?

Lecture 01

4 / 36



setosa



versicolor



virginica



Data Source: Fisher Iris Data

K. Bradbury and L. Collins

What is machine learning?

Lecture 01

5 / 36

Challenges



What
is
this?

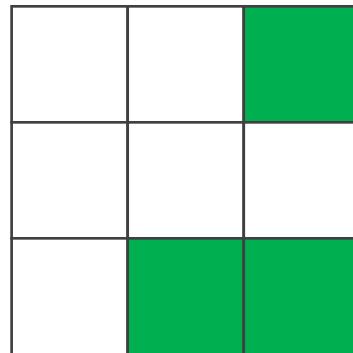
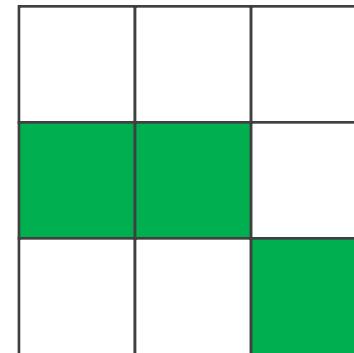
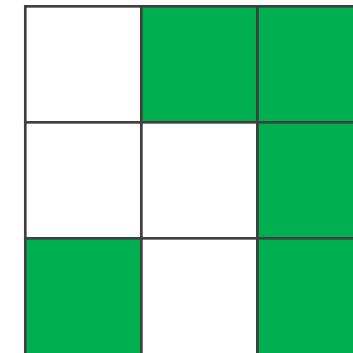
We generalize from past experiences



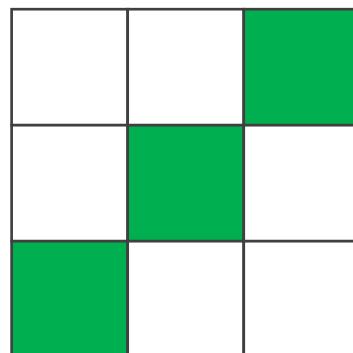
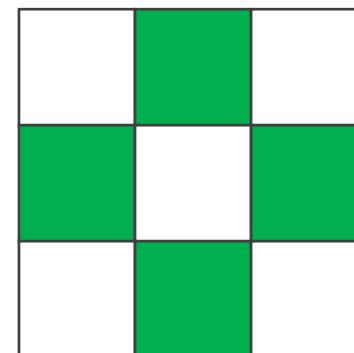
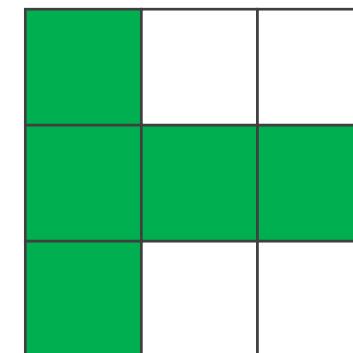
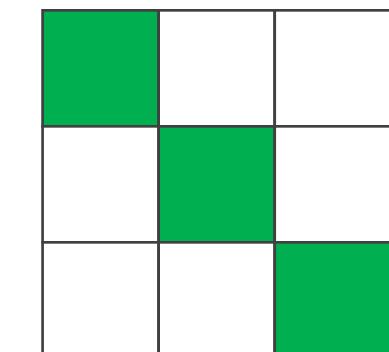
Image: "It's not what it seems" by artist Hikaru Cho

Predict which class x_{new} belongs to...

$$f(x) = 1$$

 x_0  x_2  x_4

$$f(x) = 0$$

 x_1  x_3  x_5  x_{new}

$$f(x_{\text{new}}) = ?$$

Example credit: Yaser Abu-Mostafa, 2012

Machine learning is an **ill-posed problem**

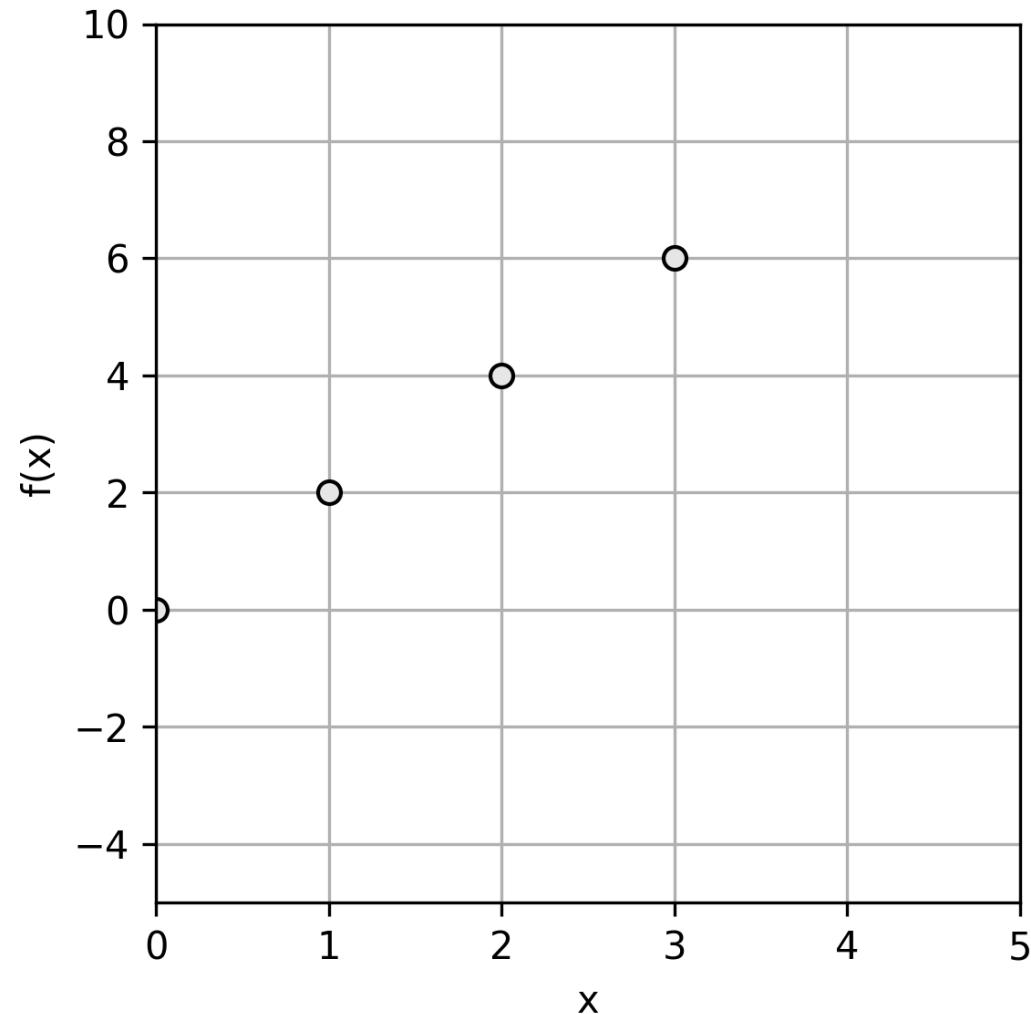
There are often **many** models that will fit your data – how do we choose which to use?

the best models
generalize well

Predict the next value in the sequence...

x	0	1	2	3	4
$f(x)$	0	2	4	6	?

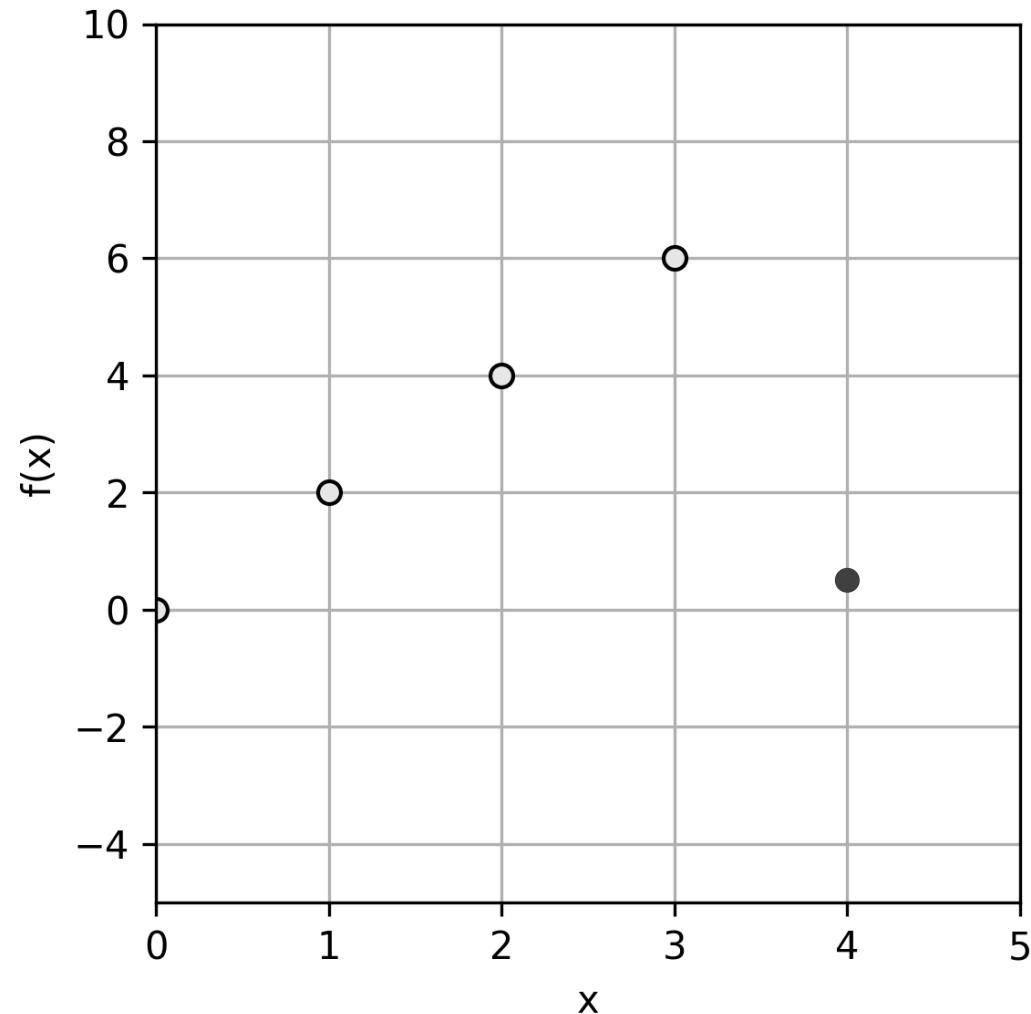
$$f(4) = ?$$



Predict the next value in the sequence...

x	0	1	2	3	4
$f(x)$	0	2	4	6	?

$$f(4) = \boxed{0.530}$$



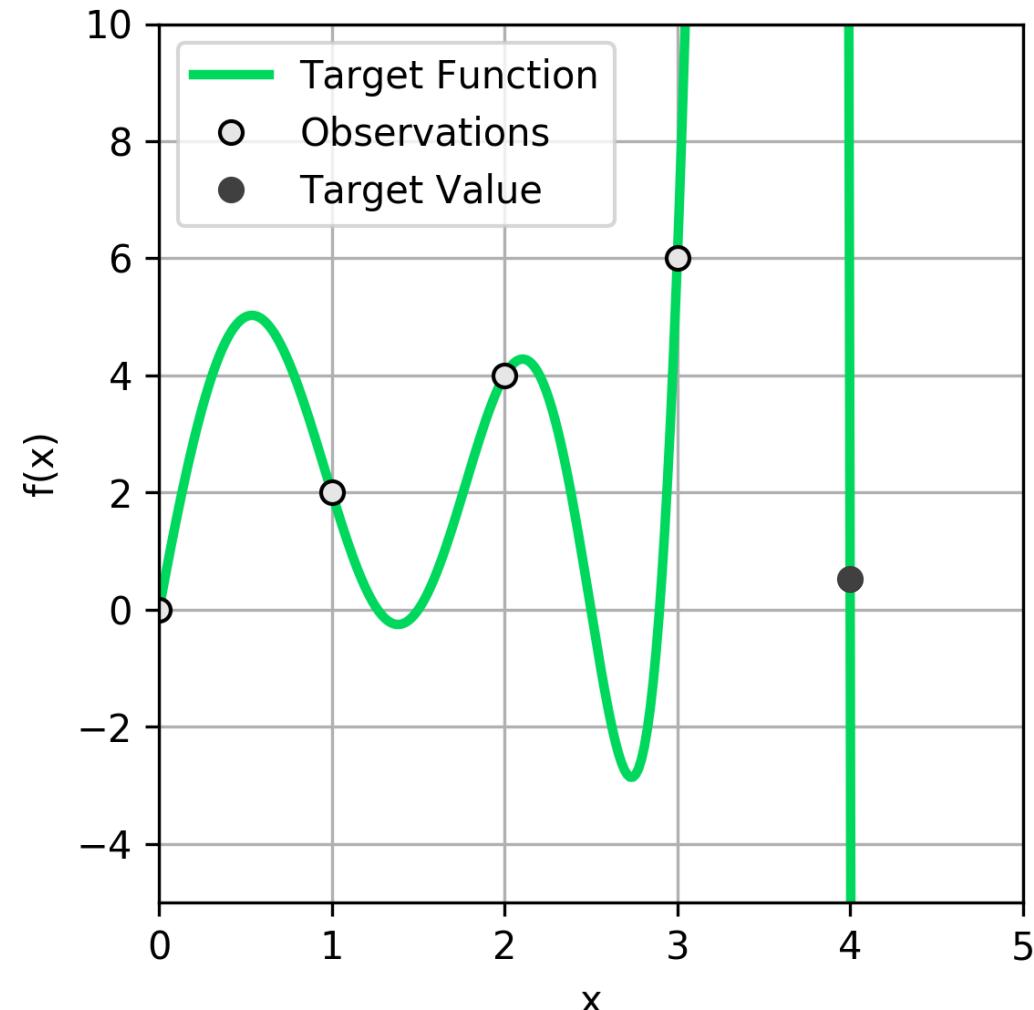
Predict the next value in the sequence...

