Machine Learning I

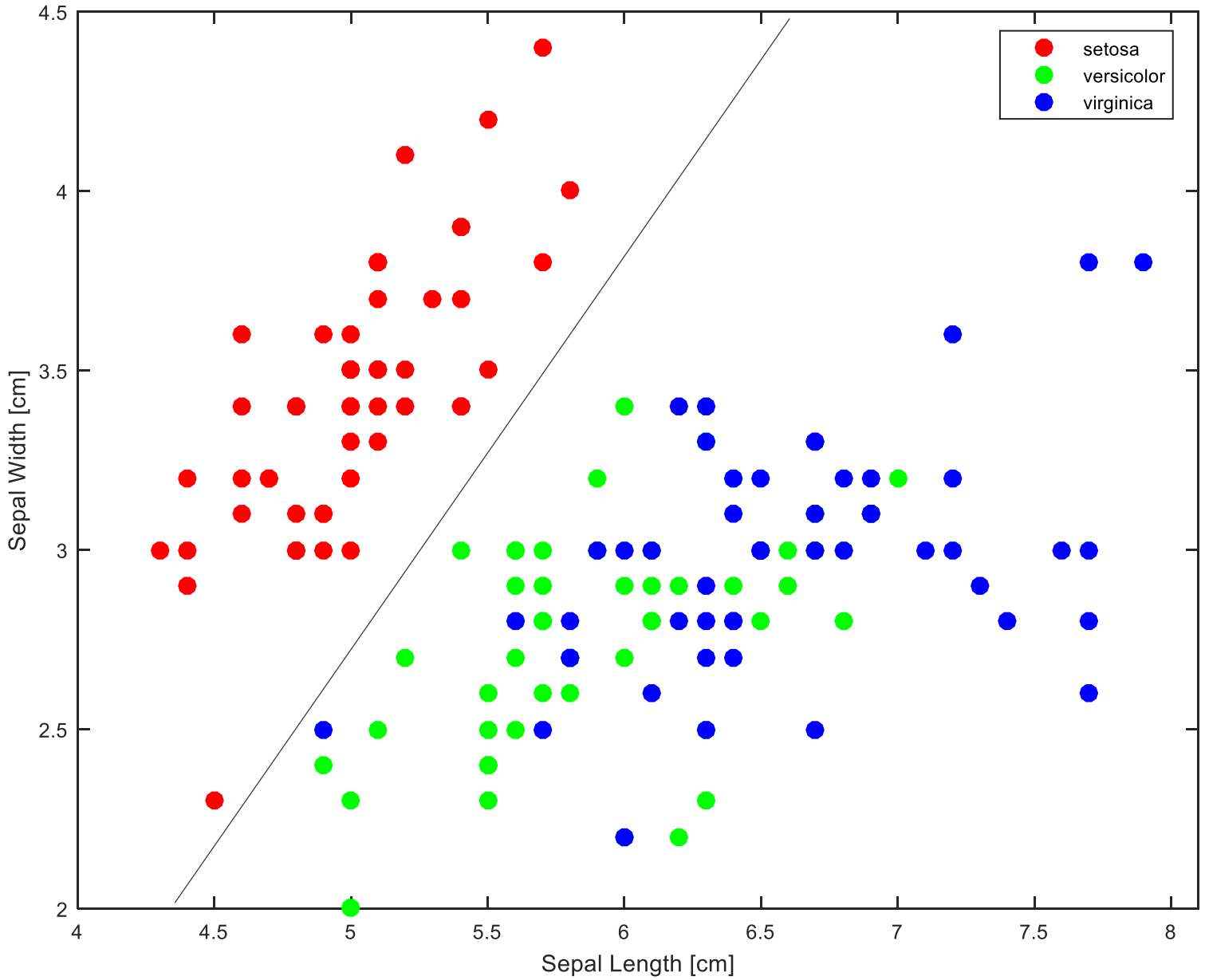
What is machine learning?

How can you tell these flowers apart?

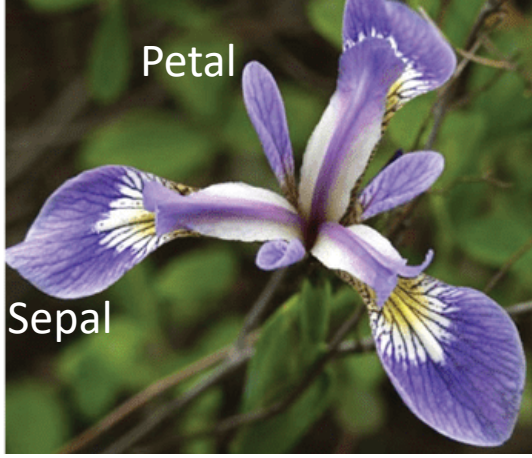
Iris setosa

Iris virginica

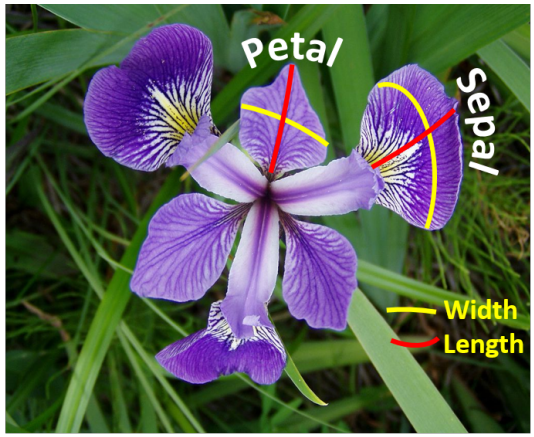
Image Sources: Srishti Sawla (setosa) and Ivo Dinov, University of Michigan SOCR (versicolor and verginica)



setosa



versicolor



virginica

Data Source: Fisher Iris Data
Image Sources: Srishti Sawla (setosa) and Ivo Dinov, University of Michigan SOCR (versicolor and verginica)

Challenges

What is this?

Image by artist Hikaru Cho

We **generalize** from past experiences

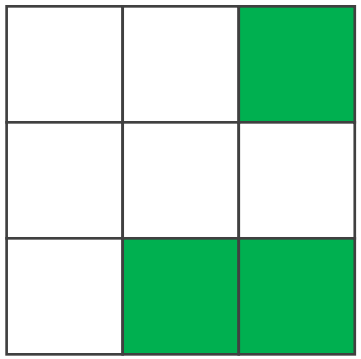
Image: “It’s not what it seems” by artist Hikaru Cho

