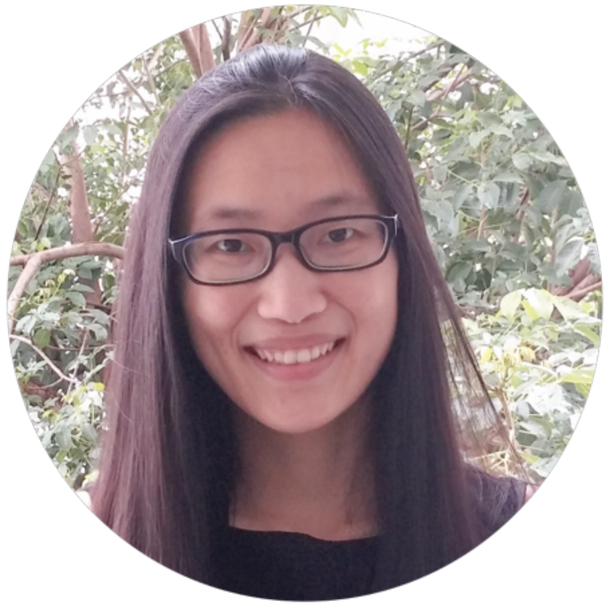
Introduction to Machine Learning

Instructor: Warasinee Chaisangmongkon, PhD



Warasinee Chaisangmongkon, PhD

Lecturer at Institute of Field Robotics, KMUTT. Data Scientist at Big Data Experience Center

- PhD from Yale University
- 10 years experiences in machine learning & data mining
- Lead scientist of corporate data science projects (banking, business digitizations)
- Found IDEA LAB - R&D in machine learning for businesses

Course Objectives

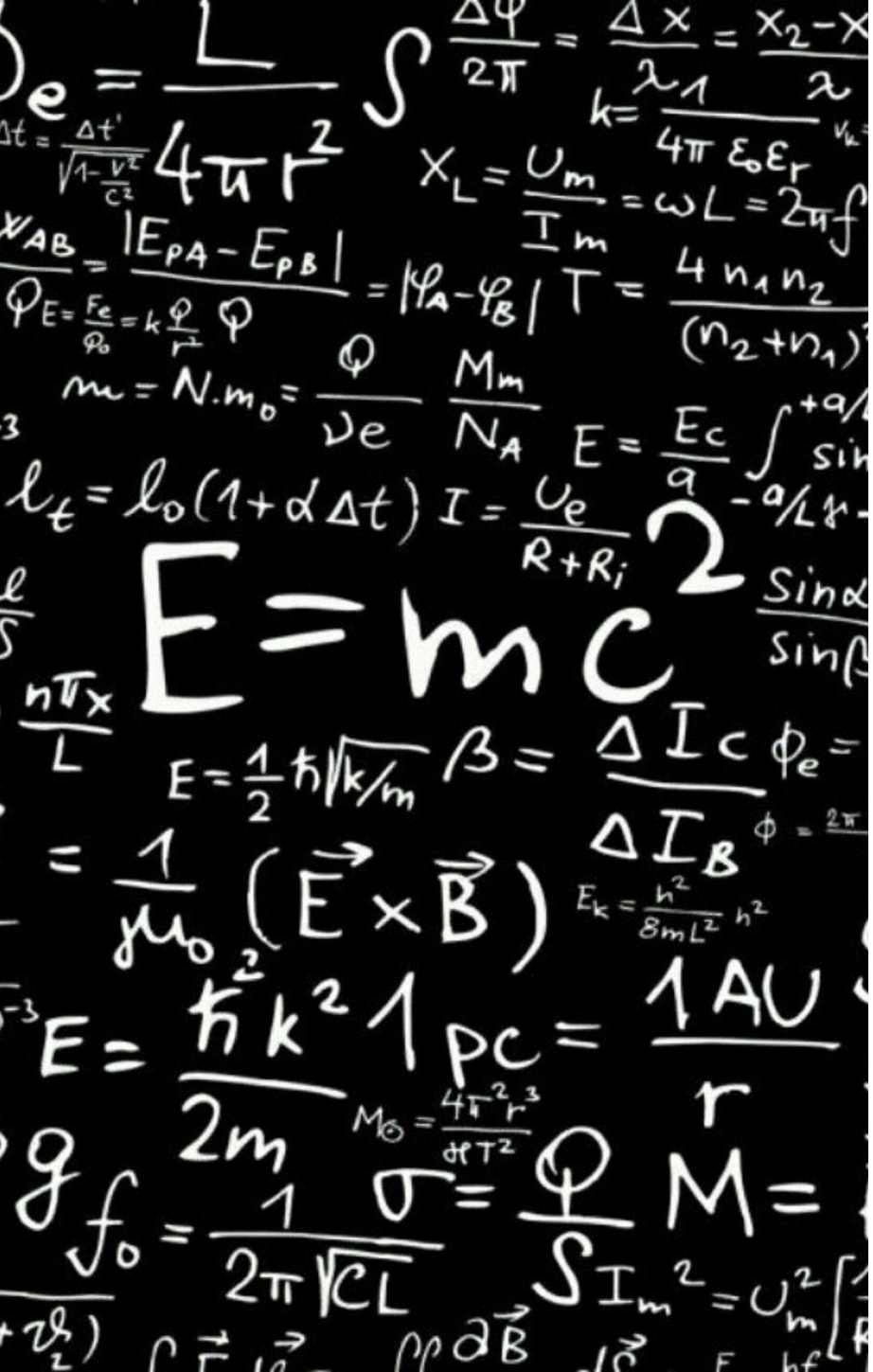
- Students understand the basic of machine learning and can design simple machine learning models for any given business problem.
- Students understand machine learning models and can code a simple model. Students understand how to tune the model to accomplish desired objectives.
- Students get exposed to more advanced machine learning techniques to prepare them for the future.
- Students are exposed to practical real-world examples of machine learning and can communicate with machine learning engineers using the same language.

Day 1-2

- Day 1 Morning : Introduction to Machine Learning
- Day 1 Afternoon : Data Preparation with Python (Refresh)
- Day 2 Morning : Linear and Logistic Regression
- Day 2 Afternoon : Decision Trees and Random Forest

Day 3-4

- Day 3 Morning : Nearest-Neighbor Methods, Feature Selection
- Day 3 Afternoon : Recommender System, Unsupervised Learning
- Day 4 Morning : Neural Network
- Day 4 Afternoon : Advanced Concepts in Machine Learning



INTRODUCTION TO MACHINE LEARNING

MACHINE LEARNING

THE SCIENCE OF GETTING COMPUTERS TO LEARN
FROM DATA WITHOUT HAVING
TO BE EXPLICITLY PROGRAMMED BY HUMANS.