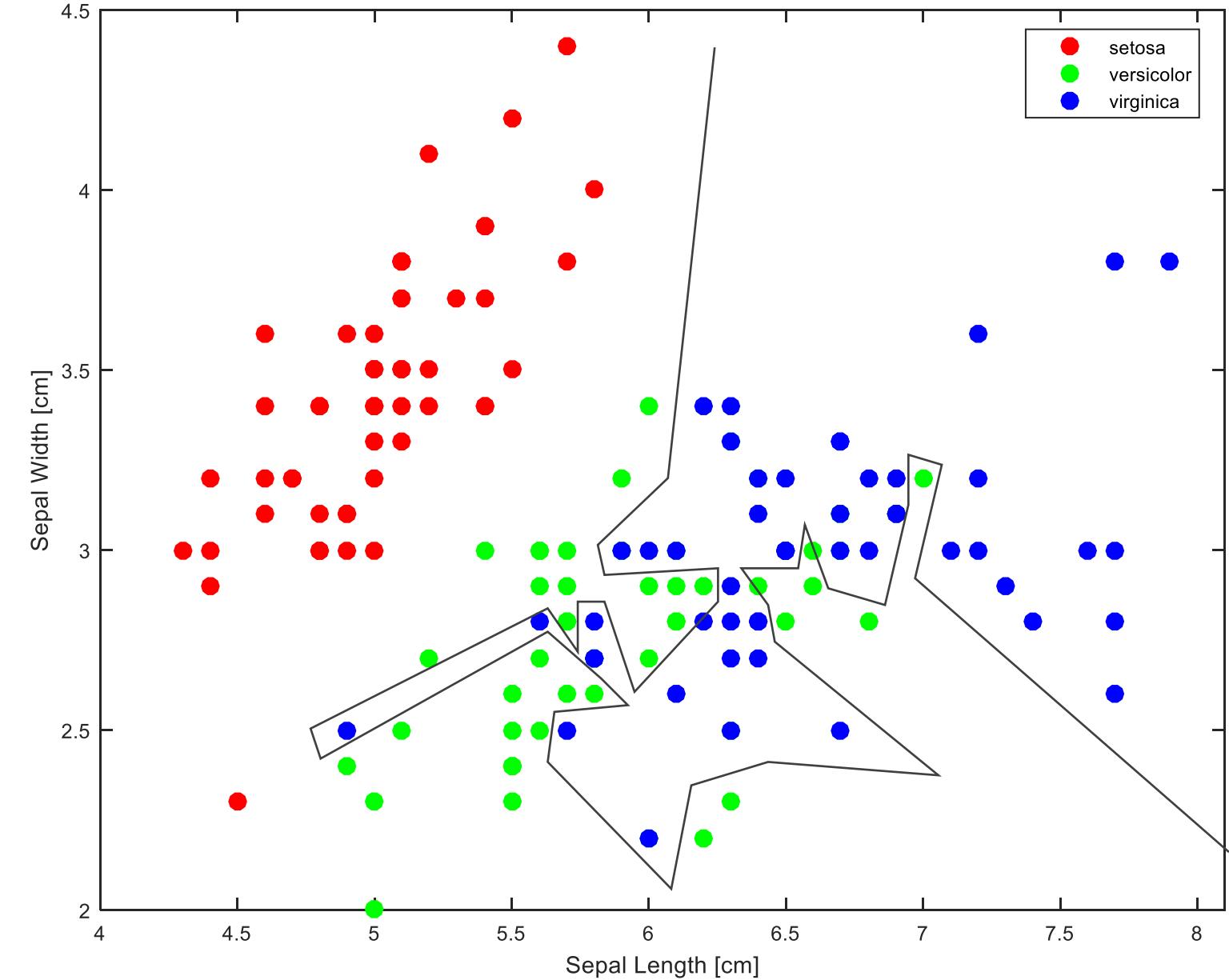
x	0	1	2	3	4
$f(x)$	0	2	4	6	?

$$f(4) = \boxed{0.530}$$

Our guess:

$$f(x) = 16.2x - 6.36x^2 - 11.9x^3 - 4.77x^4 + 7.03x^5 + 8.32x^6 - 9.01x^7 + 2.75x^8 - 0.275x^9$$





setosa



versicolor



virginica



Data Source: Fisher Iris Data

K. Bradbury and L. Collins

What is machine learning?

Lecture 01

15 / 36

Overfit works against generalization

We can use **regularization** to prevent overfit

What is machine learning?

A class of techniques where the **goal** is to **describe**, **predict**, and **strategize**...

...**based on** data, past experiences, and/or direct instruction...

...and do so **automatically**, with minimal human intervention.

Types of machine learning tools

Types of learning

Unsupervised learning

Supervised learning

Reinforcement learning

Common use case

Describe

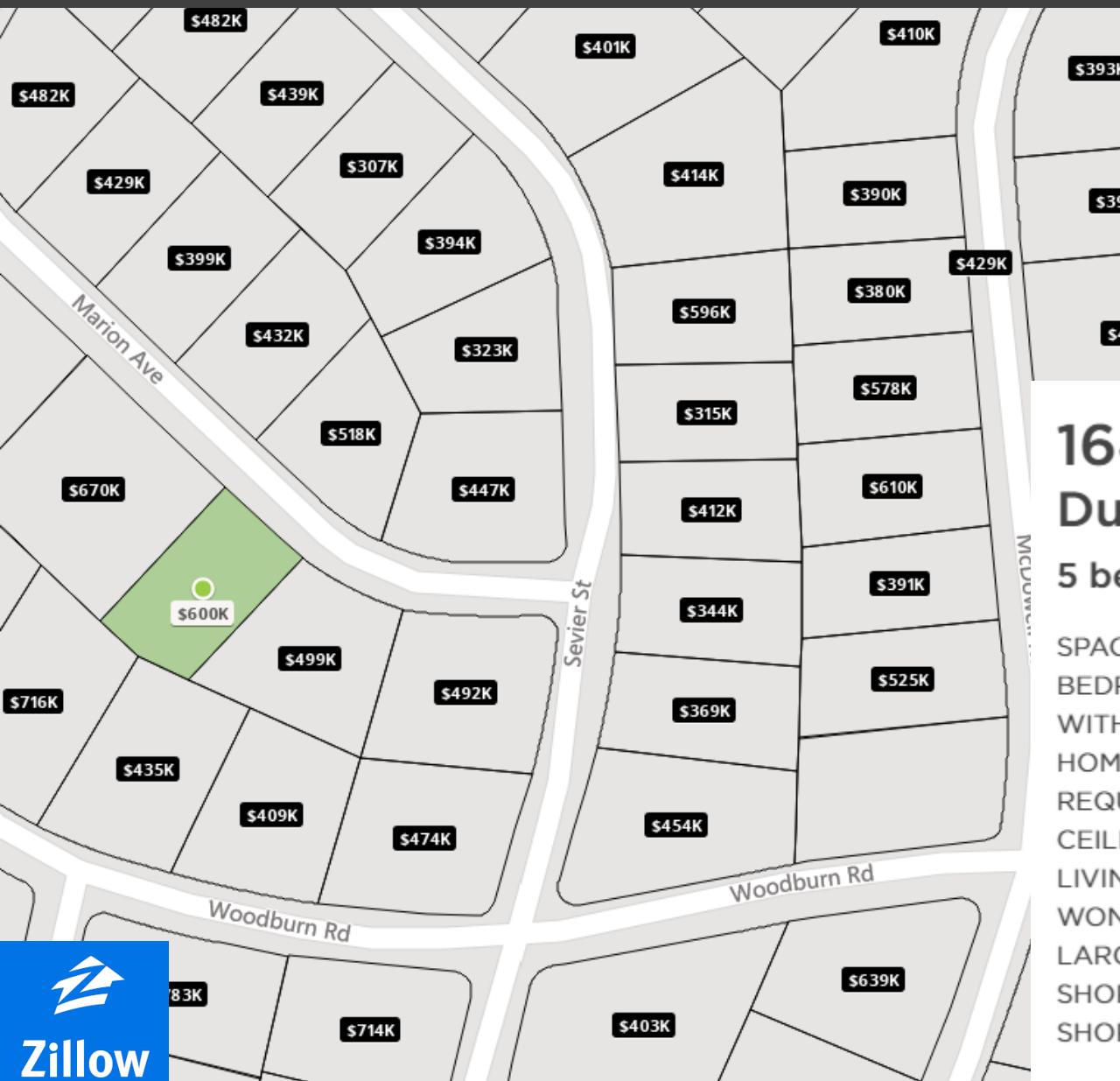
Predict

Strategize

Types of machine learning

	Unsupervised Learning	Supervised Learning	Reinforcement Learning
Goal	Describe ...structure in data	Predict ...from examples	Strategize learn through interaction
Data available	predictors, x	predictor and response pairs, (x, y)	predictors and delayed responses (called rewards)
Types	<ul style="list-style-type: none">• Density estimation• Clustering• Dimensionality reduction• Anomaly detection	<ul style="list-style-type: none">• Classification• Regression	<ul style="list-style-type: none">• Model-free learning• Model-based learning

Sale Price Prediction



27708 Real Estate

1 home for sale

Homes for You

Newest

Cheapest

More



**1640 Marion Ave,
Durham, NC 27705**

5 beds · 4 baths · 3,264 sqft

SPACIOUS RANCH W FINISHED LL WALKOUT! 5 BEDROOMS AND 4 BRAND NEW BATHS! RENOVATED WITH CUSTOM FEATURES THRUOUT! CONTEMPORARY HOME WITH MANY HANDICAP ACCESSIBLE REQUIREMENTS ALREADY IN PLACE! VAULTED CEILINGS! SECLUDED TREED LOT! GREAT HOME FOR LIVING AND ENTERTAINING WITH LARGE REAR DECK! WONDERFUL CONTEMPORARY FEEL THAT LIVES LARGE WITH EASY ACCESS TO DUKE UNIVERSITY; SHOPPING; HEALTH CARE; PARKS; RESTAURANTS AND SHOPPING; AND EASY HIGHWAY ACCESS!