our data must be

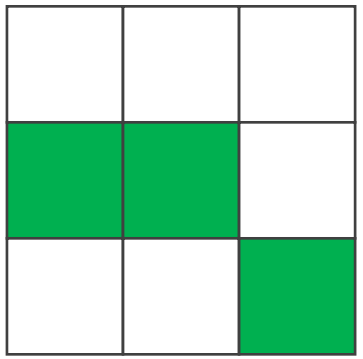
representative

Predict which class x_{new} belongs to…

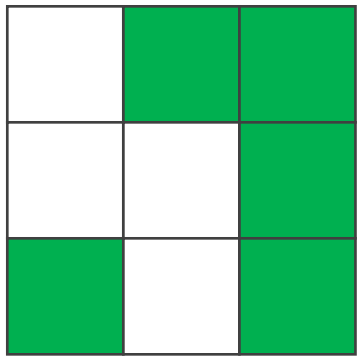
$f(\boldsymbol{x}) = 1$



$\boldsymbol{x}_0$

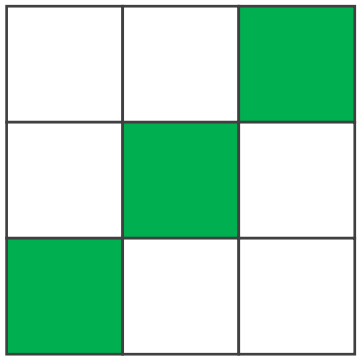


$\boldsymbol{x}_2$

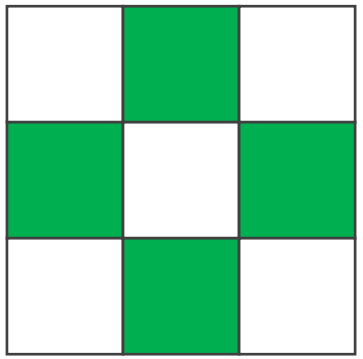


$\boldsymbol{x}_4$

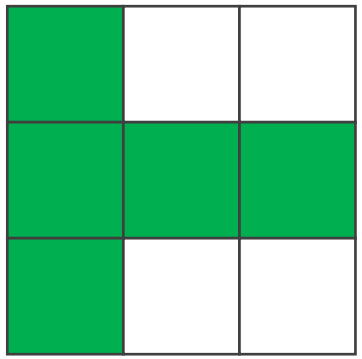
$f(\boldsymbol{x}) = 0$



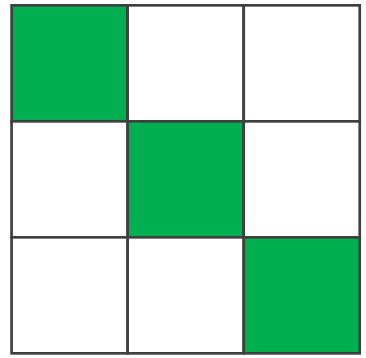
$\boldsymbol{x}_1$



$\boldsymbol{x}_3$



$\boldsymbol{x}_5$



$\boldsymbol{x}_{new}$

$f(\boldsymbol{x}_{new}) = ?$

Example credit: Yaser Abu-Mostafa, 2012

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

There are often **many** models that fit your training data similarly well

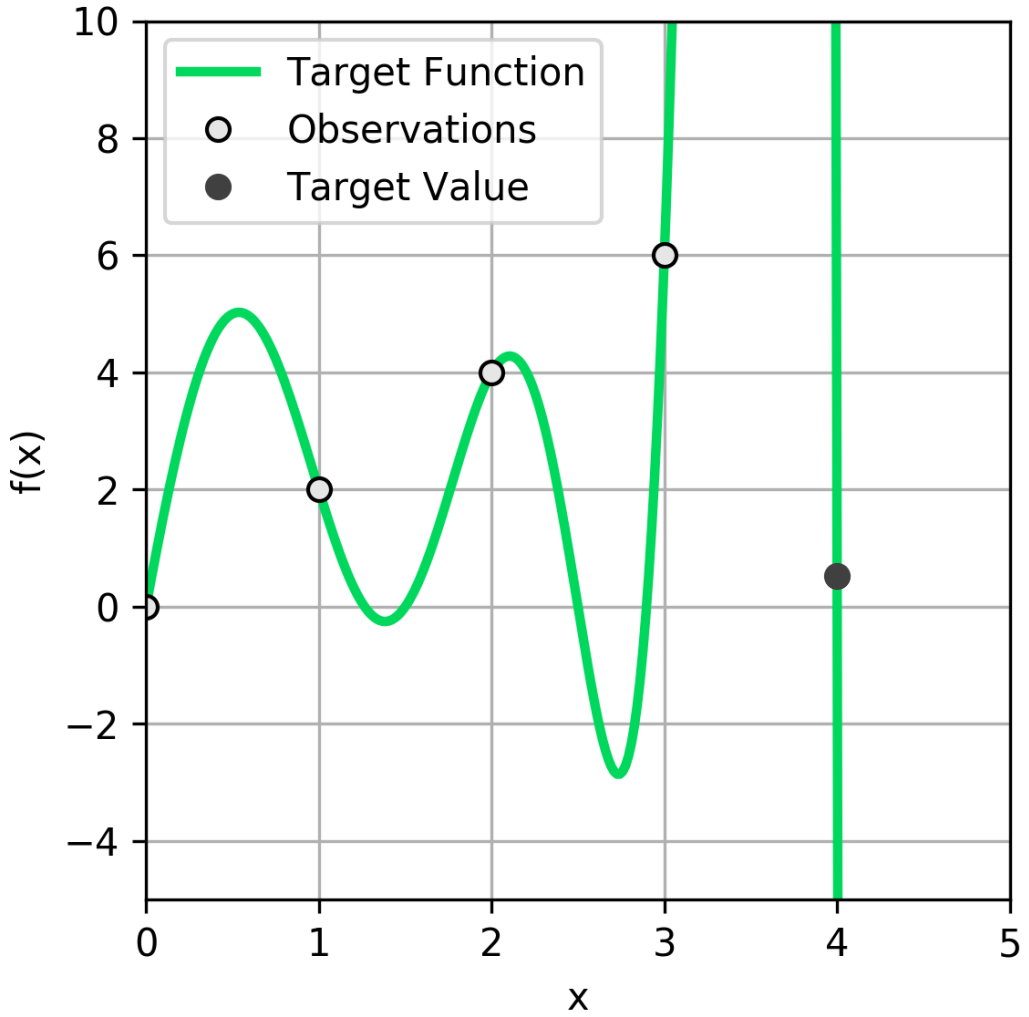
So how do we choose which to use?

the best models **generalize** well

Predict the next value in the sequence…

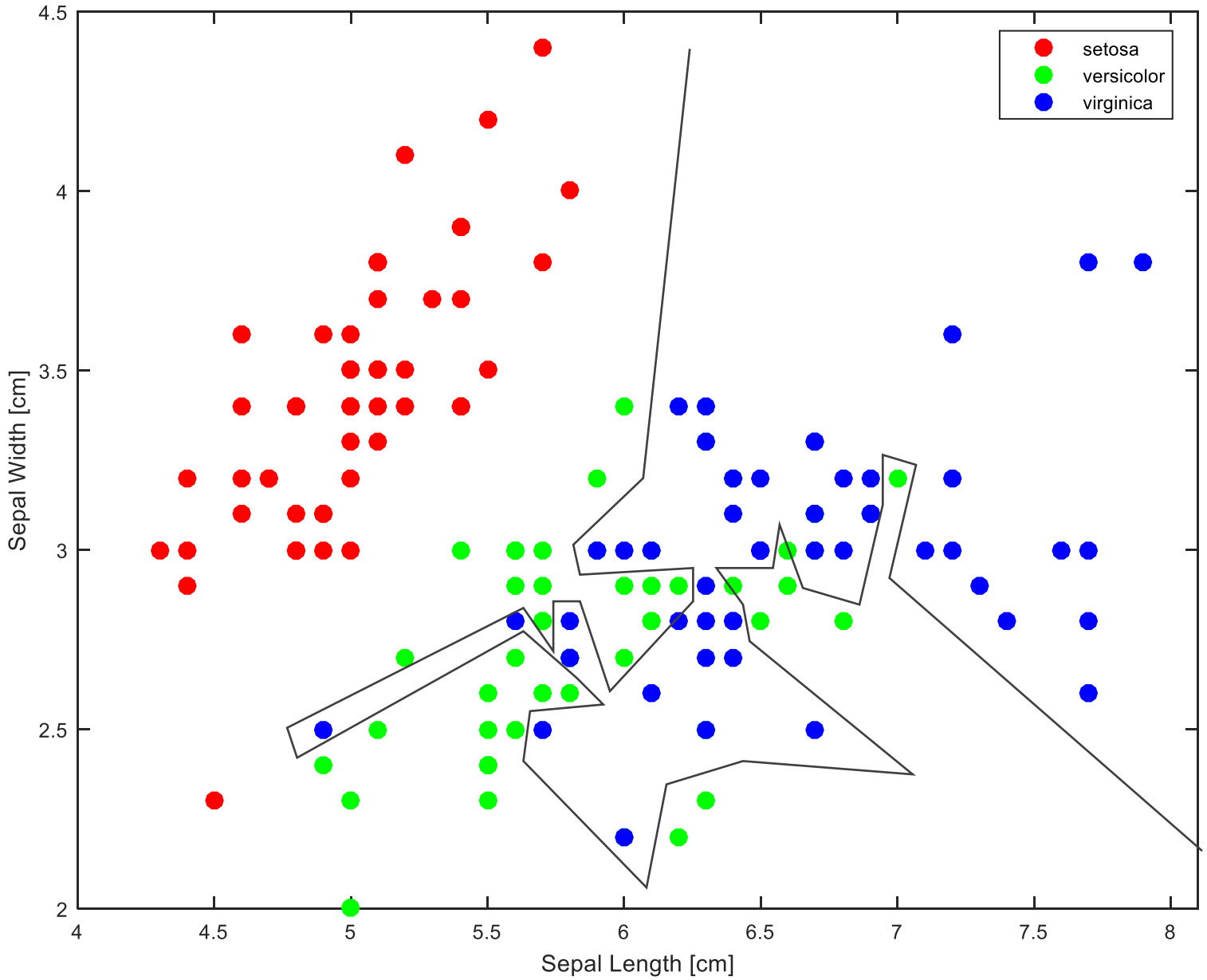
x	0	1	2	3	4
$f(x)$	**0**	**2**	**4**	**6**	**?**