THE MOST BASIC UNDERSTANDING

- It’s all about letting computer learns what ‘input’ is associated to what ‘output’.
  - Example: given a picture, computer outputs what object appears in the picture (human, car, tree?).
  - Example: given inputs from sensors and cameras, the robotic algorithm pushes out the appropriate movement.

EXAMPLES OF MACHINE LEARNING APPLICATIONS

SPAM CLASSIFICATION

- Email (text) as the input -> Go into classification model -> Output the answer whether this is spam or not.
- Big email platforms can identify spams with 99% accuracy.

machine learning - Google Search

Secure https://www.google.co.th/search?q=machine+learning&oq=machine+learning&aqs=chrome..69i57j69i60l2j69i61j0j...

Google

machine learning

All Images News Videos Books More Settings Tools

About 18,000,000 results (0.69 seconds)

Machine Learning | Coursera

https://www.coursera.org/learn/machine-learning

About this course: **Machine learning** is the science of getting computers to act without being explicitly programmed. In the past decade, **machine learning** has given us self-driving cars, practical speech recognition, effective web search, and a vastly improved understanding of the human genome.

Machine learning - Wikipedia

https://en.wikipedia.org/wiki/Machine_learning

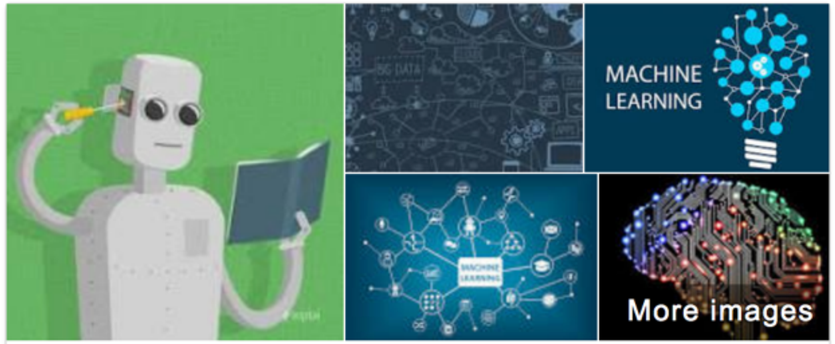
Machine learning is a field of computer science that gives computers the ability to learn without being explicitly programmed. Arthur Samuel, an American pioneer in the field of computer gaming and artificial intelligence, coined the term "**Machine Learning**" in 1959 while at IBM.

Top stories

Workday buys SkipFlag to bolster machine

Are Artificial Intelligence And Machine Learning

Unstructured content: An untapped fuel



Machine learning

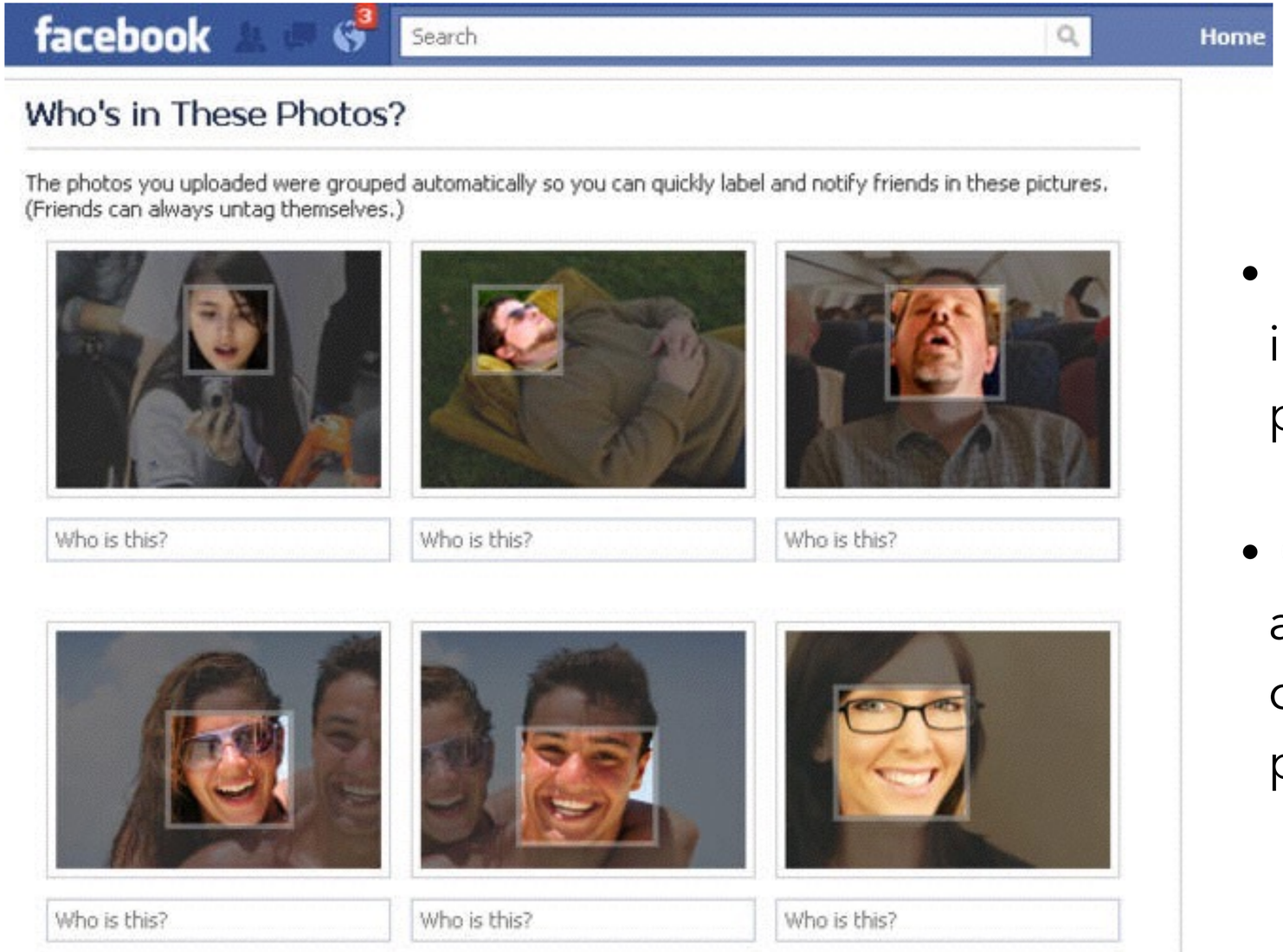
Field of study

Machine learning is a field of computer science that gives computers the ability to learn without being explicitly programmed. Wikipedia