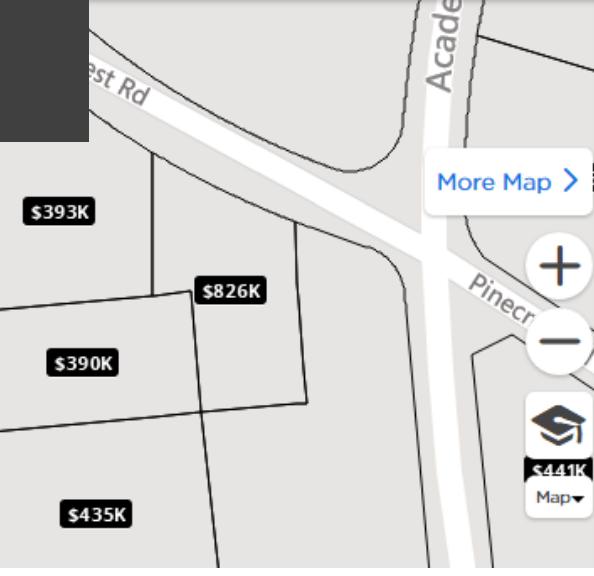
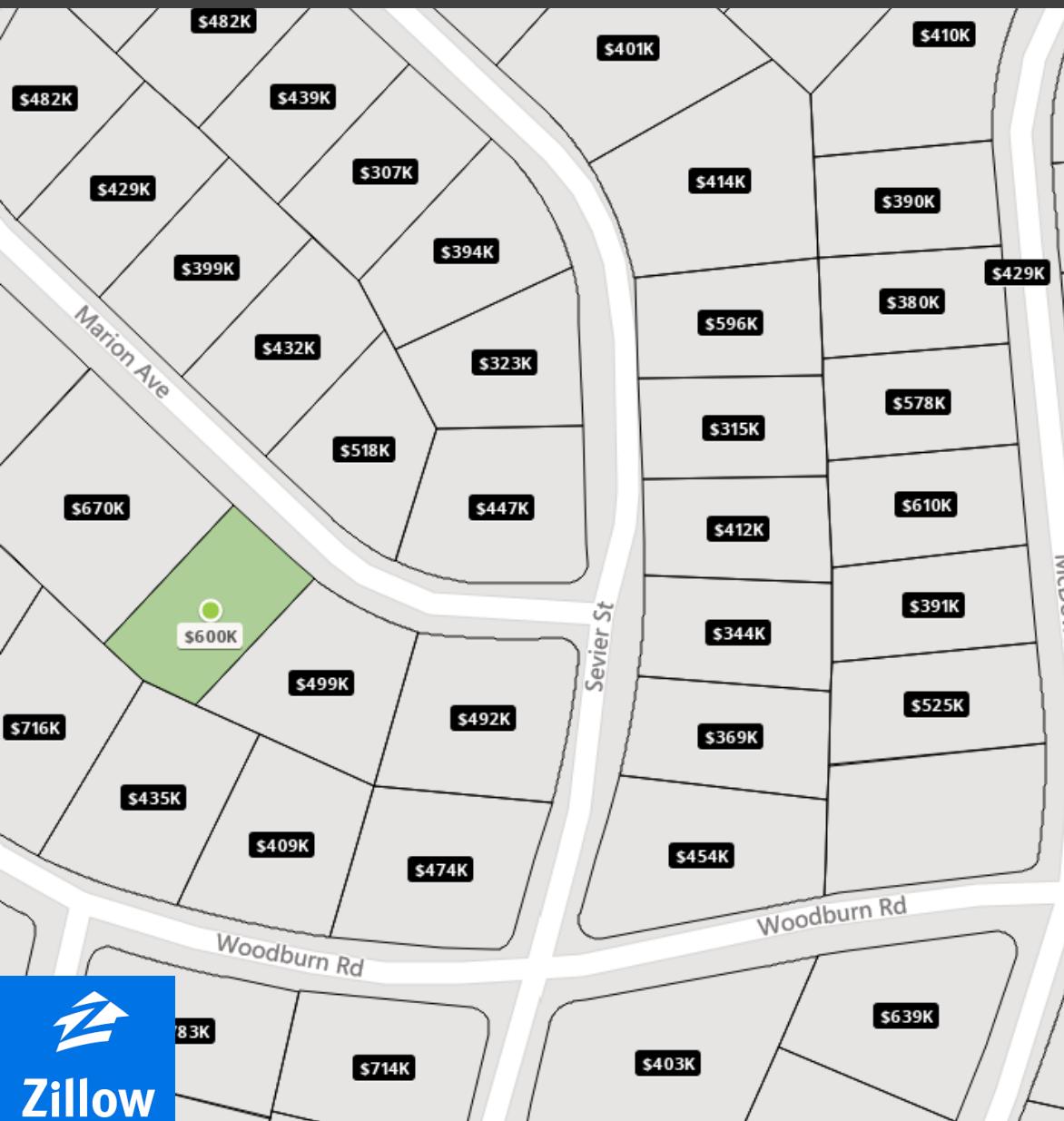
● FOR SALE
\$599,900

Price cut: -\$79,100 (6/17)
Zestimate*: \$619,585

EST. MORTGAGE

\$2,284/mo

Sale Price Prediction



27708 Real Estate

1 home for sale

Homes for You

Newest

Cheapest

More



**1640 Marion Ave,
Durham, NC 27705**

5 beds · 4 baths · 3,264 sqft

SPACIOUS RANCH W FINISHED LL WALKOUT! 5 BEDROOMS AND 4 BRAND NEW BATHS! RENOVATED WITH CUSTOM FEATURES THRUOUT! CONTEMPORARY HOME WITH MANY HANDICAP ACCESSIBLE REQUIREMENTS ALREADY IN PLACE! VAULTED CEILINGS! SECLUDED TREED LOT! GREAT HOME FOR LIVING AND ENTERTAINING WITH LARGE REAR DECK! WONDERFUL CONTEMPORARY FEEL THAT LIVES LARGE WITH EASY ACCESS TO DUKE UNIVERSITY: SHOPPING; HEALTH CARE; PARKS; R SHOPPING; AND EASY HIGHWAY AC

● FOR SALE
\$599,900

Price cut: -\$79,100 (6/17)

Zestimate®: **\$619,585**

EST. MORTGAGE

\$2,284/mo

[Get pre-qualified](#)

Zestimate®: \$619,585

Sale Price Prediction

Input Data:

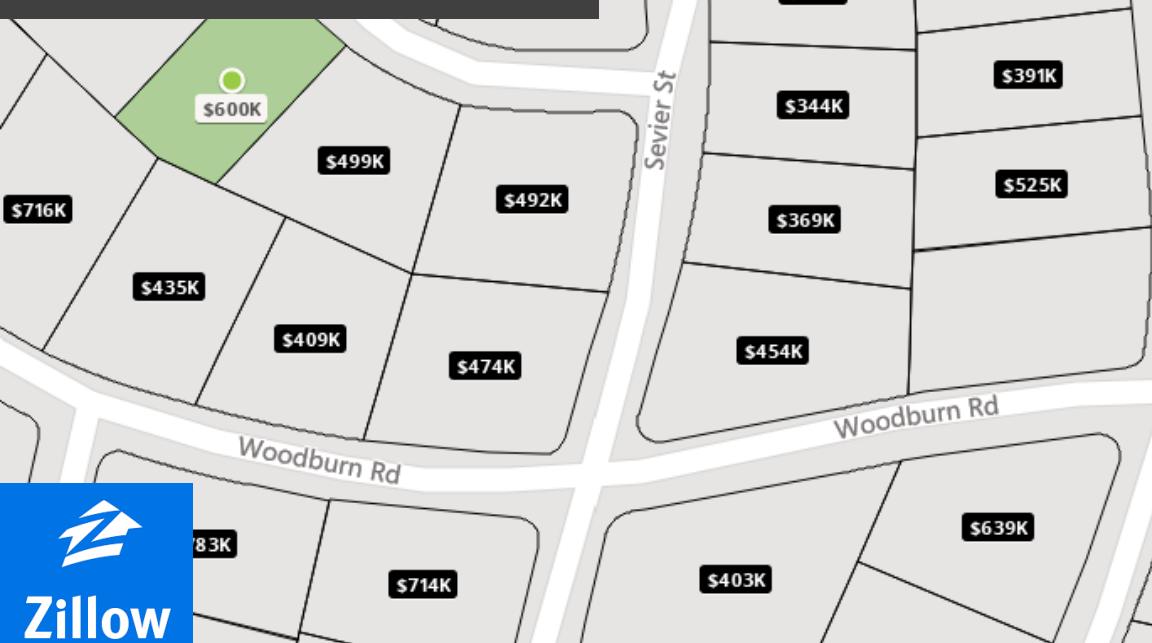
Home characteristics
(Numerical & Categorical)

Target Data:

Price estimate (numerical)

Learning Category:

Supervised Learning
Regression



27708 Real Estate 1 home for sale

Homes for You Newest Cheapest More

\$79,100 (Jun 17)

HOUSE FOR SALE
\$599,900 5 bds . 4 ba . 3,264 sqft
1640 Marion Ave, Durham, NC

● FOR SALE
\$599,900
Price cut: -\$79,100 (6/17)
Zestimate®: \$619,585

EST. MORTGAGE
\$2,284/mo Get pre-qualified

1640 Marion Ave, Durham, NC 27705

5 beds · 4 baths · 3,264 sqft

SPACIOUS RANCH W FINISHED LL WALKOUT! 5 BEDROOMS AND 4 BRAND NEW BATHS! RENOVATED WITH CUSTOM FEATURES THRUOUT! CONTEMPORARY HOME WITH MANY HANDICAP ACCESSIBLE REQUIREMENTS ALREADY IN PLACE! VAULTED CEILINGS! SECLUDED TREED LOT! GREAT HOME FOR LIVING AND ENTERTAINING WITH LARGE REAR DECK! WONDERFUL CONTEMPORARY FEEL THAT LIVES LARGE WITH EASY ACCESS TO DUKE UNIVERSITY: SHOPPING; HEALTH CARE; PARKS; R

Zestimate®: \$619,585



Video Recommendations



Sherlock

97% Match 2017 TV-14 4 Series

Next Up

S3:E2 "The Sign of Three"

While mortal danger stalks Watson's wedding reception, Sherlock faces his biggest challenge of all: delivering a best man's speech!



Season 3's episode "The Abominable Bride," which originally aired as a TV movie, won two Emmys.

MY LIST



NETFLIX

OVERVIEW

EPISODES

MORE LIKE THIS

DETAILS

Video Recommendations



Sherlock

97% Match 2017 TV-14 4 Series

97% Match

Wedding reception, Sherlock
faces his biggest challenge yet: delivering a best man's speech!



Season 3's episode "The Abominable Bride," which originally aired as a TV movie, won two Emmys.

MY LIST



NETFLIX

OVERVIEW

EPISODES

MORE LIKE THIS

DETAILS

Video Recommendations



Sherlock

97% Match 2017 TV-14 4 Series

97% Match

Wedding reception, Sherlock
isn't the biggest challenge of all... delivering a best man's speech



Season 3's episode "The Abominable Bride," which originally aired as a TV movie, won two Emmys.

+ MY LIST



NETFLIX

OVERVIEW

EPISODES

MORE LIKE THIS

DETAILS

What is machine learning?