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



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

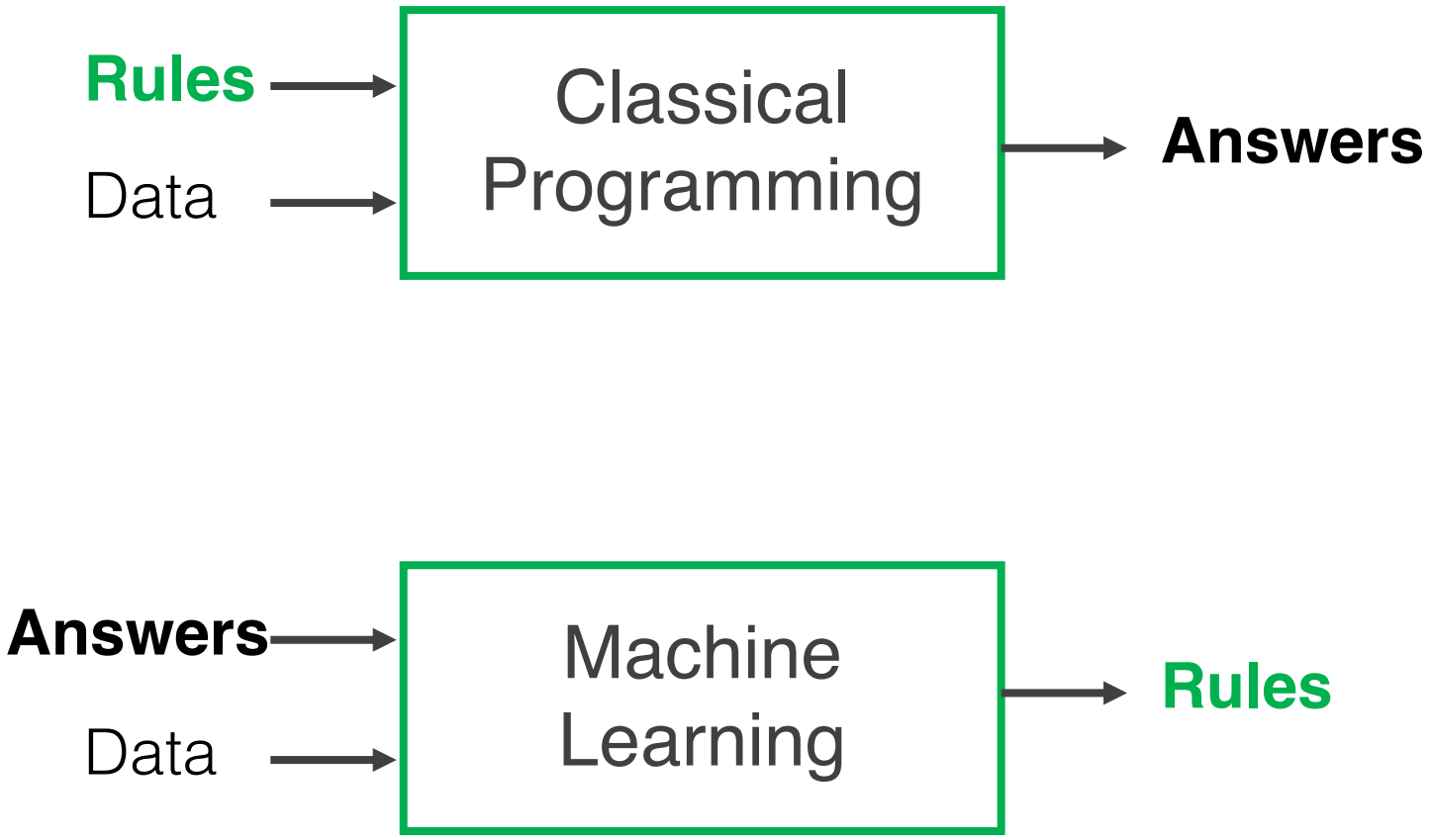
Data Source: Fisher Iris Data

Complex models overfit to the data

overfit works against generalization

Learning simpler representations of data enables learning

Machine learning suggests an alternative programming paradigm



François Chollet, *Deep Learning with Python*, 2017

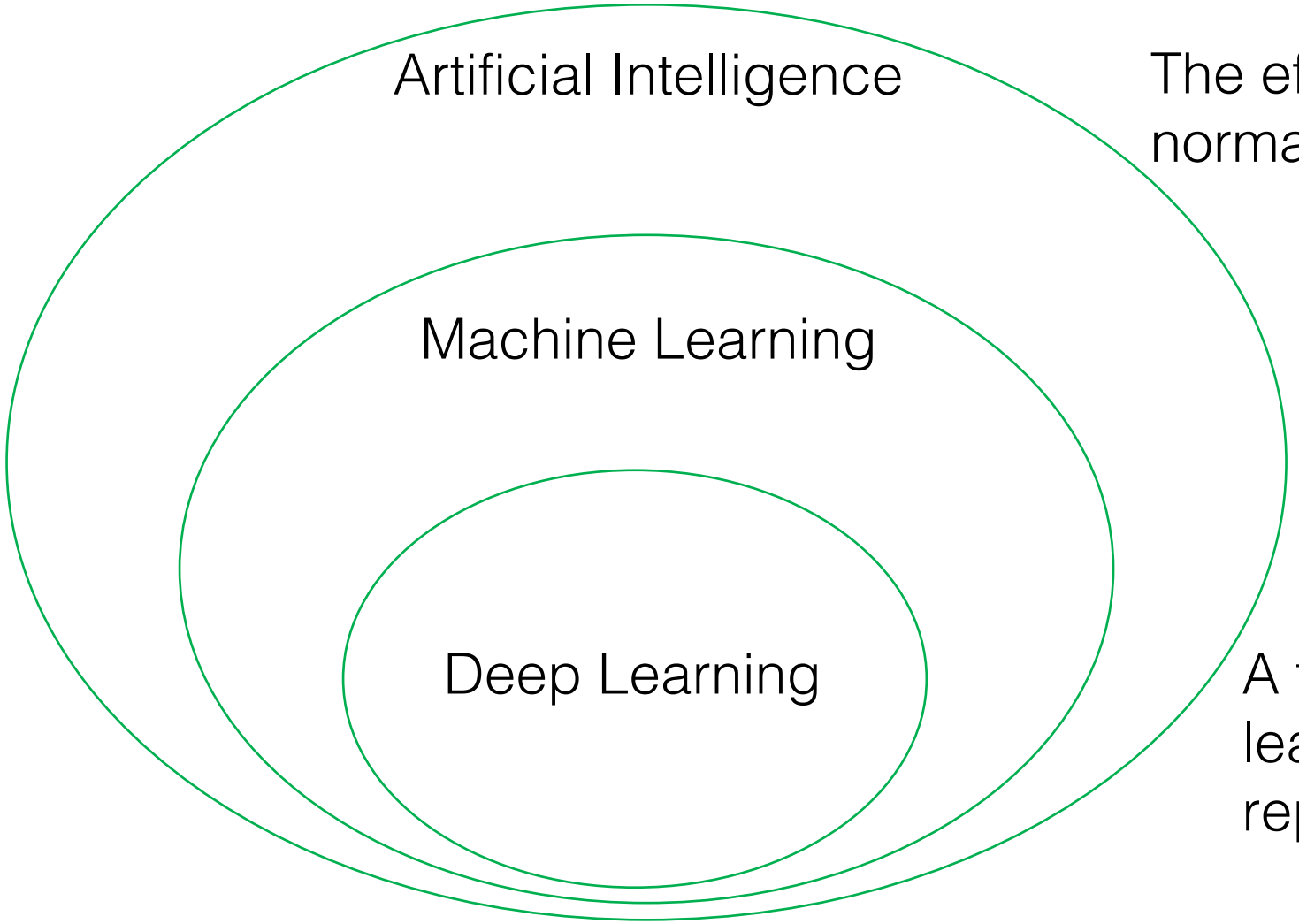
What is machine learning?