Feedback

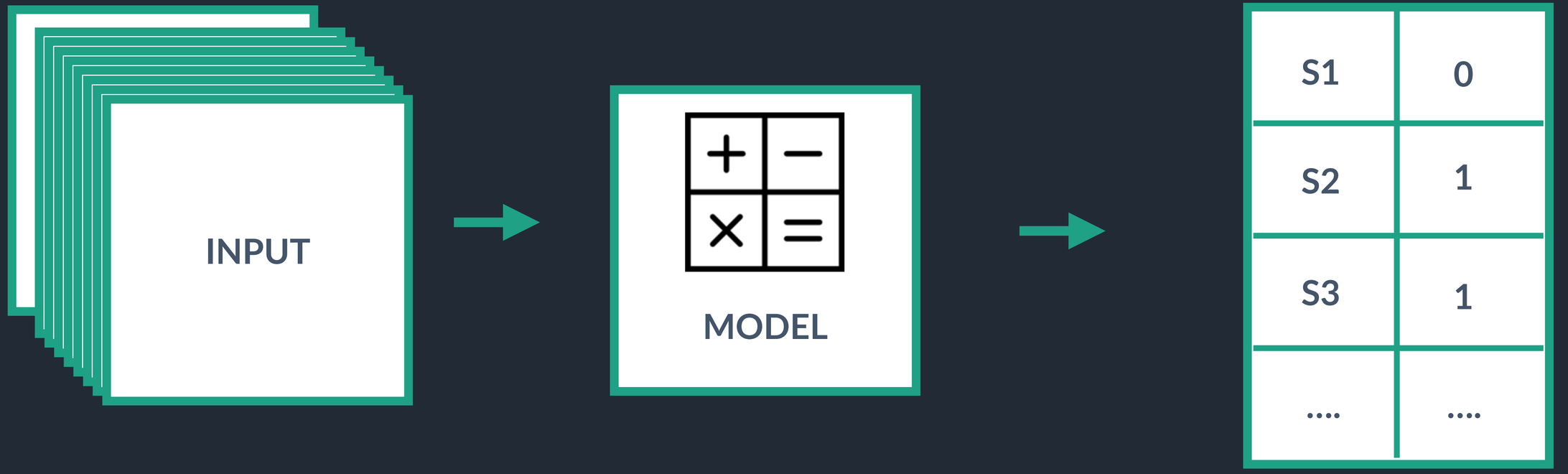
GOOGLE SEARCH ENGINE

FACEBOOK FACE TAGGING



- People provide Facebook the images and tags of names in the photos.
- Over time Facebook learned to associate names with faces and can automatically recognize these people.