Lecture 01

Input Data:

User video ratings
(numerical and categorical)



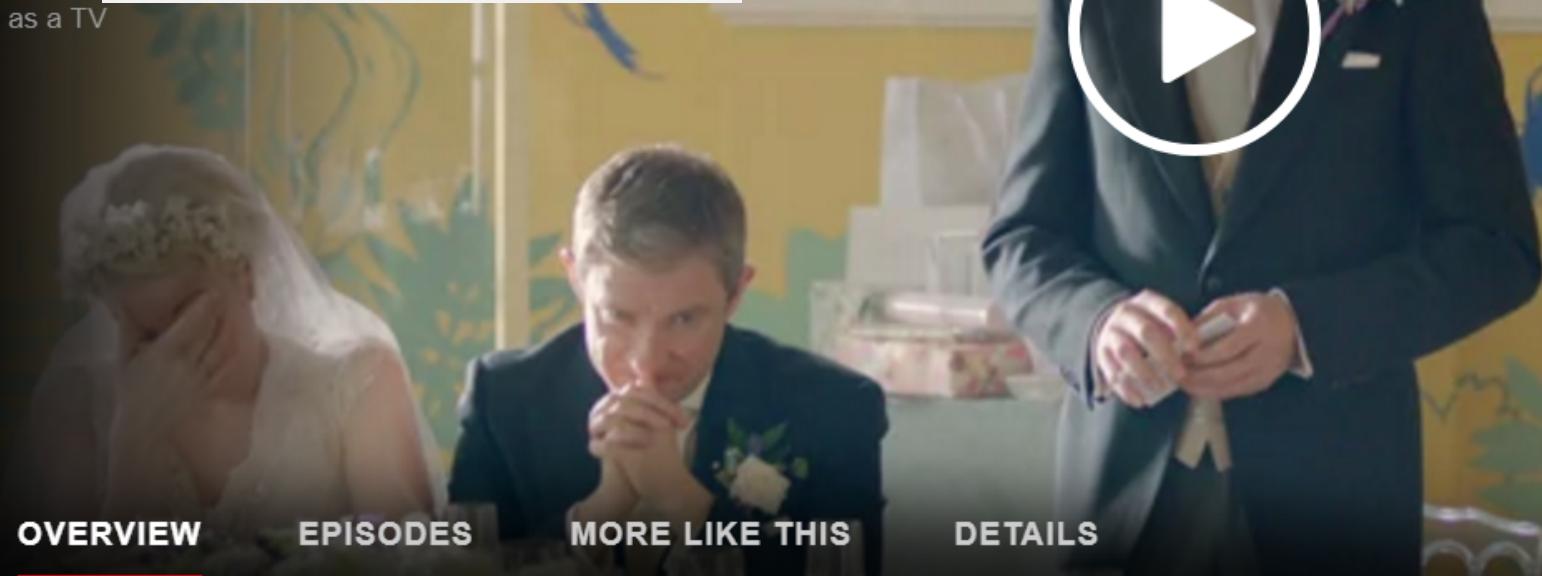
Target Data:

User rating of video
(numerical)



Learning Category:

Supervised Learning
Regression/Classification



Spam Filters

From: Internal Revenue Service
[mailto:yourtaxrefund@InternalRevenueService.com]
Sent: Tuesday, July 22, 2008 9:47 AM
Subject: Get your tax refund now
Importance: High

After the last annual calculations of your account activity we have determined that you are eligible to receive a tax refund of \$479.30 .

Please submit the tax refund request and allow us 2-6 days in order to process it.

A refund can be delayed for a variety of reasons. For example submitting invalid records or applying after the deadline.

To access the form for your tax refund, please click here (<http://e-dlogs.rta.mi.th:84/www.irs.gov/>)

Note: Deliberate wrong inputs will be prosecuted by law.

Regards,

Internal Revenue Service

Spam example source: itservices.uchicago.edu

Spam Filters

From: Internal Revenue Service
[mailto:yourtaxrefund@InternalRevenueService.com]

Sent: Tuesday, July 22, 2008 9:47 AM

Subject: Get your tax refund now

Importance: High

After the last annual calculations of your account activity we have determined that you are eligible to receive a tax refund of \$479.30 .

Please submit the tax refund request and allow us 2-6 days in order to process it.

A refund can be delayed for a variety of reasons. For example submitting invalid records or applying after the deadline.

To access the form for your tax refund, please click here (<http://e-dlogs.rta.mi.th:84/www.irs.gov/>)

Note: Deliberate wrong inputs will be prosecuted by law.

Regards,

Internal Revenue Service

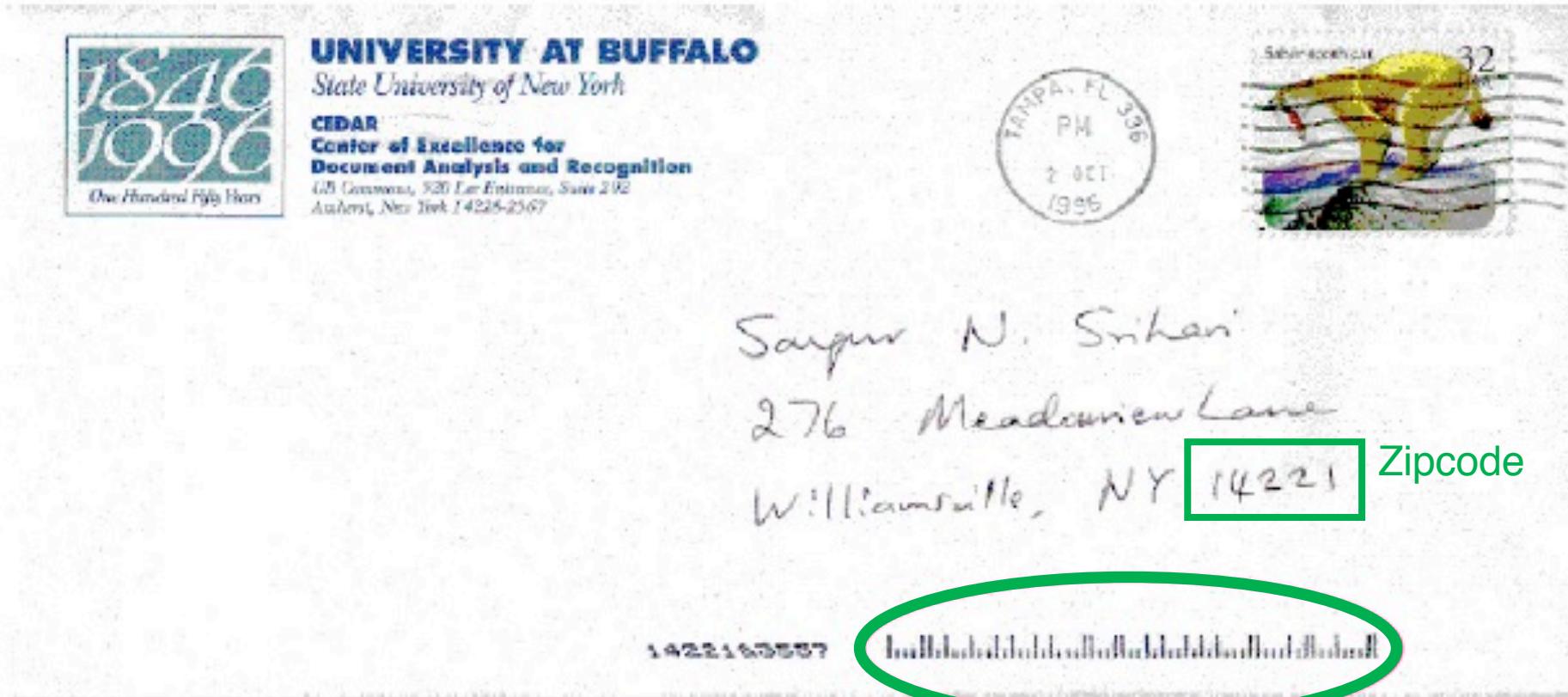
Input Data:
Email text (text)

Target Data :
Spam/not spam
(category)

Learning Category:
Supervised Learning
Classification (binary)

Spam example source: itservices.uchicago.edu

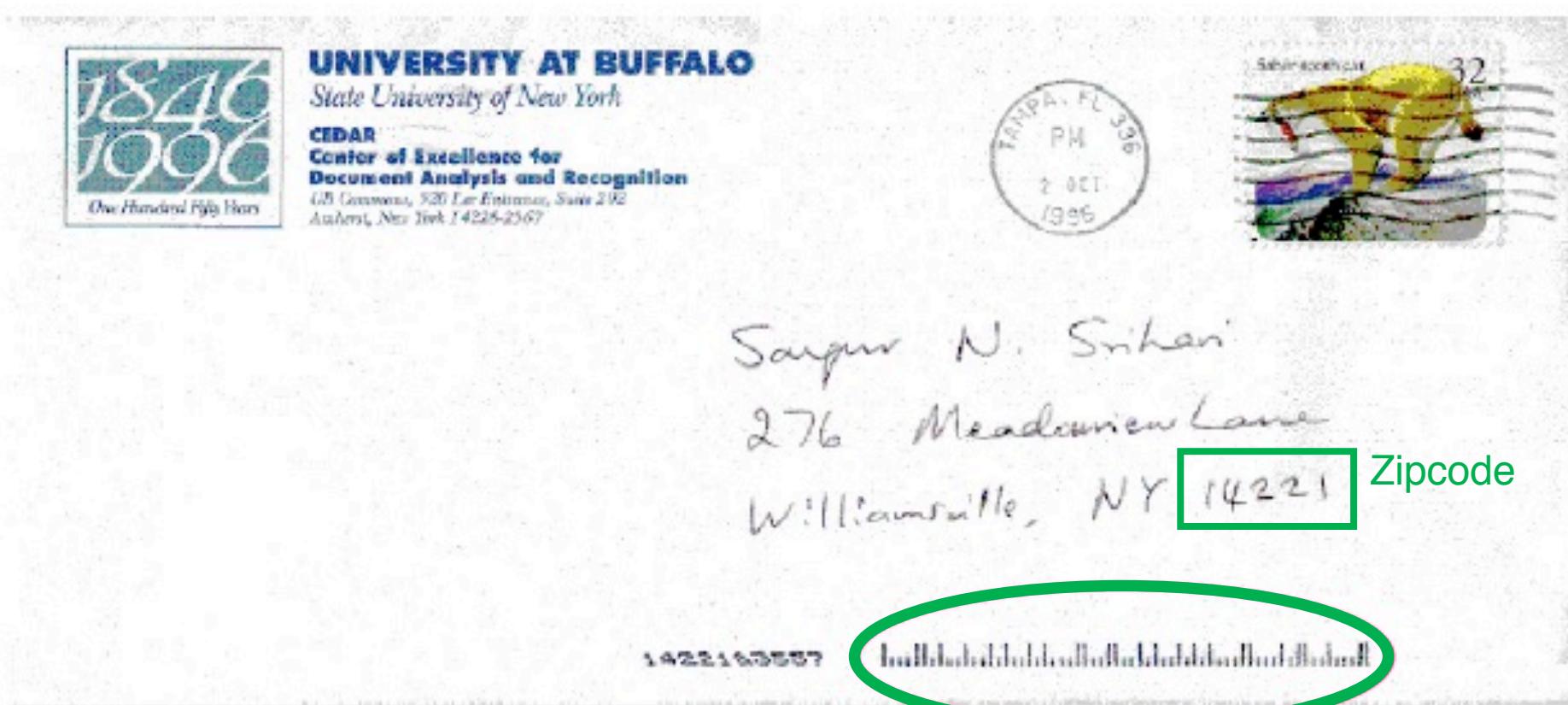
Handwriting and Optical Character Recognition



Among the first handwritten addresses sorted automatically in October 1996

Image source: Sargur Srihari, SUNY

Handwriting and Optical Character Recognition



Input Data:
Imagery

Target Data:
Text Characters

Learning Category:
Supervised Learning
Classification (multiclass)

Among the first handwritten addresses sorted automatically in October 1996

Image source: Sargur Srihari, SUNY

Where's Waldo?