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

…**based on** data and past experience…

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

What is machine learning?



The effort to automate intellectual tasks normally performed by humans

Techniques to automatically describe, predict, and strategize based on data, without being explicitly programmed

A type of machine learning that learns successive layers of data representations

François Chollet, *Deep Learning with Python*, 2017

Types of machine learning tools

Types of learning	Common use case
Unsupervised learning	**Describe**
Supervised learning	**Predict**
Reinforcement learning	**Strategize**

Types of machine learning

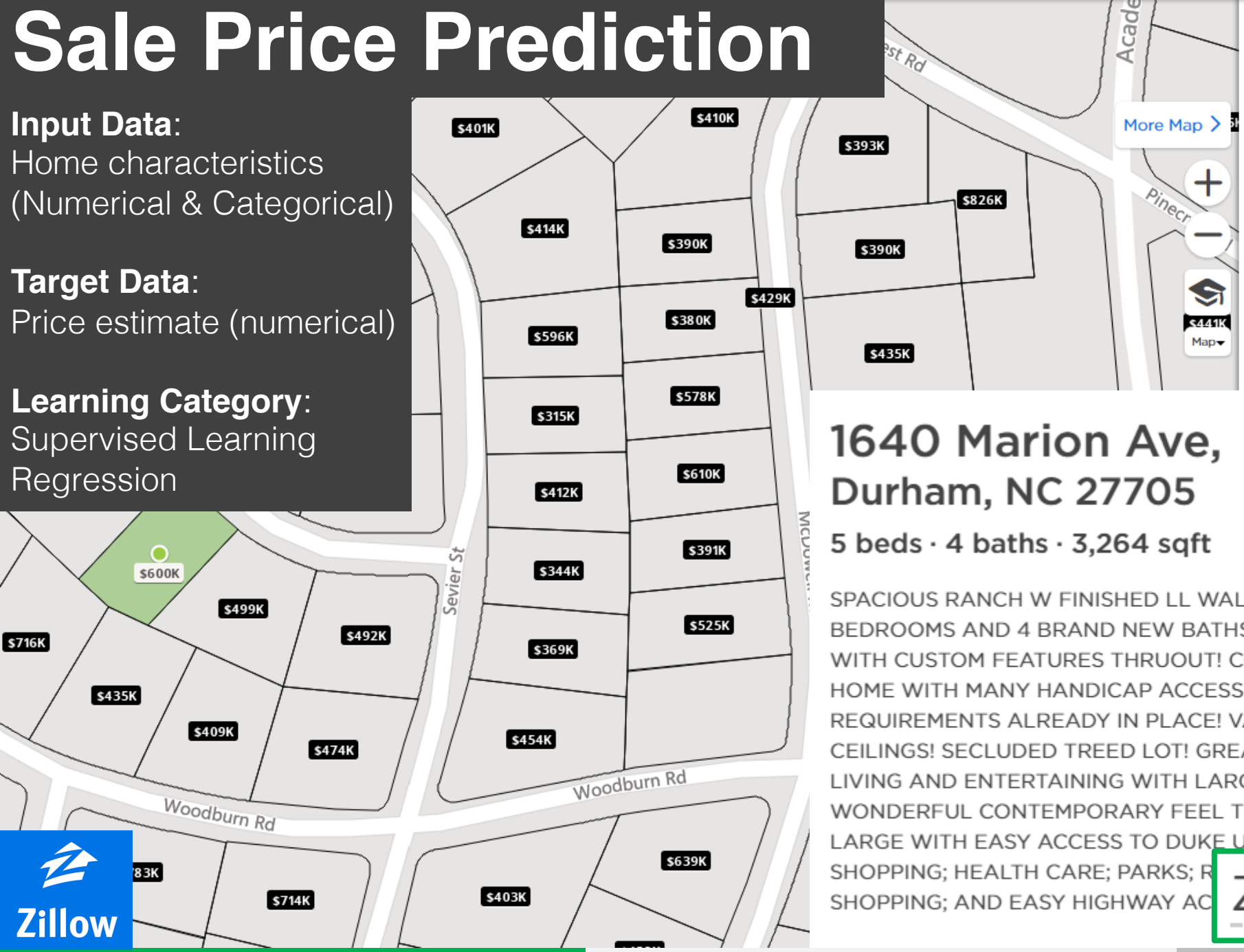
	Supervised Learning	Unsupervised Learning	Reinforcement Learning
Goal	**Predict** …from examples	**Describe** ...structure in data	**Strategize** learn by trial and error
Data	(x, y)	x	delayed feedback
Types	• Classification • Regression	• Density estimation • Clustering • Dimensionality reduction • Anomaly detection	• Model-free learning • Model-based learning

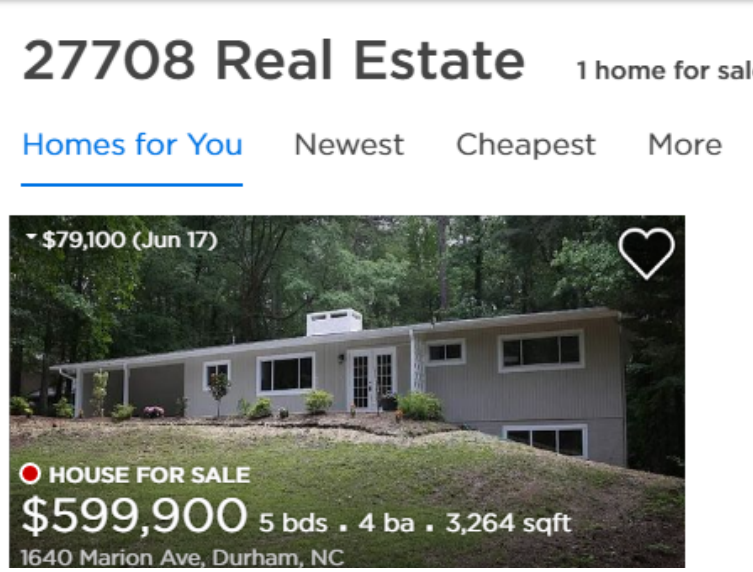
Sale Price Prediction

Input Data:
Home characteristics
(Numerical & Categorical)

Target Data:
Price estimate (numerical)