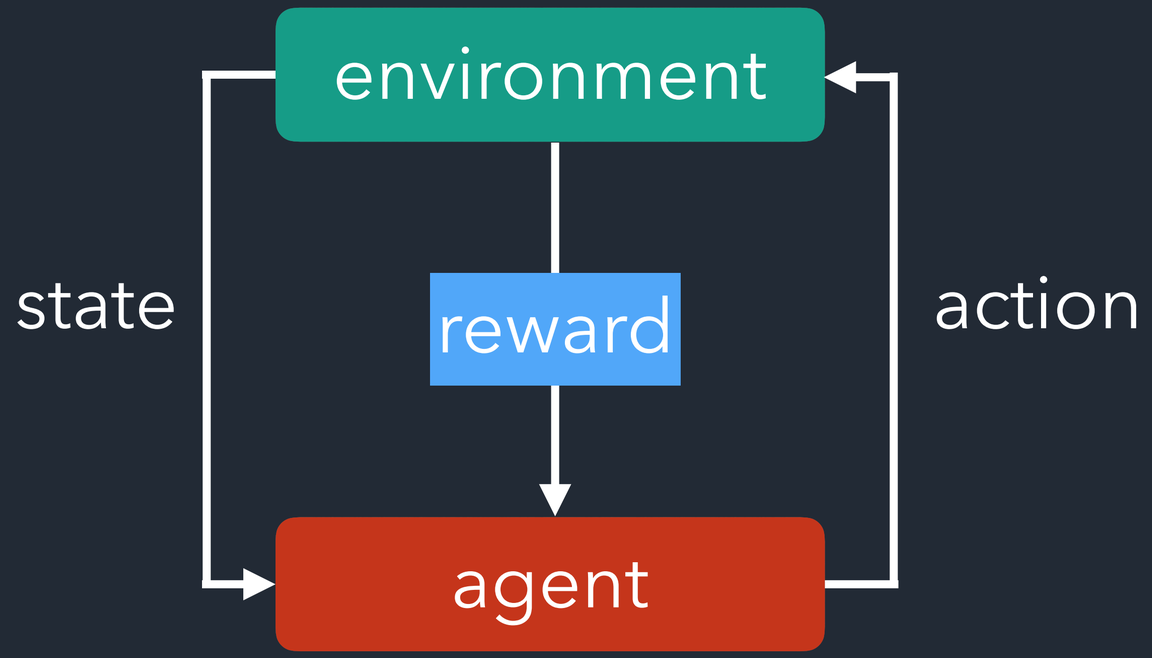
INPUT - MODEL - OUTPUT

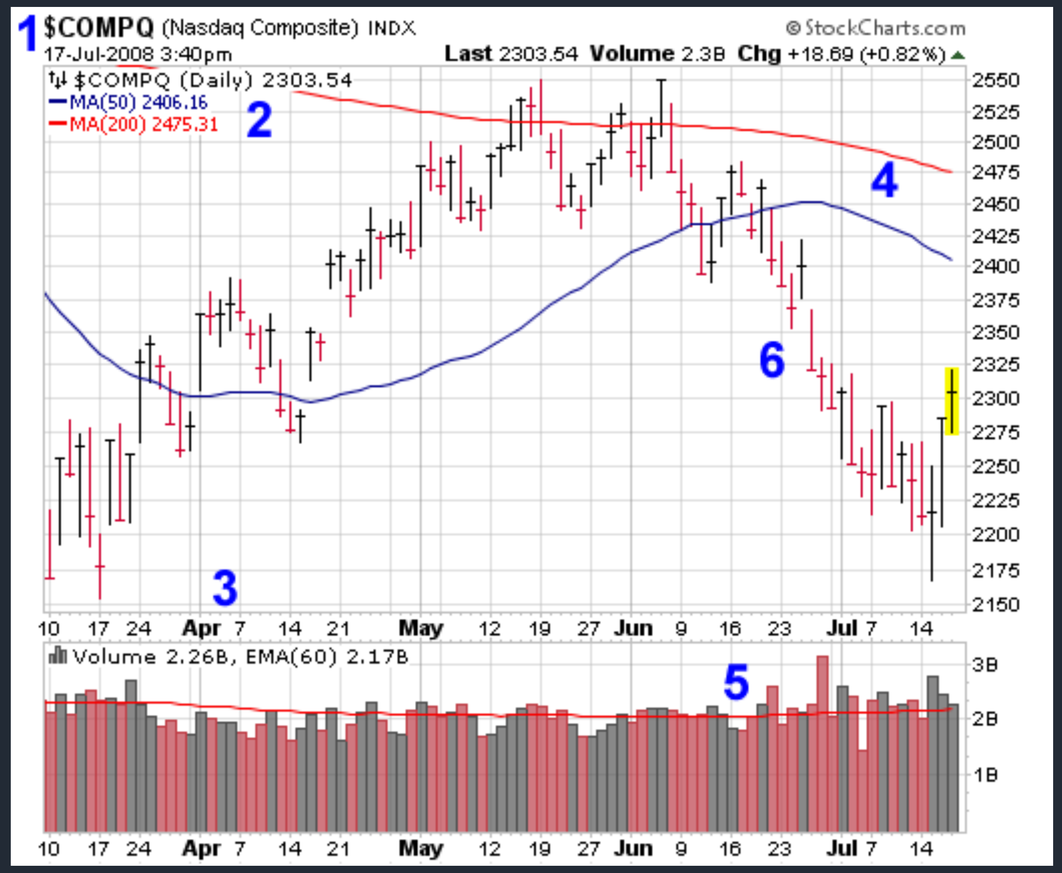


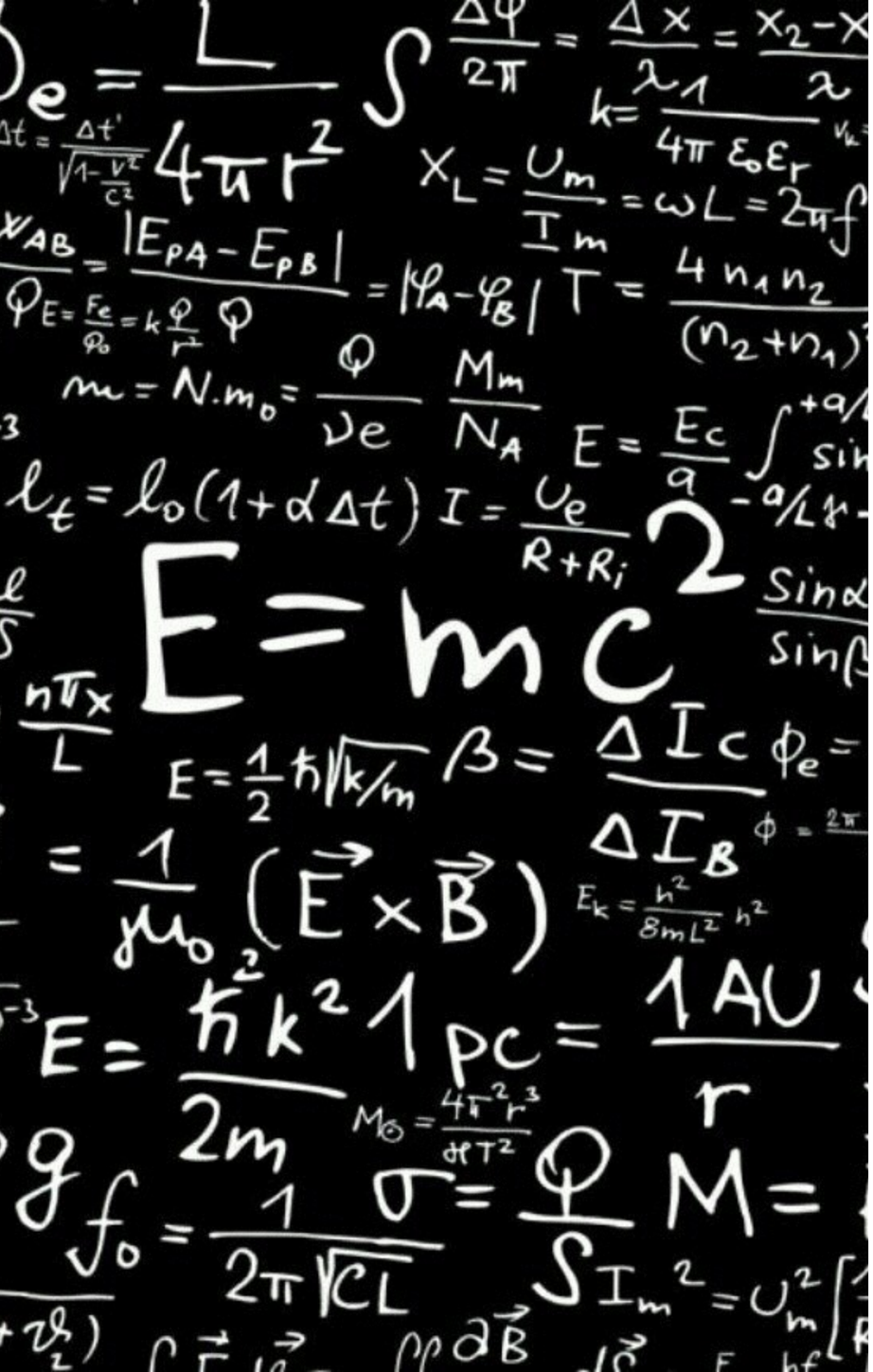
ANOTHER VIEW OF MACHINE LEARNING

TEACHING THE COMPUTER TO LEARN FROM EXPERIENCES AND OPTIMIZE A GIVEN PERFORMANCE INDEX AS THEY PRACTICE.