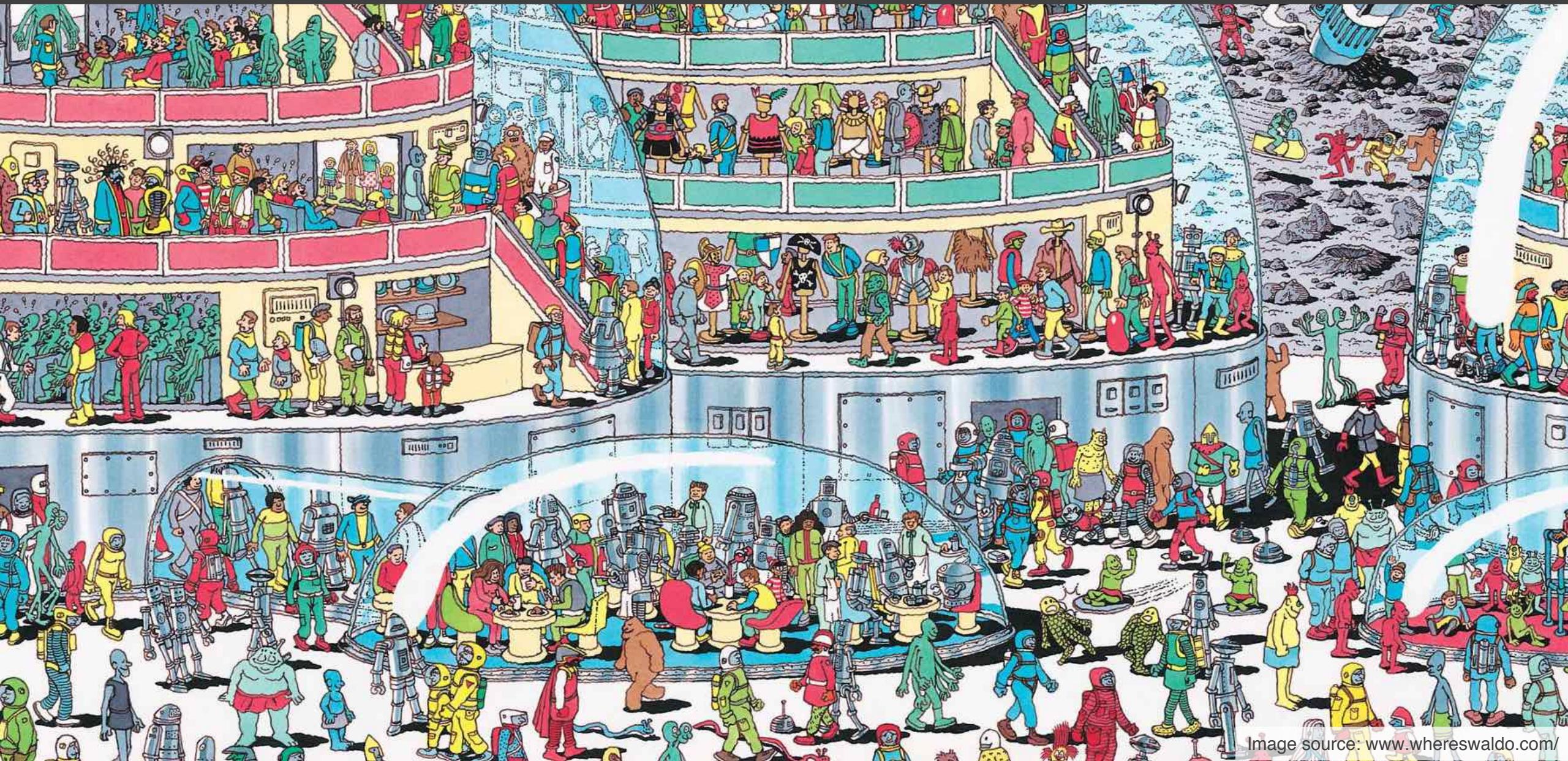


Image source: www.whereswaldo.com/

Where's Waldo = Computer Vision Problem



Image source: www.whereswaldo.com/

Where's Waldo = Computer Vision Problem



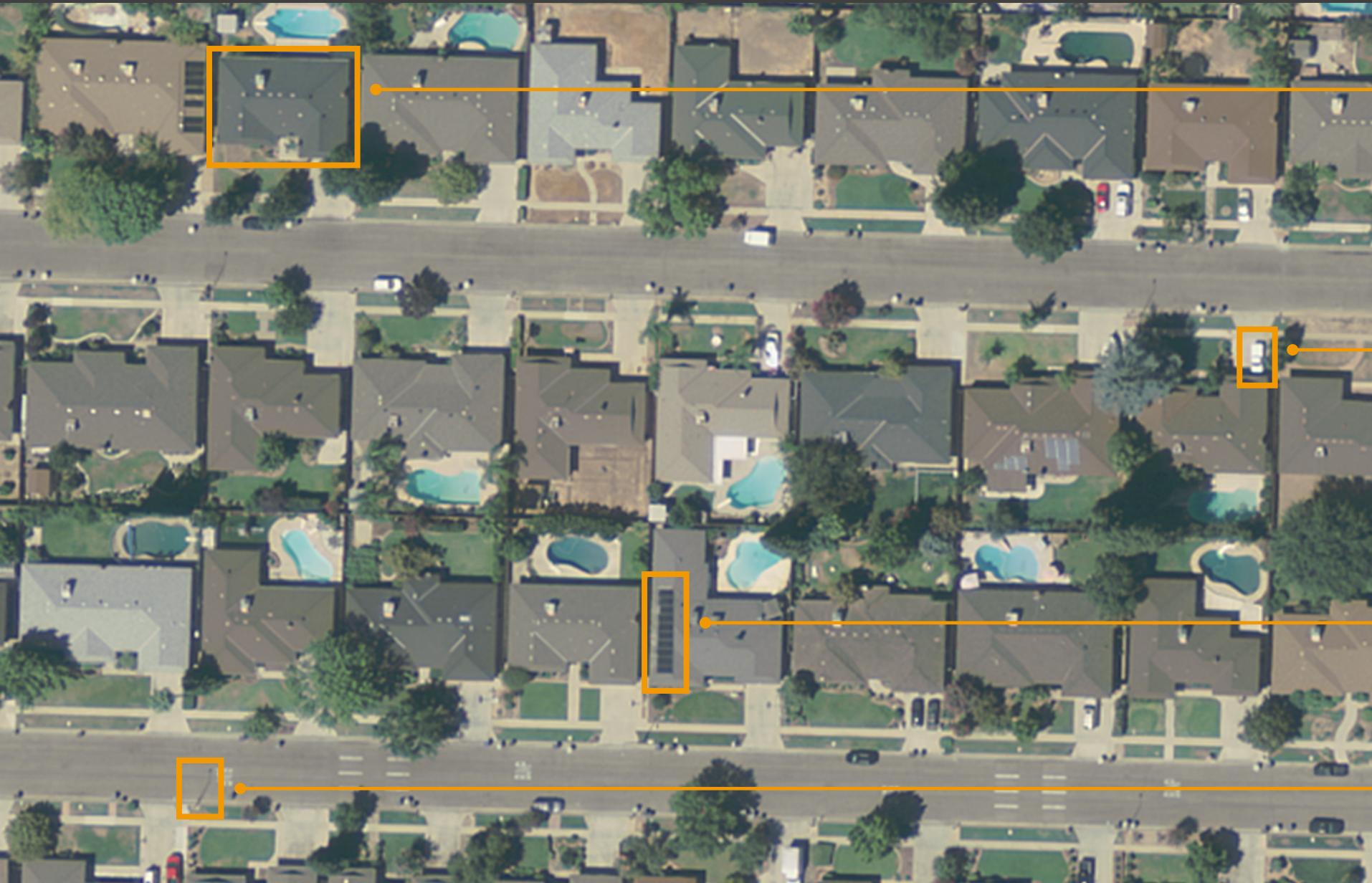
Input Data:
Color Imagery (Image)

Target Data:
Locations in an image
(label for each pixel)

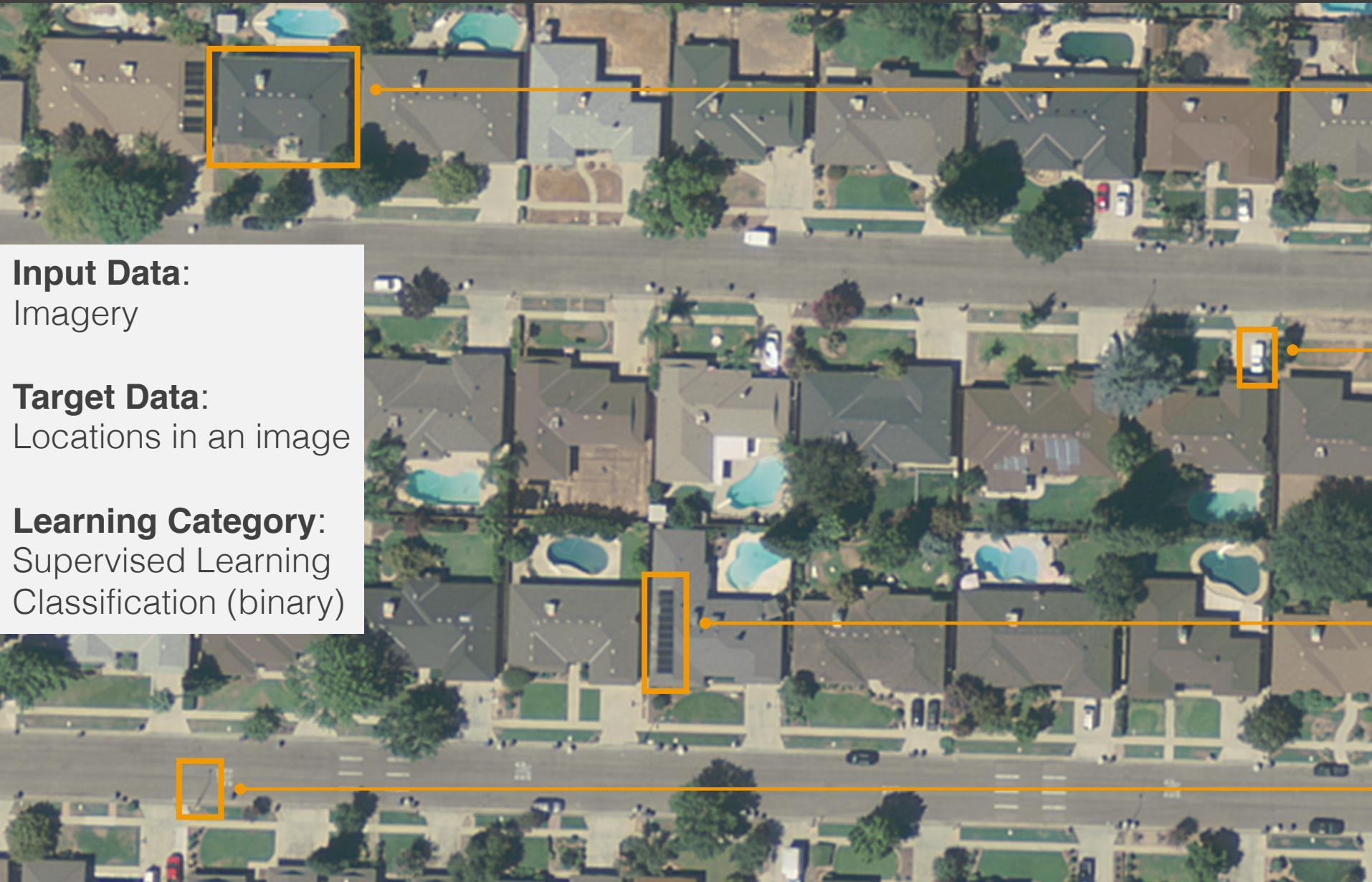
Learning Category:
Supervised Learning
Classification (binary)

Image source: www.whereswaldo.com/

Object Recognition: Energy Systems



Object Recognition: Energy Systems



Input Data:

Imagery

Target Data:

Locations in an image

Learning Category:

Supervised Learning
Classification (binary)