Learning Category:
Supervised Learning
Regression





1640 Marion Ave, Durham, NC 27705

5 beds · 4 baths · 3,264 sqft

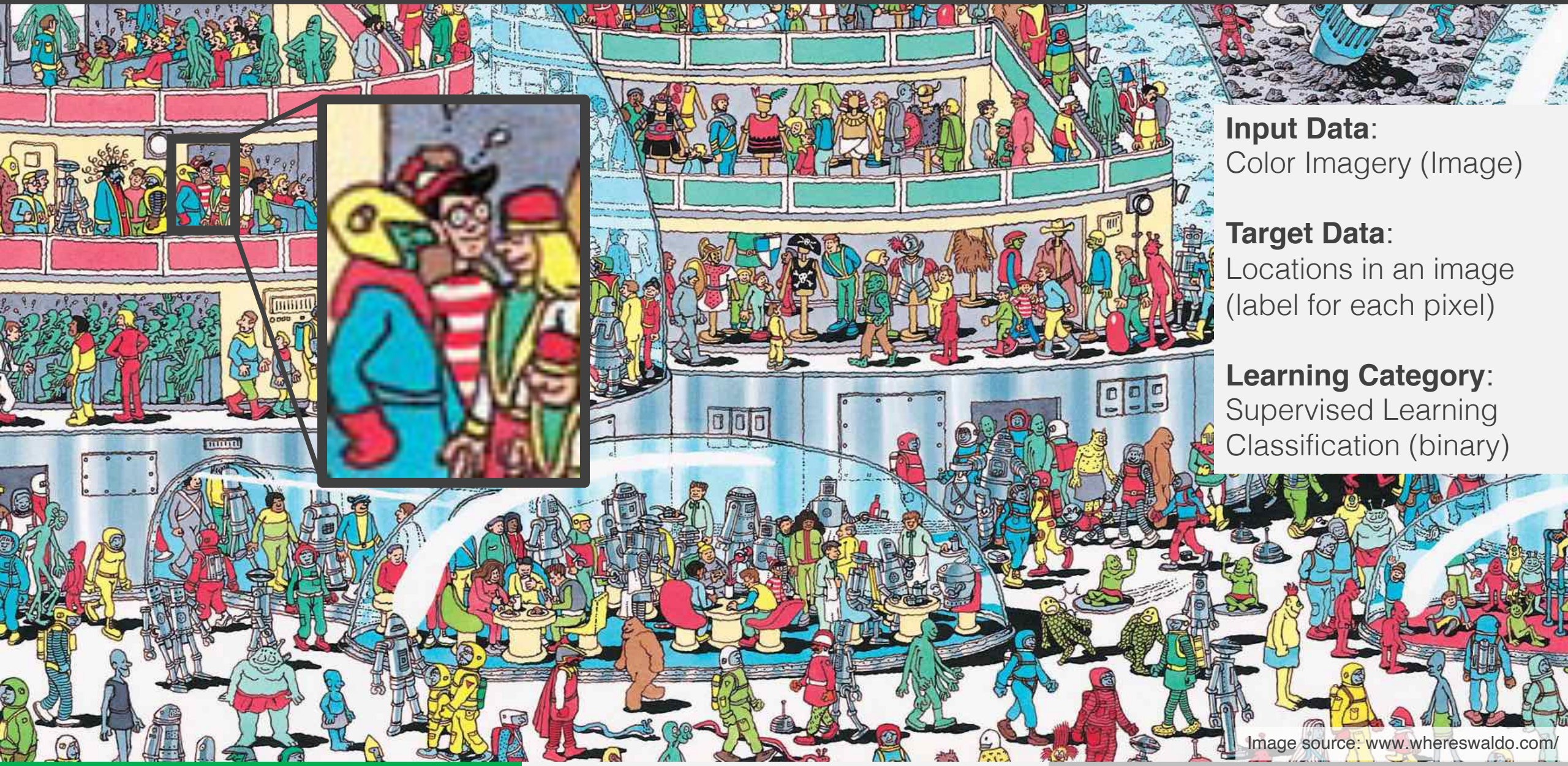
SPACIOUS RANCH W FINISHED LL WALKOUT! 5 BEDROOMS AND 4 BRAND NEW BATHS! RENOVATED WITH CUSTOM FEATURES THRUOUT! CONTEMPORARY HOME WITH MANY HANDICAP ACCESSIBLE REQUIREMENTS ALREADY IN PLACE! VAULTED CEILINGS! SECLUDED TREED LOT! GREAT HOME FOR LIVING AND ENTERTAINING WITH LARGE REAR DECK! WONDERFUL CONTEMPORARY 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

Where's Waldo = Computer Vision Problem



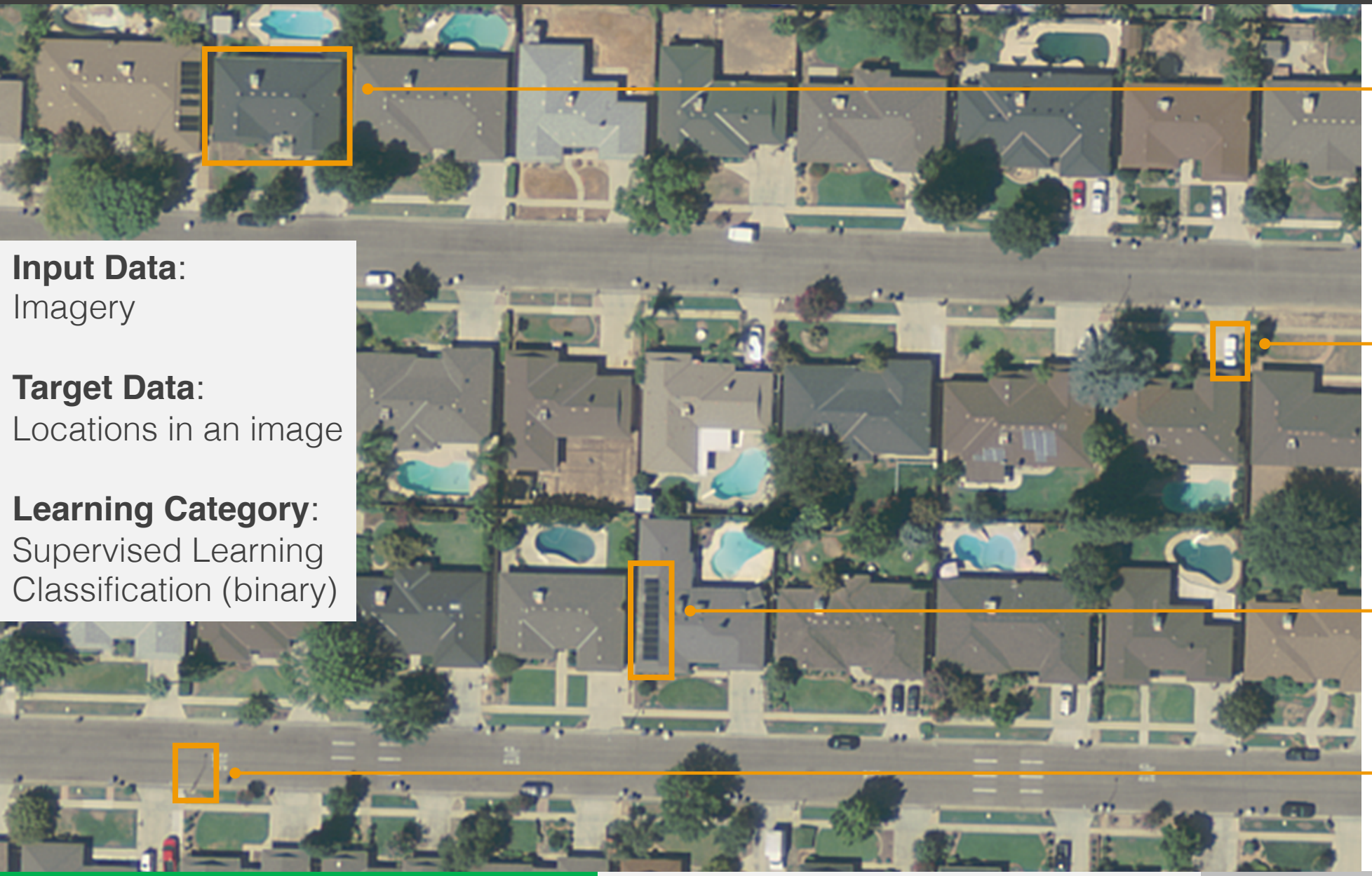
Input Data:
Color Imagery (Image)