INTELLIGENT SYSTEM WITH MACHINE LEARNING



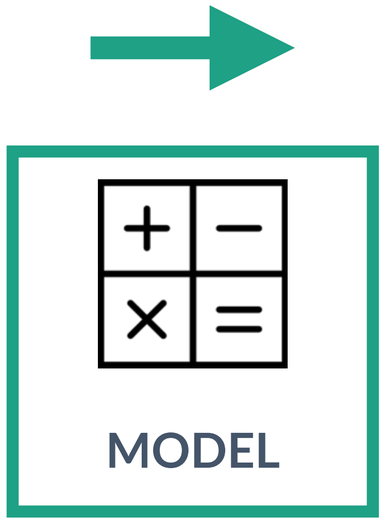




CLASSIFICATION AND REGRESSION

REGRESSION PROBLEM

- Property size
- Property age
- Bedrooms
- Bathrooms
- Parking size

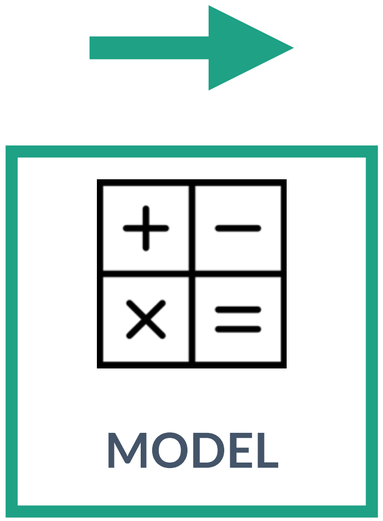


How much should we sell the property? (the answers range from 0 to 1B)

Regression problems are the type of problems where model's answers are continuous numbers.

CLASSIFICATION PROBLEM

- Property size
- Property age
- Bedrooms
- Bathrooms
- Parking size

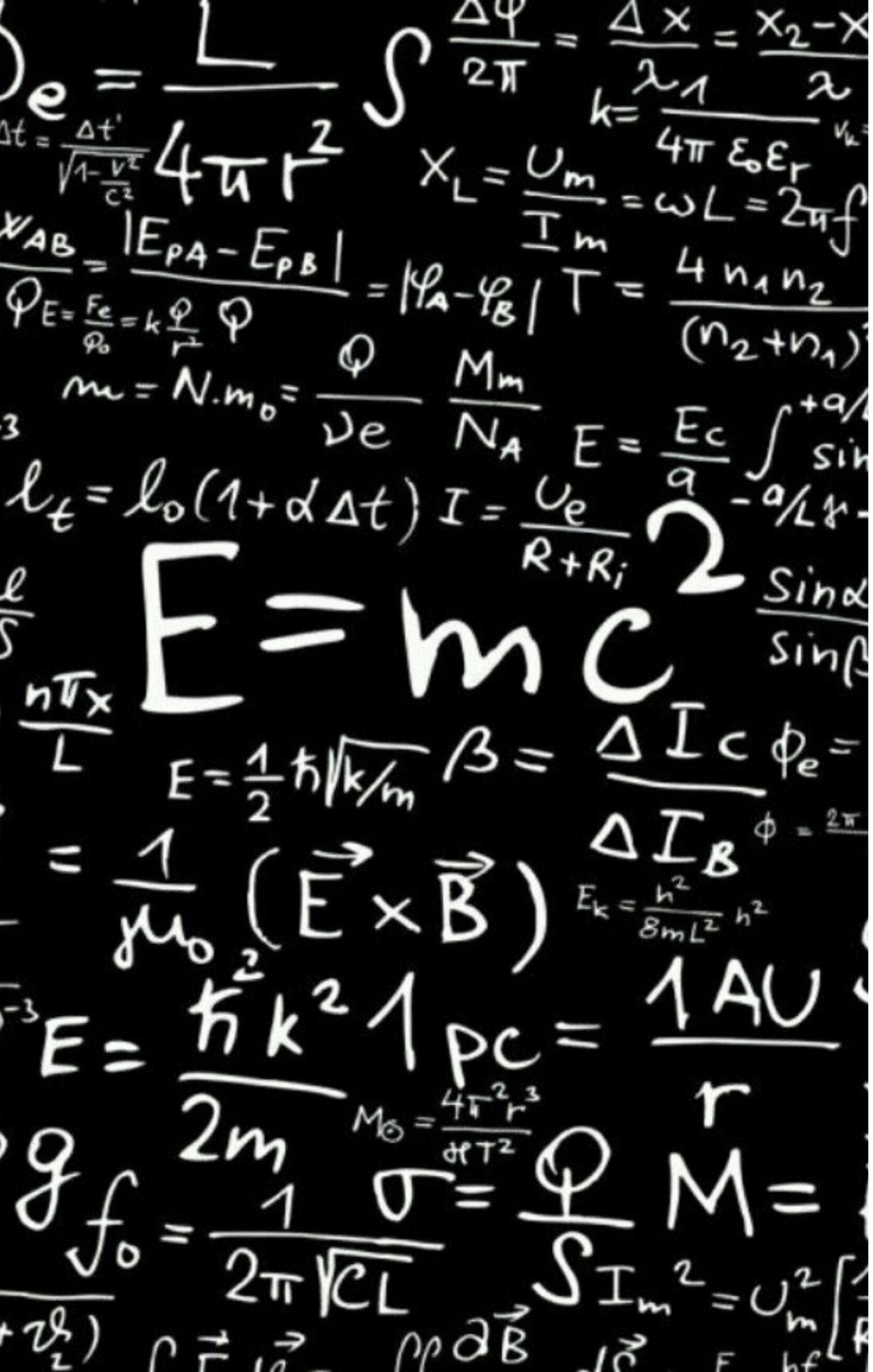


Tell me, what type of property is this? (residential or commercial)

Classification problems are the type of problems where model's answers are discrete categories.

SUPERVISED V.S. UNSUPERVISED LEARNING

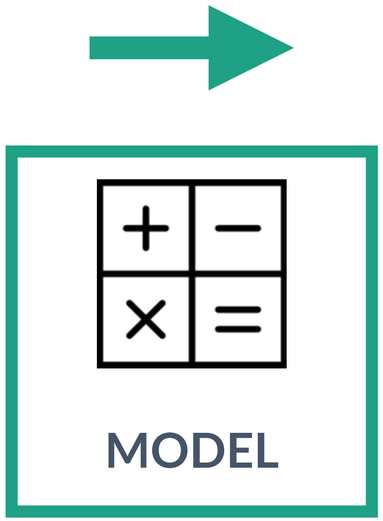
- **Regression Problem**
  - The answers models come up with are continuous numbers.
- **Classification Problem**
  - The answers models come up with are discrete categories.
- Note that you can apply both approaches to the same dataset!