Building
behind-the-meter
energy consumption

Car
transportation
energy consumption

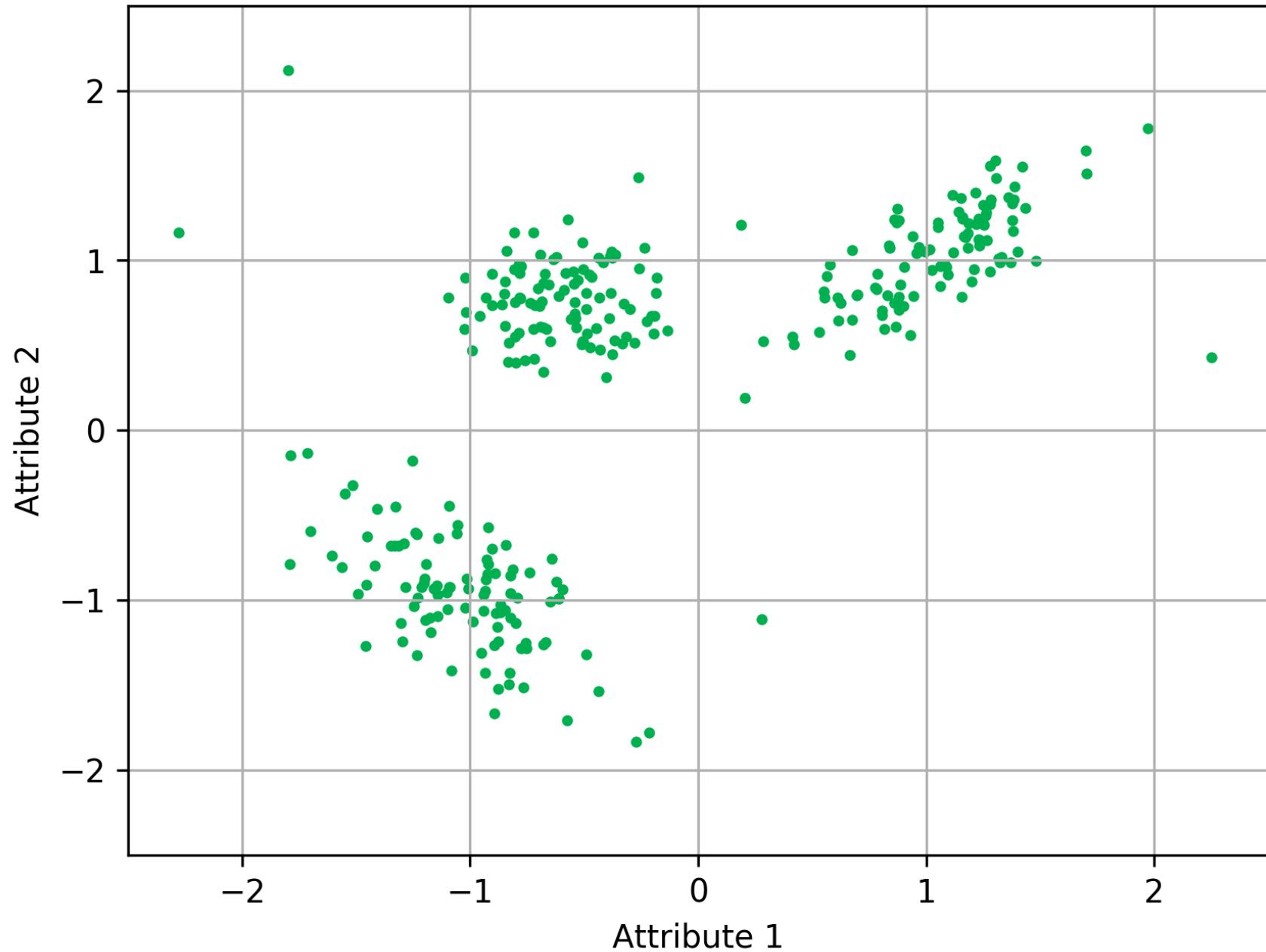
Solar Array
distributed energy
resources

Light Pole
access to electricity

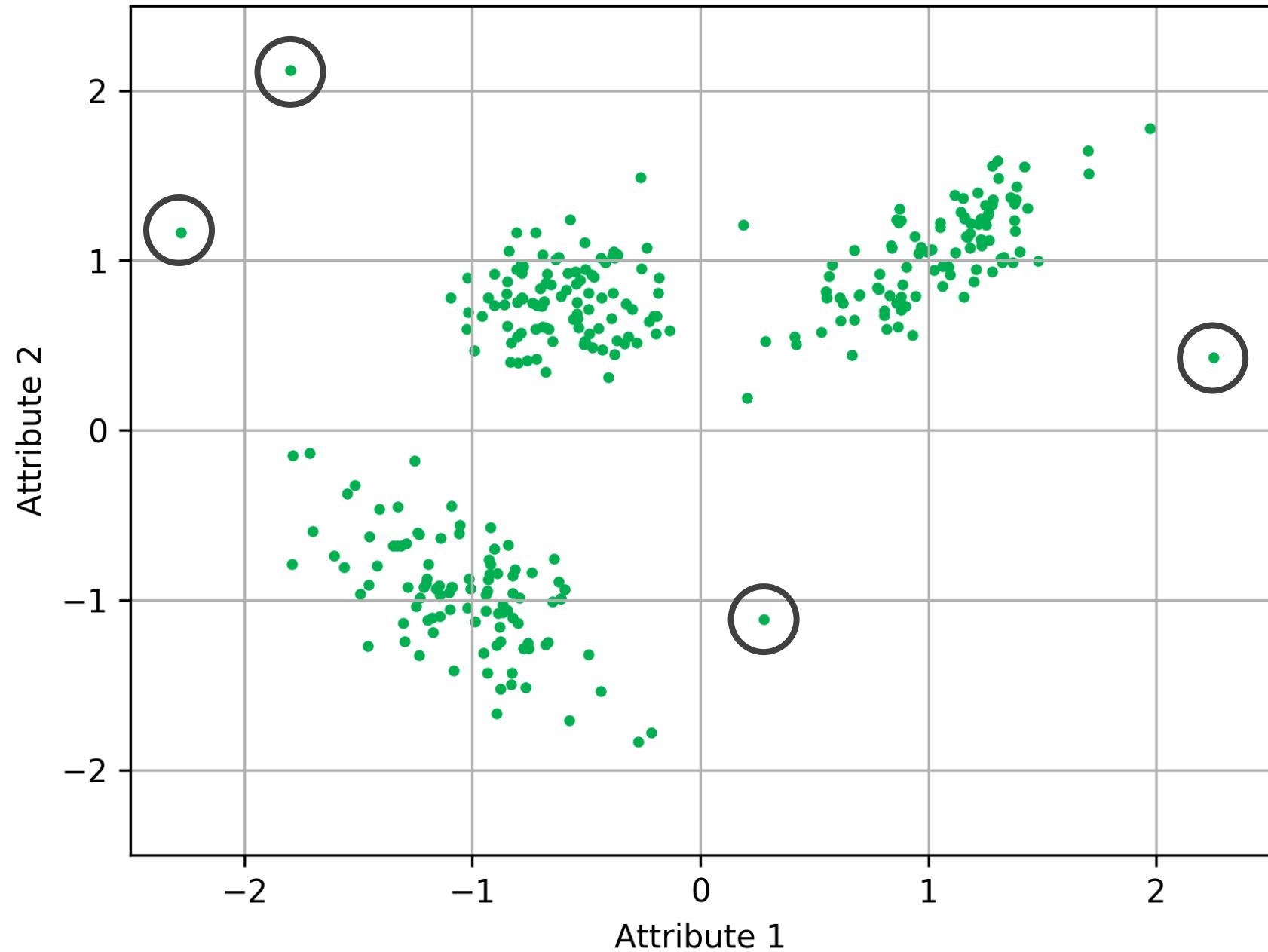
Types of machine learning

	Unsupervised Learning	Supervised Learning	Reinforcement Learning
Goal	Describe	Predict	Strategize
Data available	predictors, x	predictor and response pairs, (x, y)	predictors and delayed responses (called rewards)
Types	<ul style="list-style-type: none">• Density estimation• Clustering• Dimensionality reduction• Anomaly detection	<ul style="list-style-type: none">• Classification• Regression	<ul style="list-style-type: none">• Model-free learning• Model-based learning

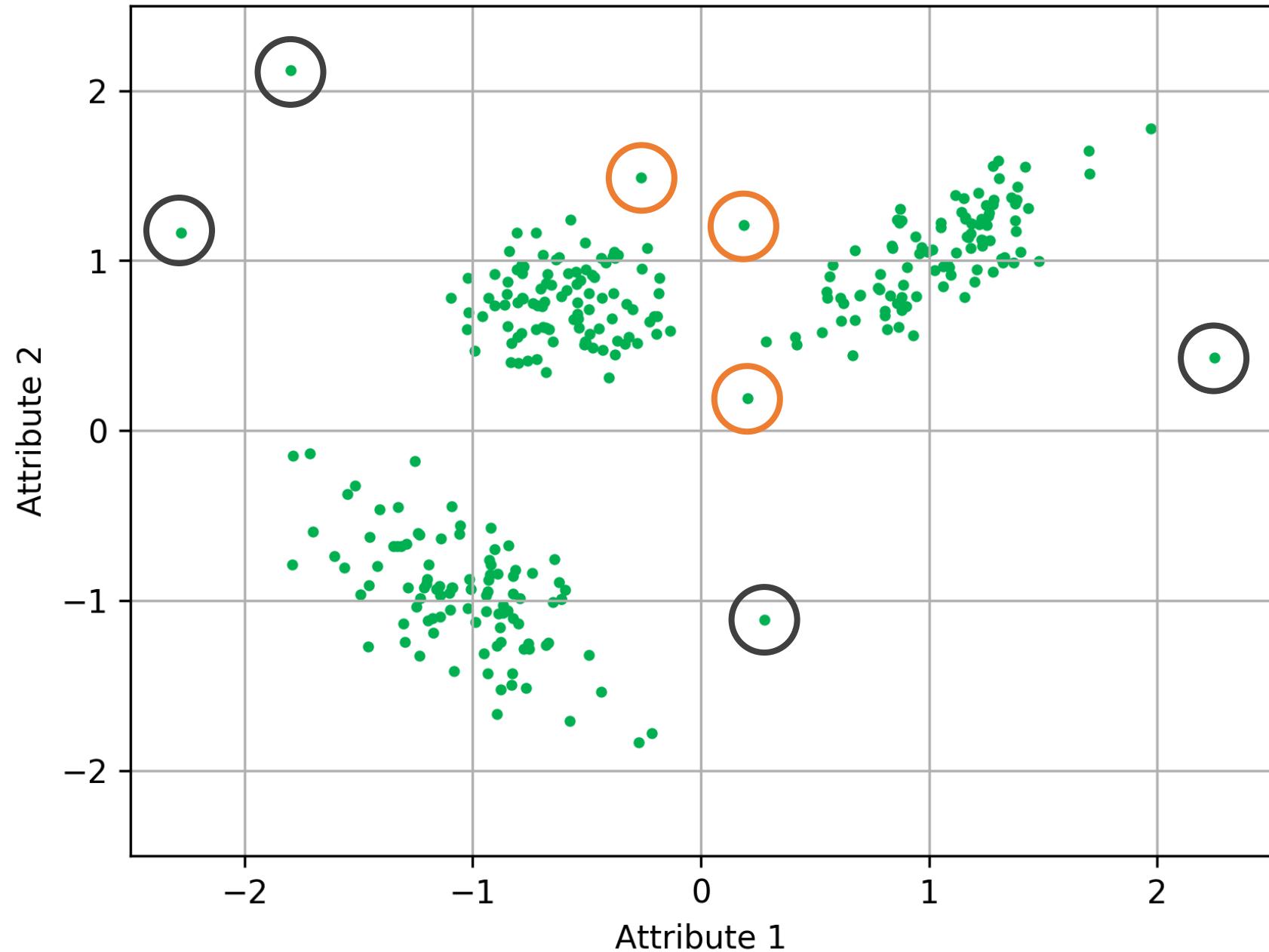
Credit Fraud



Credit Fraud



Credit Fraud



Credit Fraud

Input Data:

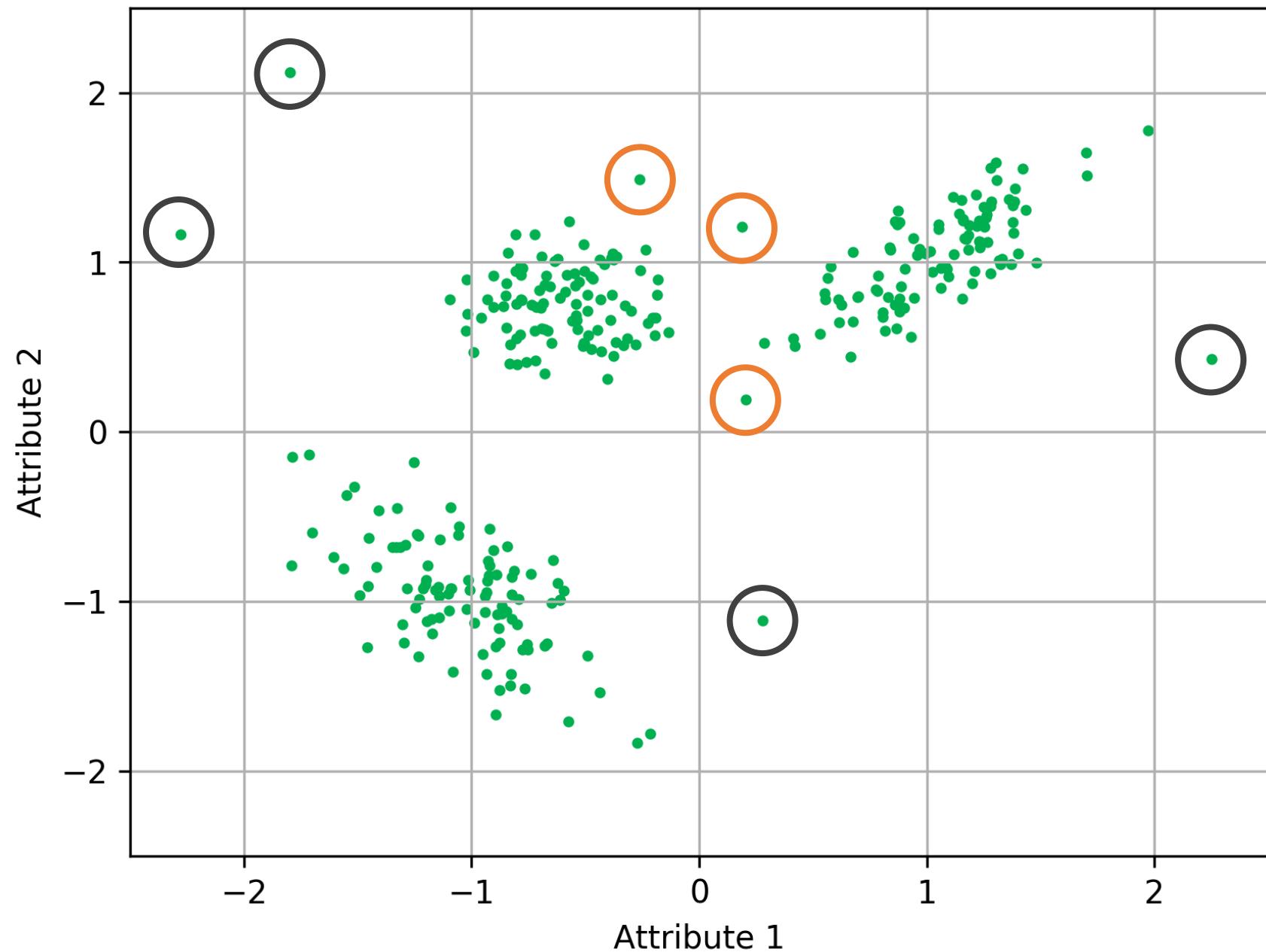
Account transactions, dates,
locations, demographic
information
(Numerical and categorical)

Target Data:

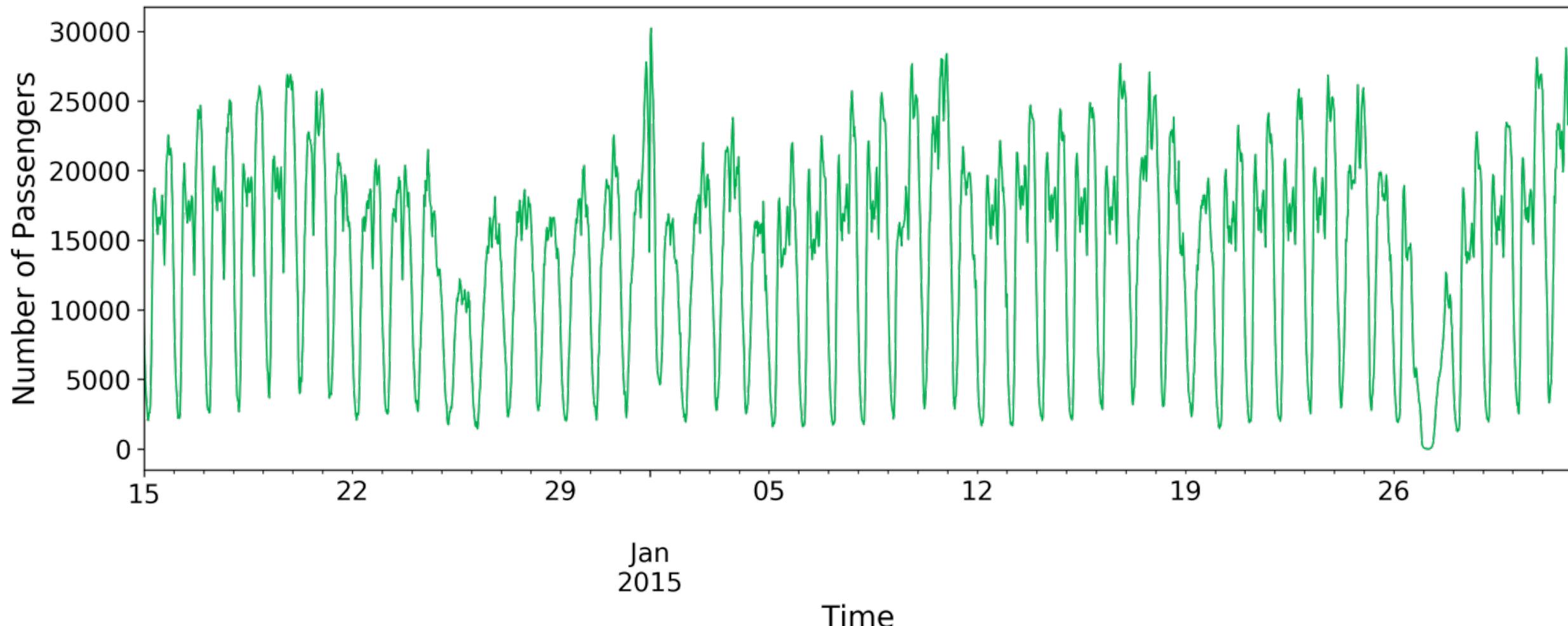
Anomalous transactions

Learning Category:

Unsupervised Learning
Clustering, Density
Estimation

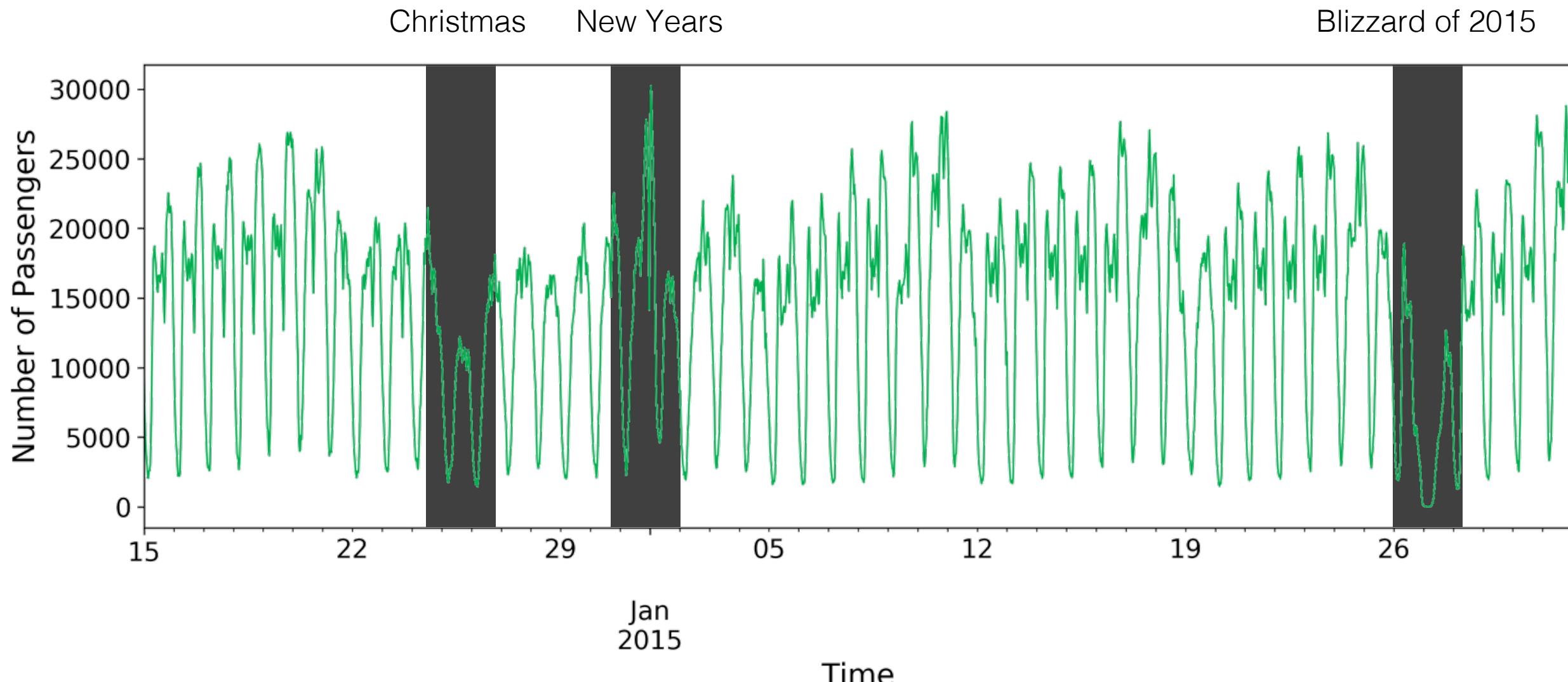


Anomalous Event Detection: NYC Taxis



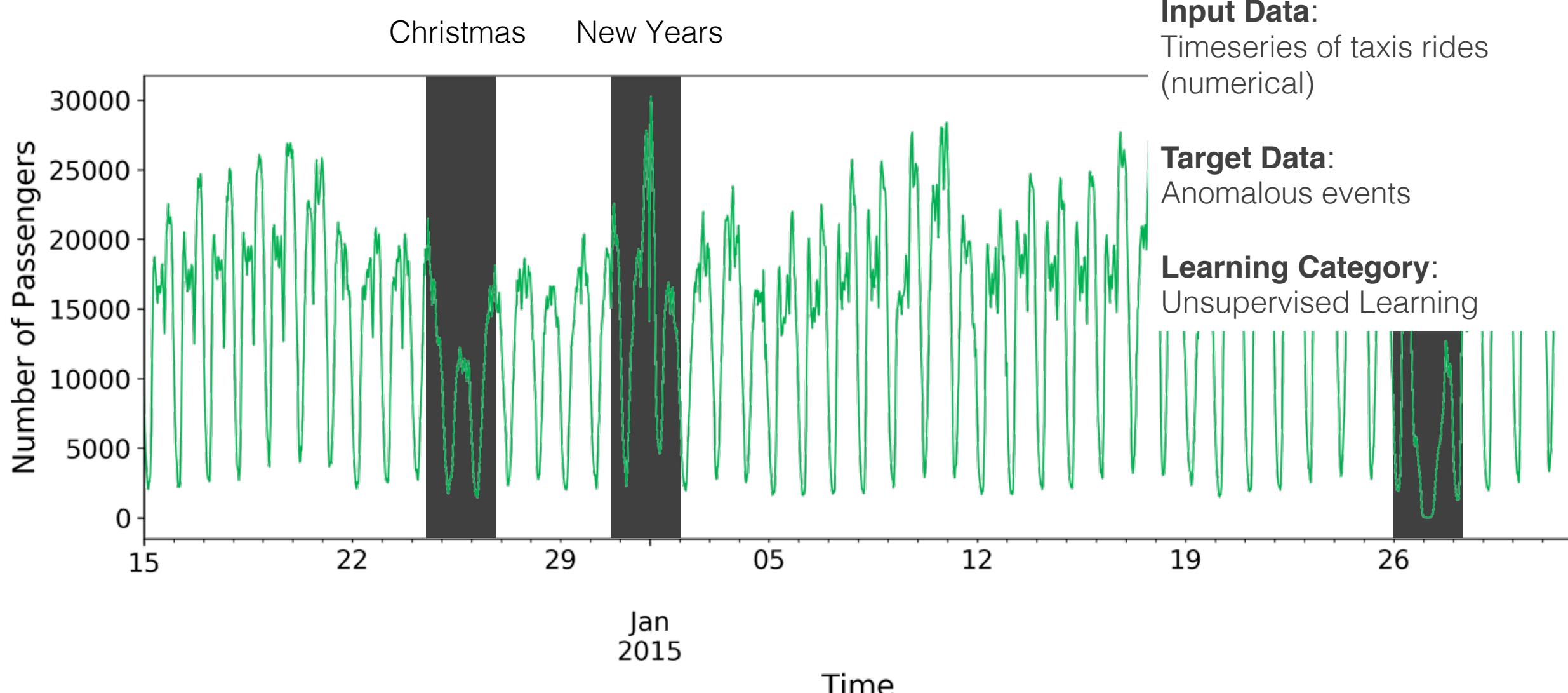
Data source: Numenta Anomaly Benchmark (NAB), from kaggle.com