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



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

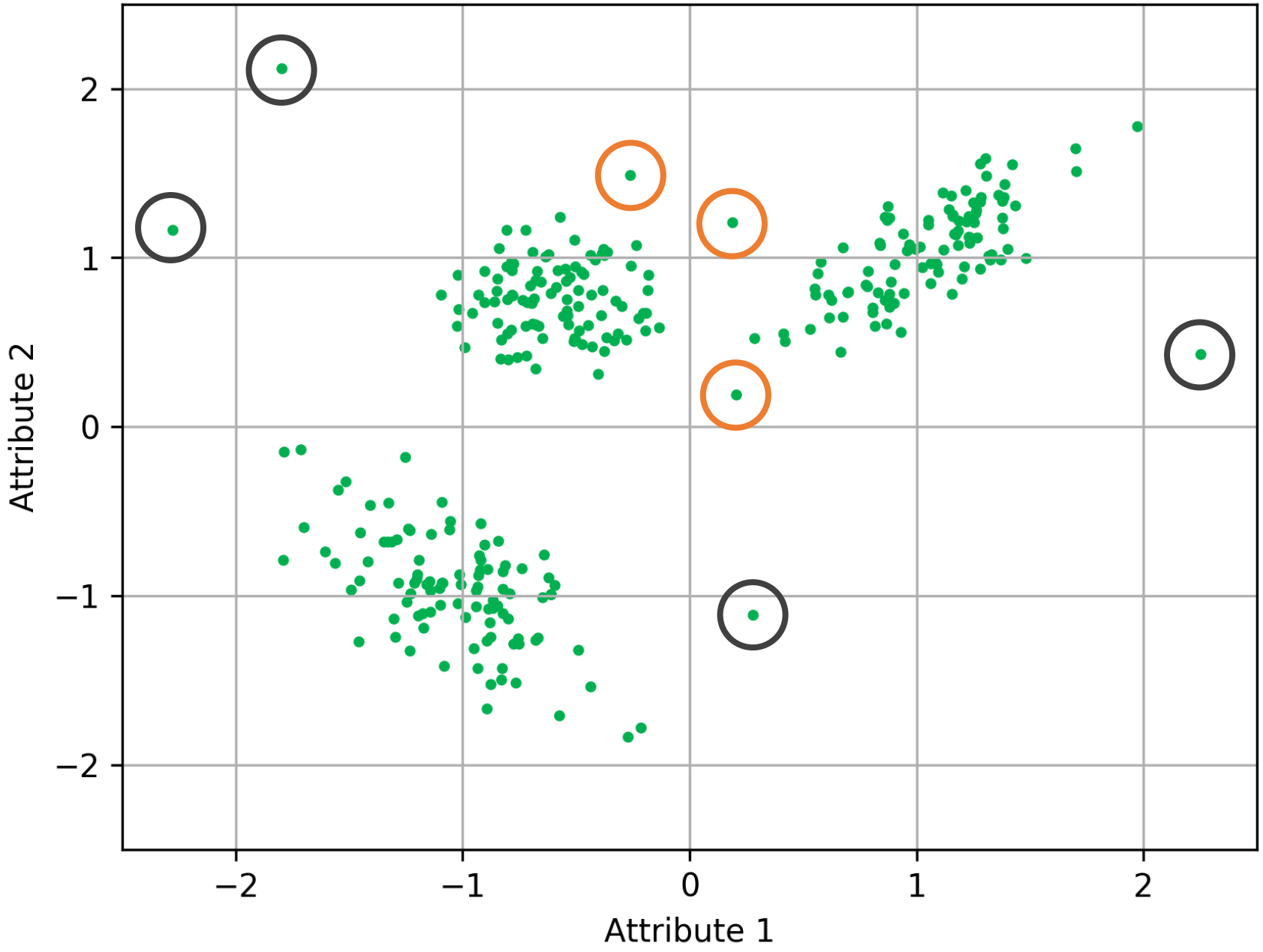
Light Pole
access to electricity

Credit Fraud

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

Target Data:
Anomalous transactions

Learning Category:
Unsupervised Learning
Clustering, Density Estimation

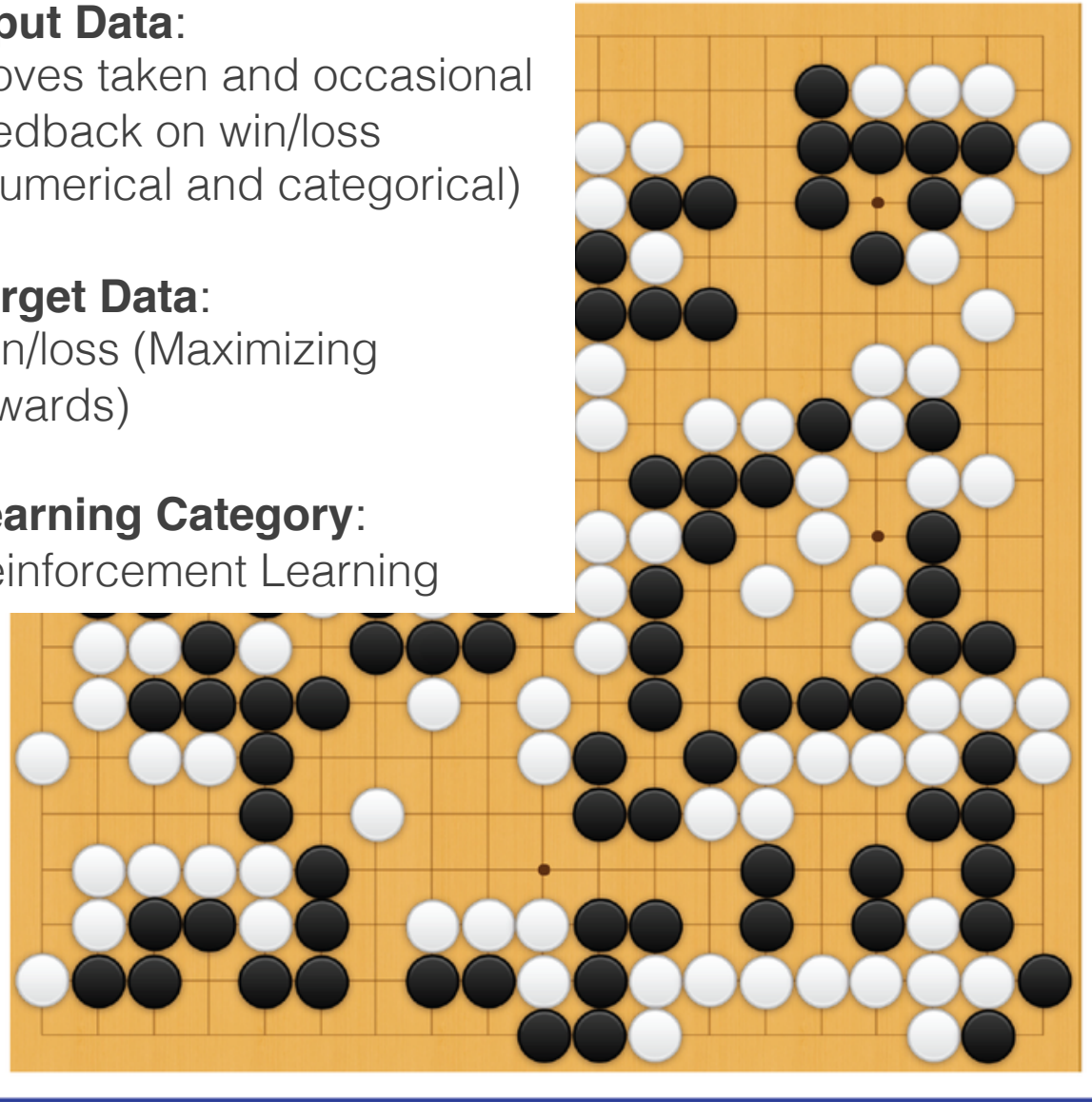


Learning a strategy to master games

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

Target Data:
Win/loss (Maximizing rewards)

Learning Category:
Reinforcement Learning



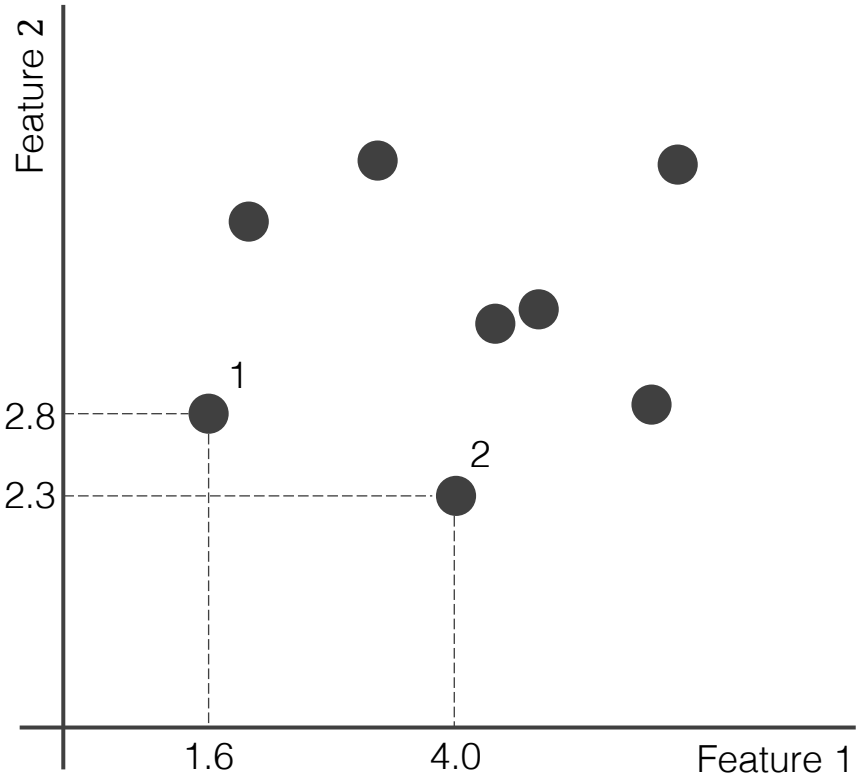
Types of machine learning

	Supervised Learning	Unsupervised Learning	Reinforcement Learning
Goal	**Predict** …from examples	**Describe** ...structure in data	**Strategize** learn by trial and error
Data	(x, y)	x	delayed feedback
Types	• Classification • Regression	• Density estimation • Clustering • Dimensionality reduction • Anomaly detection	• Model-free learning • Model-based learning