SUPERVISED AND UNSUPERVISED LEARNING

SUPERVISED LEARNING

- Property size
- Property age
- Bedrooms
- Bathrooms
- Parking size



REGRESSION

How much should we sell the property? (the answers range from 0 to 1B)

CLASSIFICATION

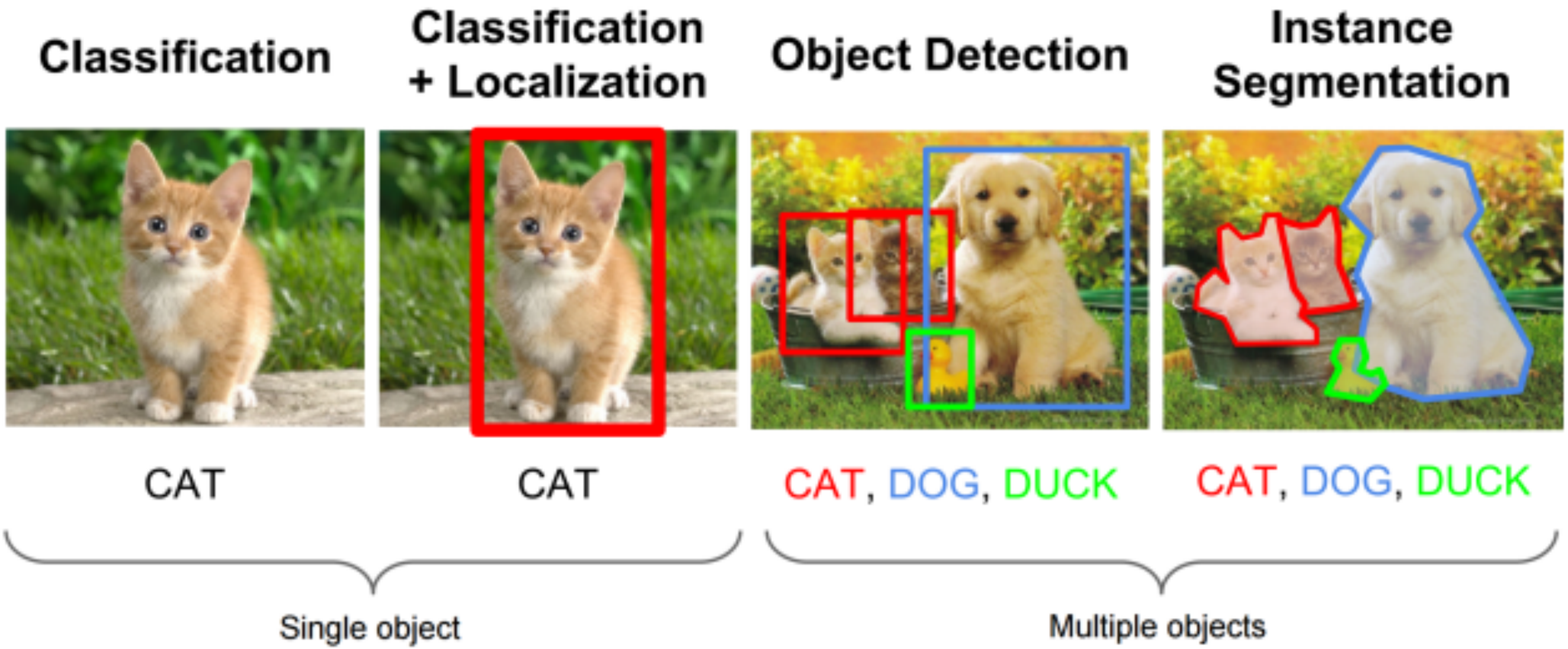
Tell me, what type of property is this? (residential or commercial)

SUPERVISED LEARNING

- We collect a lot of data points from the past,
  e.g. Collecting property qualities, property prices, and property types.
- We use the past data to fit the model
  e.g. Teaching it to understand what quality map with what prices and what types.
- When the model encounters new samples where the answers are not available, it will use knowledge from past data to provide answers.

Supervised learning problems are those problems where the answers that the model predict are already included in the training data.