Anomalous Event Detection: NYC Taxis



Data source: Numenta Anomaly Benchmark (NAB), from kaggle.com

Anomalous Event Detection: NYC Taxis



Data source: Numenta Anomaly Benchmark (NAB), from kaggle.com

Types of machine learning

	Unsupervised Learning	Supervised Learning	Reinforcement Learning
Goal	Describe ...structure in data	Predict ...from examples	Strategize learn through interaction
Data available	predictors, x	predictor and response pairs, (x, y)	predictors and delayed responses (called rewards)
Types	<ul style="list-style-type: none">• Density estimation• Clustering• Dimensionality reduction• Anomaly detection	<ul style="list-style-type: none">• Classification• Regression	<ul style="list-style-type: none">• Model-free learning• Model-based learning

Learning a strategy to master games



THE ULTIMATE GO CHALLENGE
GAME 3 OF 3
27 MAY 2017

vs

 AlphaGo
Winner of Match 3

 Ke Jie

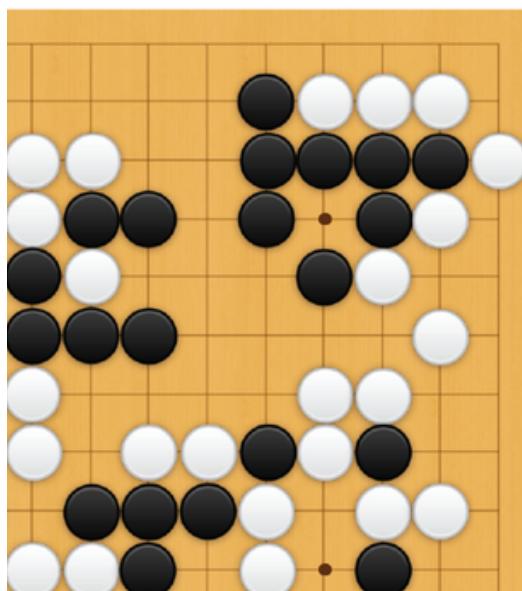
RESULT B + Res

 Google DeepMind

Learning a strategy to master games

Input Data:

Moves taken and occasional feedback on win/loss
(Numerical and categorical)

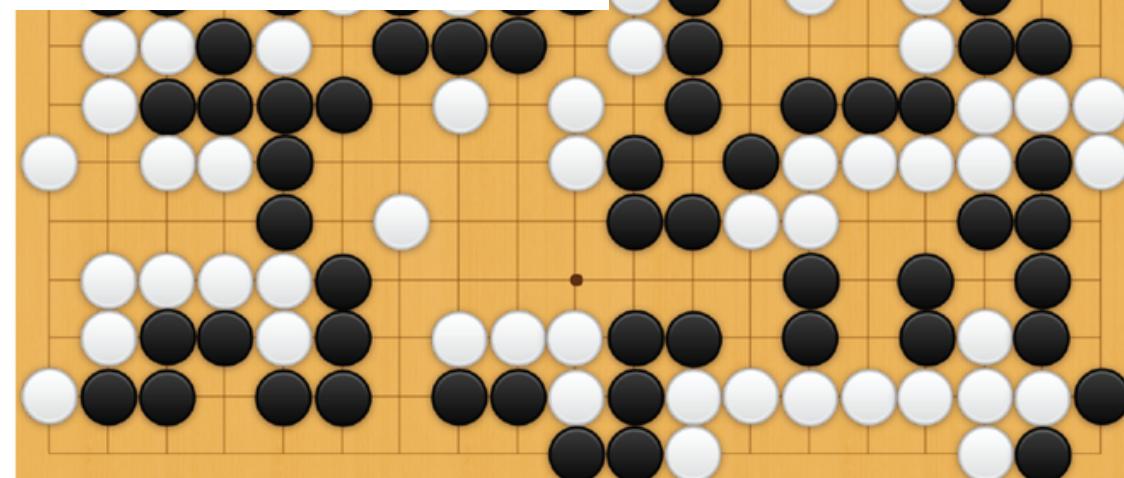


Target Data:

Win/loss (Maximizing rewards)

Learning Category:

Reinforcement Learning



THE ULTIMATE GO CHALLENGE

GAME 3 OF 3

27 MAY 2017



vs



AlphaGo

Winner of Match 3

Ke Jie

RESULT B + Res



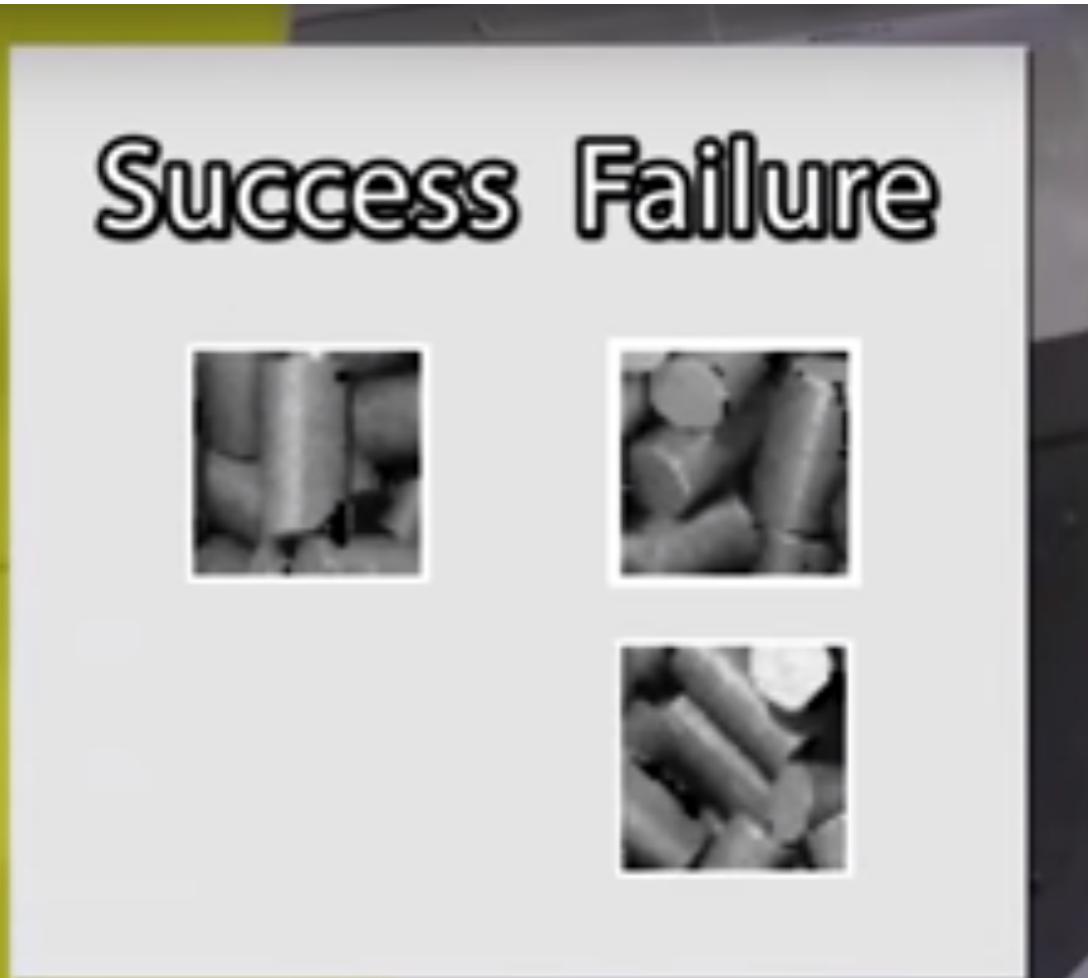
Google DeepMind

Manufacturing – learn to pick up iron cylinders



Source: MIT Technology Review; Company: **FANUC**

Manufacturing – learn to pick up iron cylinders



Input Data:

Actions taken and occasional feedback on success/failure
(Numerical and categorical)

Target Data:

Success/failure (Maximizing rewards)

Learning Category:

Reinforcement Learning



Types of machine learning

	Unsupervised Learning	Supervised Learning	Reinforcement Learning
Goal	Describe ...structure in data	Predict ...from examples	Strategize learn through interaction
Data available	predictors, x	predictor and response pairs, (x, y)	predictors and delayed responses (called rewards)
Types	<ul style="list-style-type: none">• Density estimation• Clustering• Dimensionality reduction• Anomaly detection	<ul style="list-style-type: none">• Classification• Regression	<ul style="list-style-type: none">• Model-free learning• Model-based learning

Course logistics

Learning objectives

- Fundamentals of **machine learning** and **data science**
- **Structure** a machine learning problem
- Automatically make **decisions** from data
- Understand the techniques/algorithms and when to use them
- **Communicate** and effectively **interpret** machine learning decisions

Through this course you will...

- **Apply** machine learning techniques and **communicate** findings
- Compete in a **Kaggle machine learning competition**
- Create an **online data science portfolio**
- Implement your own **end-to-end machine learning project**
- Create a **video** about your project

Graded Components

Class participation	5%
In-class quizzes	10%
Assignments (5 assignments)	35%
Kaggle competition	20%
Final project	30%

Course website

kylebradbury.github.io/ece590

Action items

- Get the textbook (also available free online)
- Log onto Piazza (make sure you can access it and ask questions)
- Complete the first reading
- Create an account at Poll Everywhere using [this link](#). Be sure to use (a) your real name, and (b) your Duke netid-based email address (e.g. mjh8@duke.edu)
- Begin working on Assignment #1

THIS IS YOUR MACHINE LEARNING SYSTEM?

YUP! YOU POUR THE DATA INTO THIS BIG
PILE OF LINEAR ALGEBRA, THEN COLLECT
THE ANSWERS ON THE OTHER SIDE.

WHAT IF THE ANSWERS ARE WRONG?

JUST STIR THE PILE UNTIL
THEY START LOOKING RIGHT.

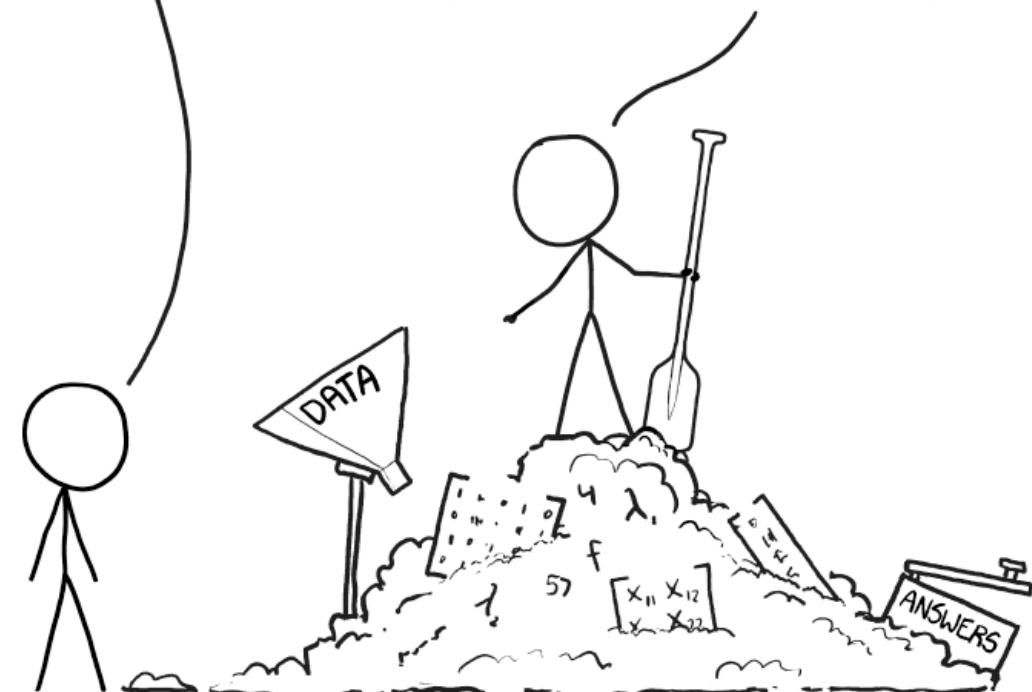


Image: xkcd.com