Supervised Learning

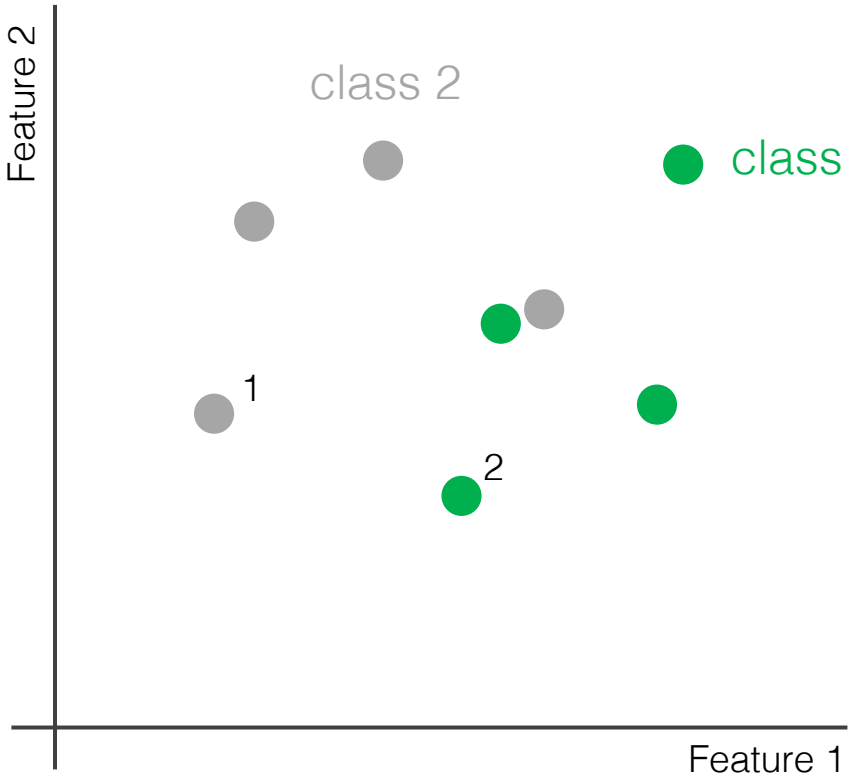
Components of supervised learning

Data



	Feature 1	Feature 2
Data Sample 1	1.6	2.8
Data Sample 2	4.0	2.3
…	###	###

Labels



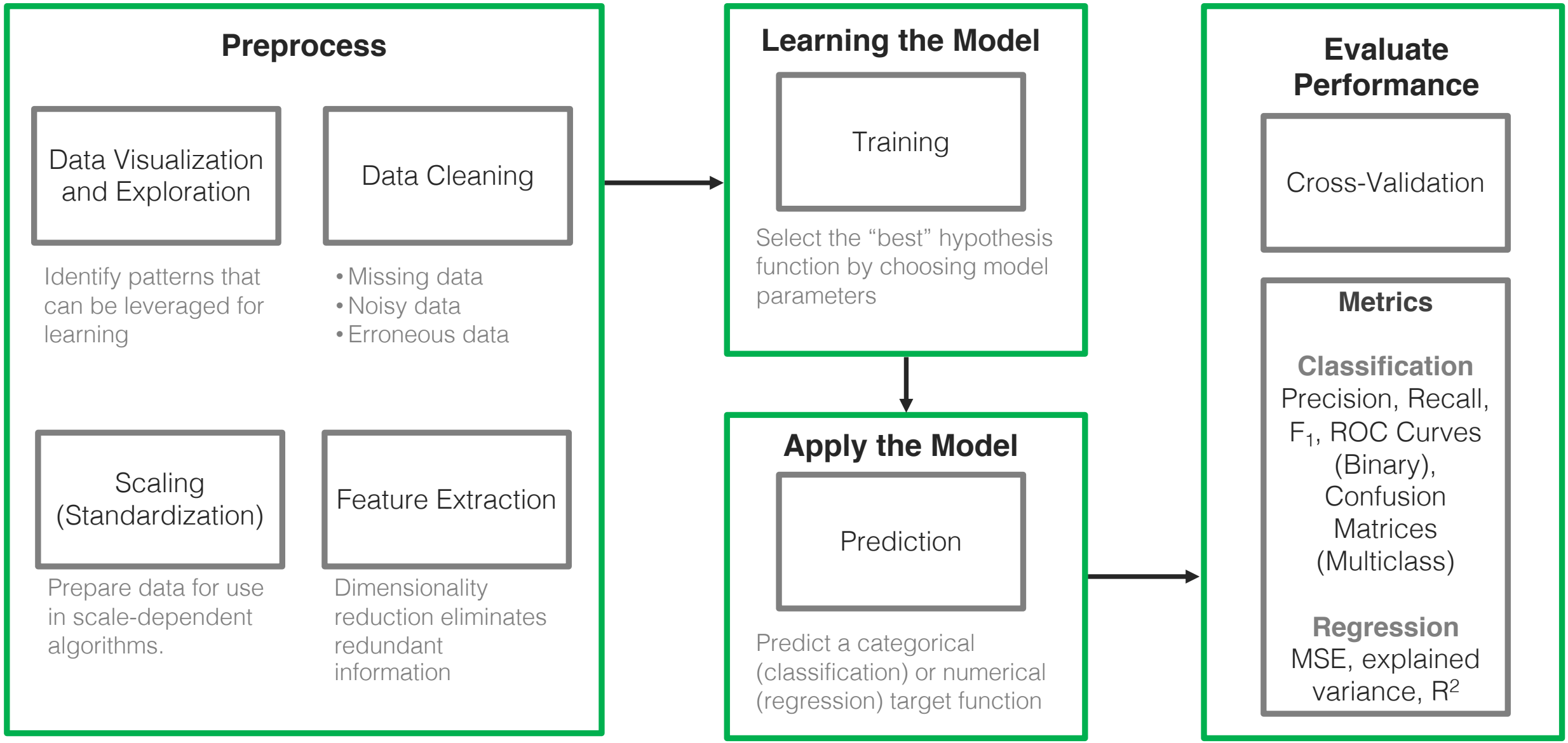
	Labels
Data Sample 1	Class 1
Data Sample 2	Class 2
…	###

Performance Metric

% of samples correctly classified
(classification)

Mean squared error
(regression)