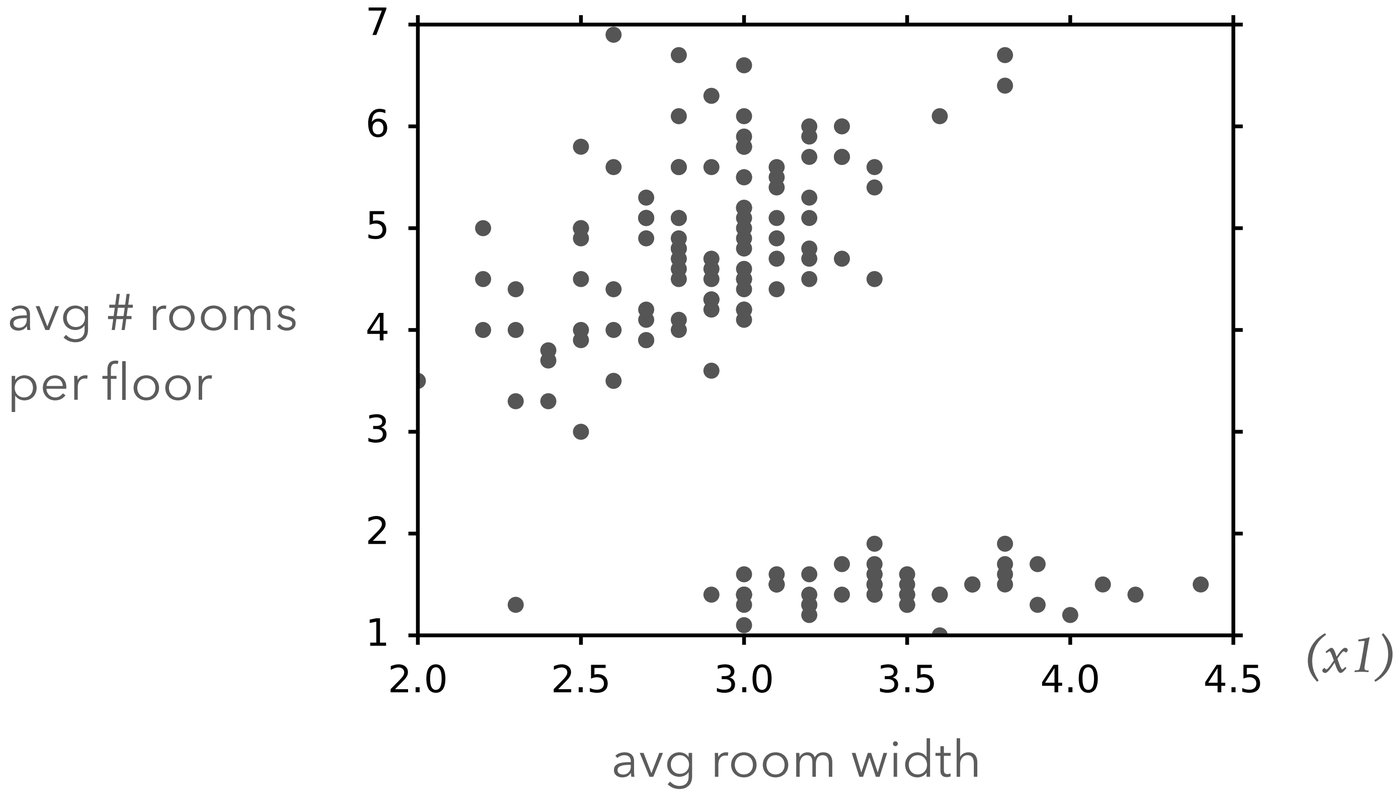
SUPERVISED LEARNING



SUPERVISED LEARNING

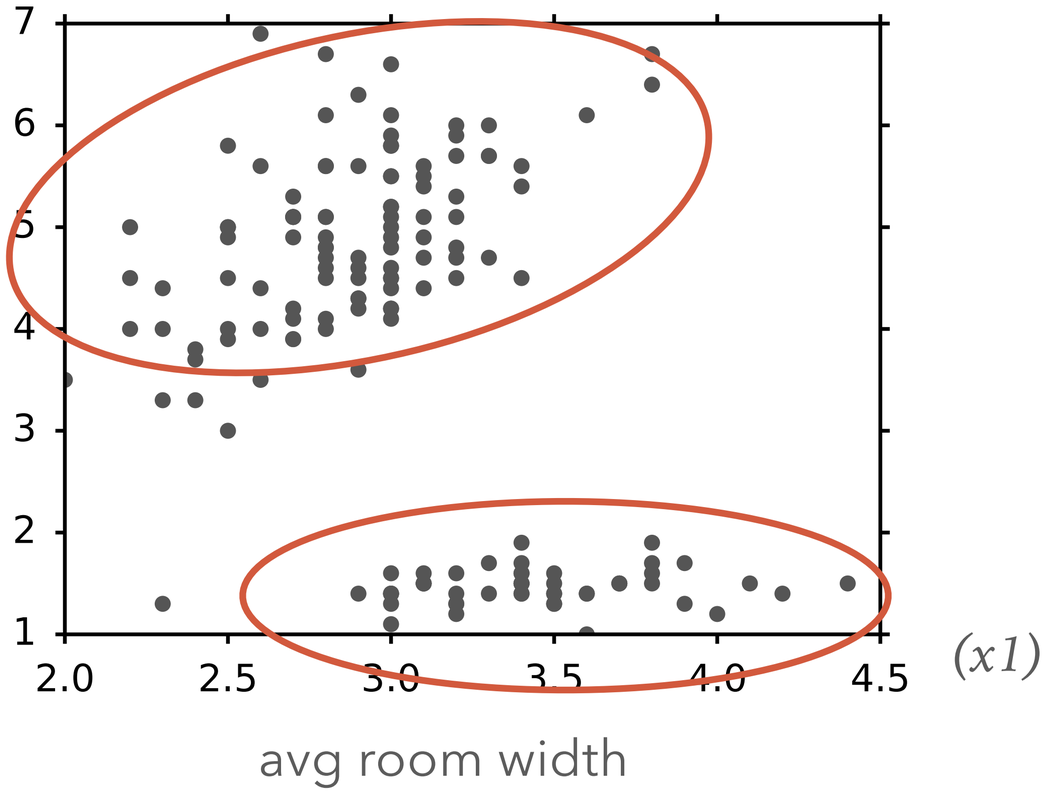
- Requires a lot of manually-labeled data.
- Requires business people to help decide “what to predict” (sometimes it’s hard to know what is the most useful thing to predict).
- Requires business people to identify and gather “appropriate inputs”.
- If done right, they are the most simple and reliable techniques to use. They are the core of most AI systems we see today.

UNSUPERVISED LEARNING



- What if you want to predict property categories, but don't know the answers in advance?
- You have to infer categories from the structure of the data.
- You are going to use unsupervised learning in this case.