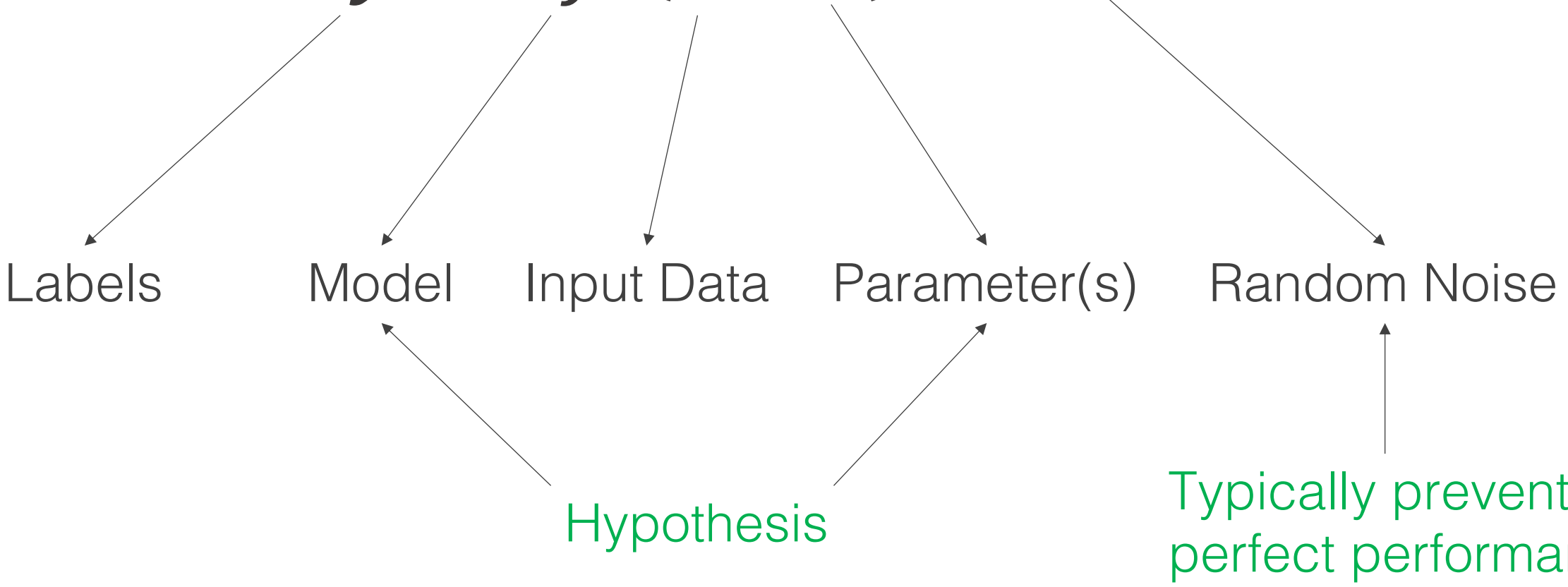
Supervised learning in practice



Supervised machine learning model

We search for the model that best fits our data

$$y = f(x, \boldsymbol{w}) + \epsilon$$



Algorithm Development

Application

(iterate as needed)

1. **Load training data**
2. **Preprocess & extract features**
3. **Choose learning model**
4. **Evaluate performance**
5. **Make predictions on unseen data**

y

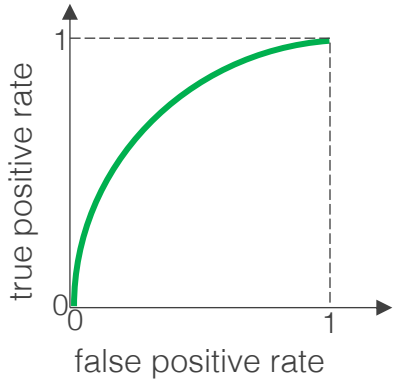
1
1
0
0
1
0

X

X'

x_1	0.38	0.39	0.85	0.78
x_2	0.81	0.91	0.97	0.53
x_3	0.65	0.59	0.91	0.11
x_4	0.94	0.05	0.40	0.26
x_5	0.27	0.19	0.03	0.64
x_6	0.02	0.98	0.36	0.11

Fisher's linear discriminant
perceptron
logistic regression
decision trees
random forests
support vector machine
k nearest neighbors
neural networks



$\mathbf{x}_{new}$ $\hat{y}_{new}$


→ 1


→ 0

Components of supervised learning

Input	$\boldsymbol{x}$	
Output	y	
Training Data	$(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_N, y_N)$	
Target function	$f(\boldsymbol{x}) \rightarrow y$	This is unknown, but the best you could ever do
Hypothesis set	$f_i(\boldsymbol{x}) \rightarrow \hat{y}$	Functions to consider in trying to approximate $f(x)$
Learning algorithm	Optimization technique that searches the hypothesis set for the function f_i that best approximates f (typically by choosing parameters in a model)	

Supervised Learning

Unobservable | Observable

Data Generating Process

$p(X, Y)$

Training Data

$(x_1, y_1), \ldots, (x_N, y_N)$

Learning Algorithm

Chooses a hypothesis, $\hat{f} = f_i$ based on the training data such that $\hat{f}(x) \approx f(x)$

Final Hypothesis

$\hat{f}(x) \rightarrow \hat{y}$

predictions

Target Function

The best function predicting y from x

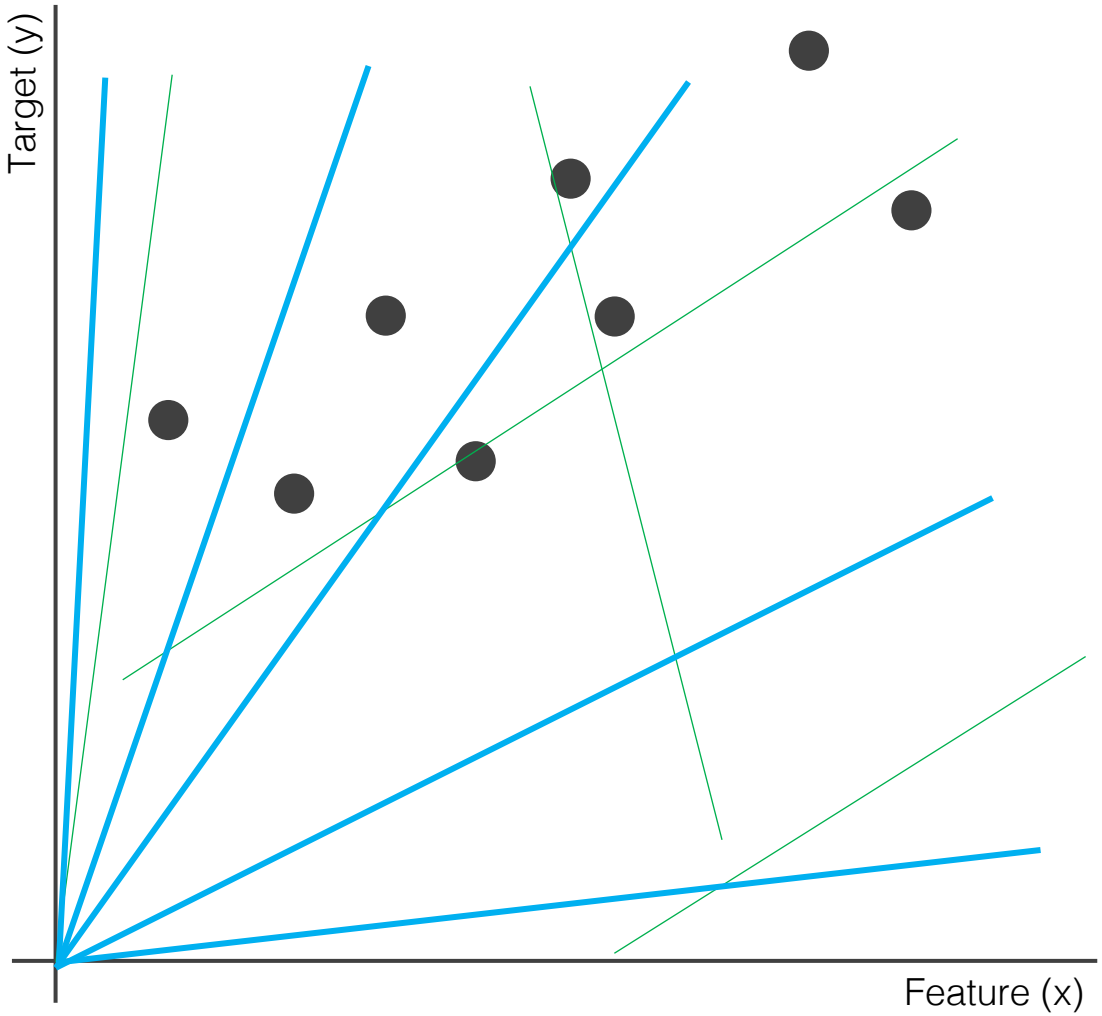
$f(x) \rightarrow y$

Hypothesis Functions Set

$f_1, f_2, f_3, \ldots$

- Need to select the hypothesis functions (models to train)
- Need to select the learning algorithm (for fitting the models to the data)

Example: linear regression



Using any line as a hypothesis function, how many possible hypothesis functions apply here?

Infinitely many

Using a the line $y = wx$ as the family of hypothesis functions, how many possible hypothesis functions apply here?

Infinitely many

Which set contains the better hypothesis?

Which set has more options to consider?

What is our learning algorithm?