UNSUPERVISED LEARNING

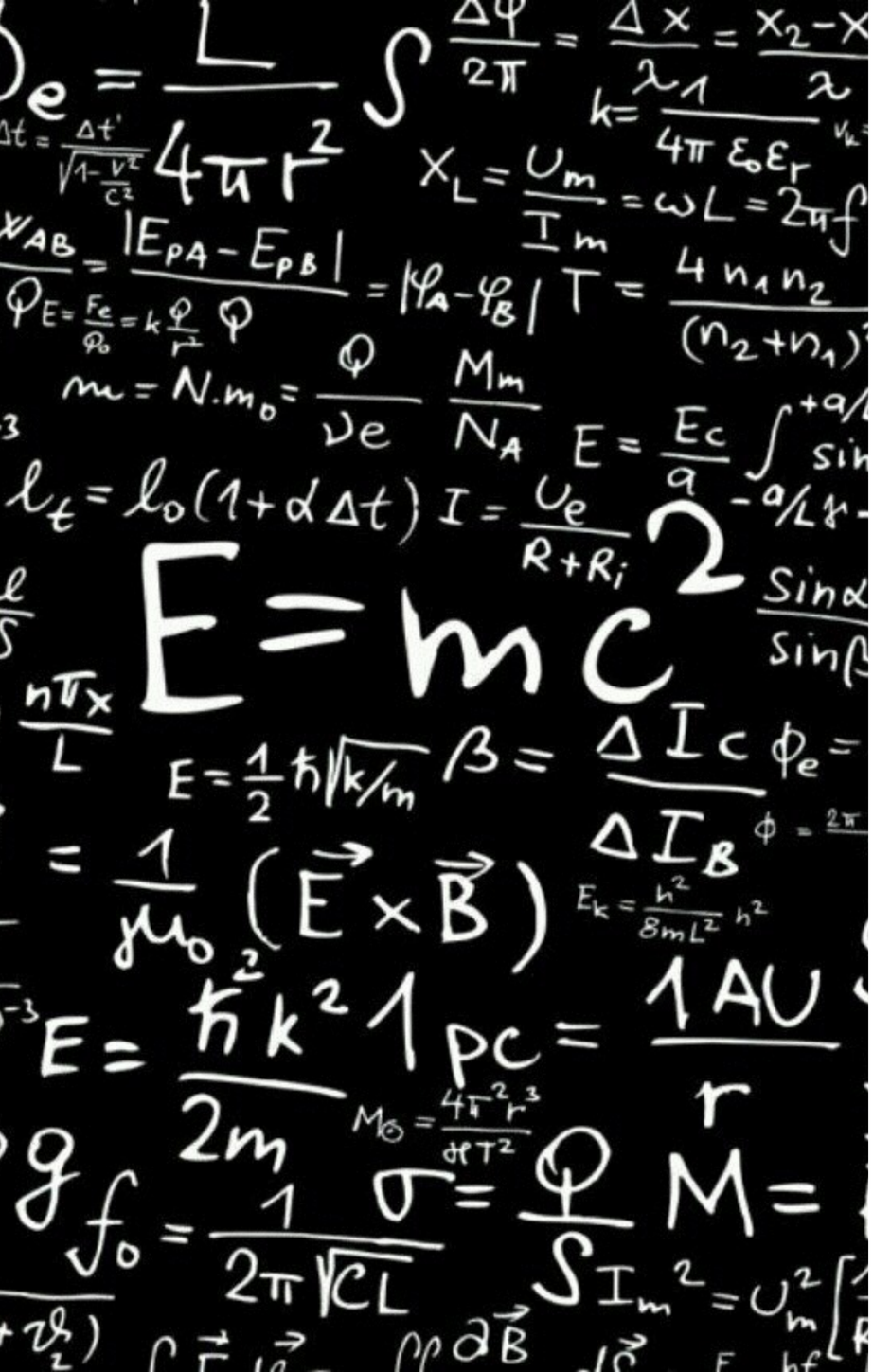


- What if you want to predict property categories, but don't know the answers in advance?
- You have to infer categories from the structure of the data.
- You are going to use unsupervised learning in this case.

Unsupervised learning problems are those problems where the answers that the model predict are not available in the training set. We infer categories from data structure.

SUPERVISED V.S. UNSUPERVISED LEARNING

- **Supervised learning**
  - The answers are included in the training data, note that answers can be numerical or categorical.
- **Unsupervised learning**
  - You would like to discover the categories, you usually don’t even know how many categories or what categories are there.
- Note that you can apply both approaches to the same dataset!



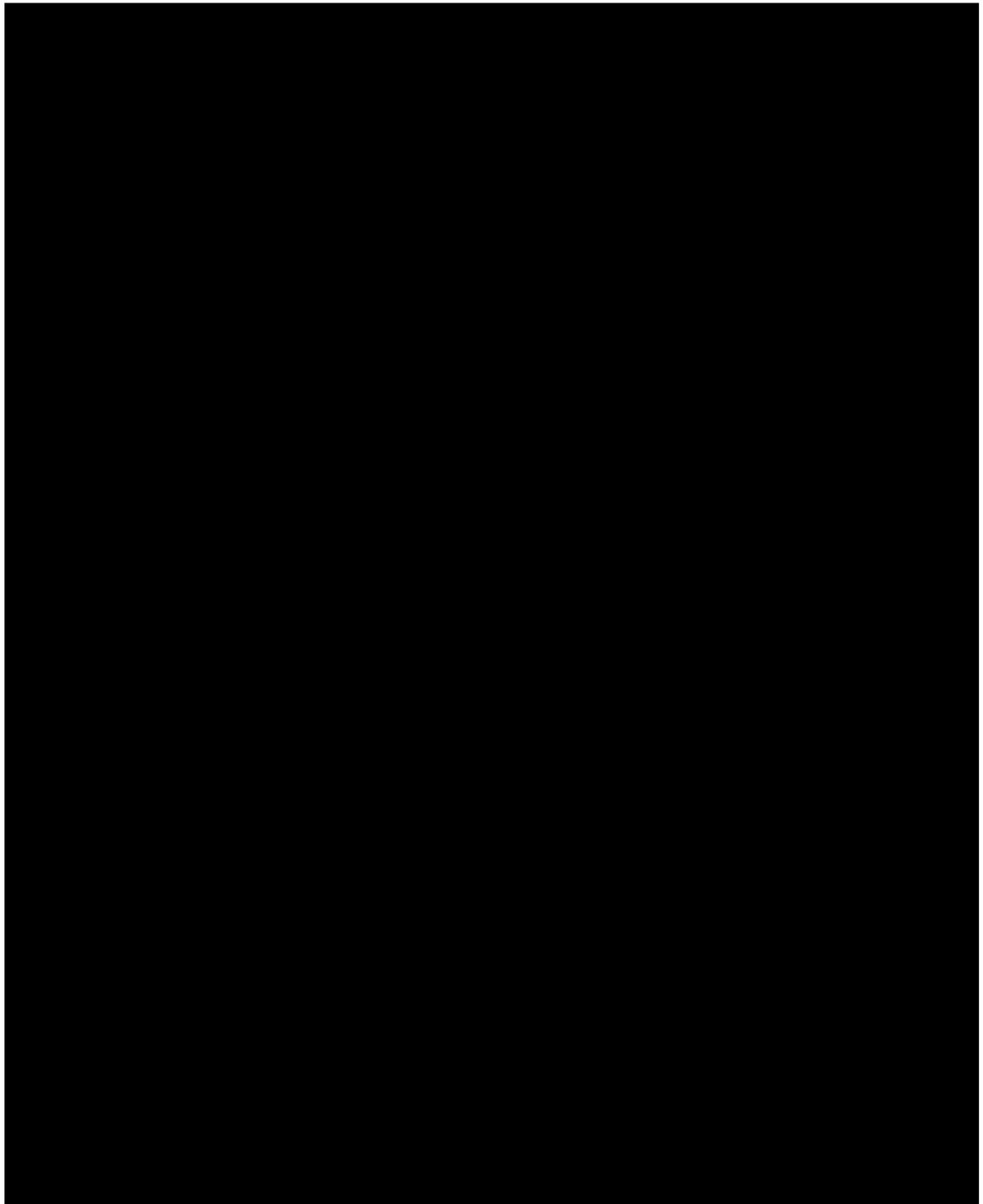
REINFORCEMENT LEARNING

REINFORCEMENT LEARNING

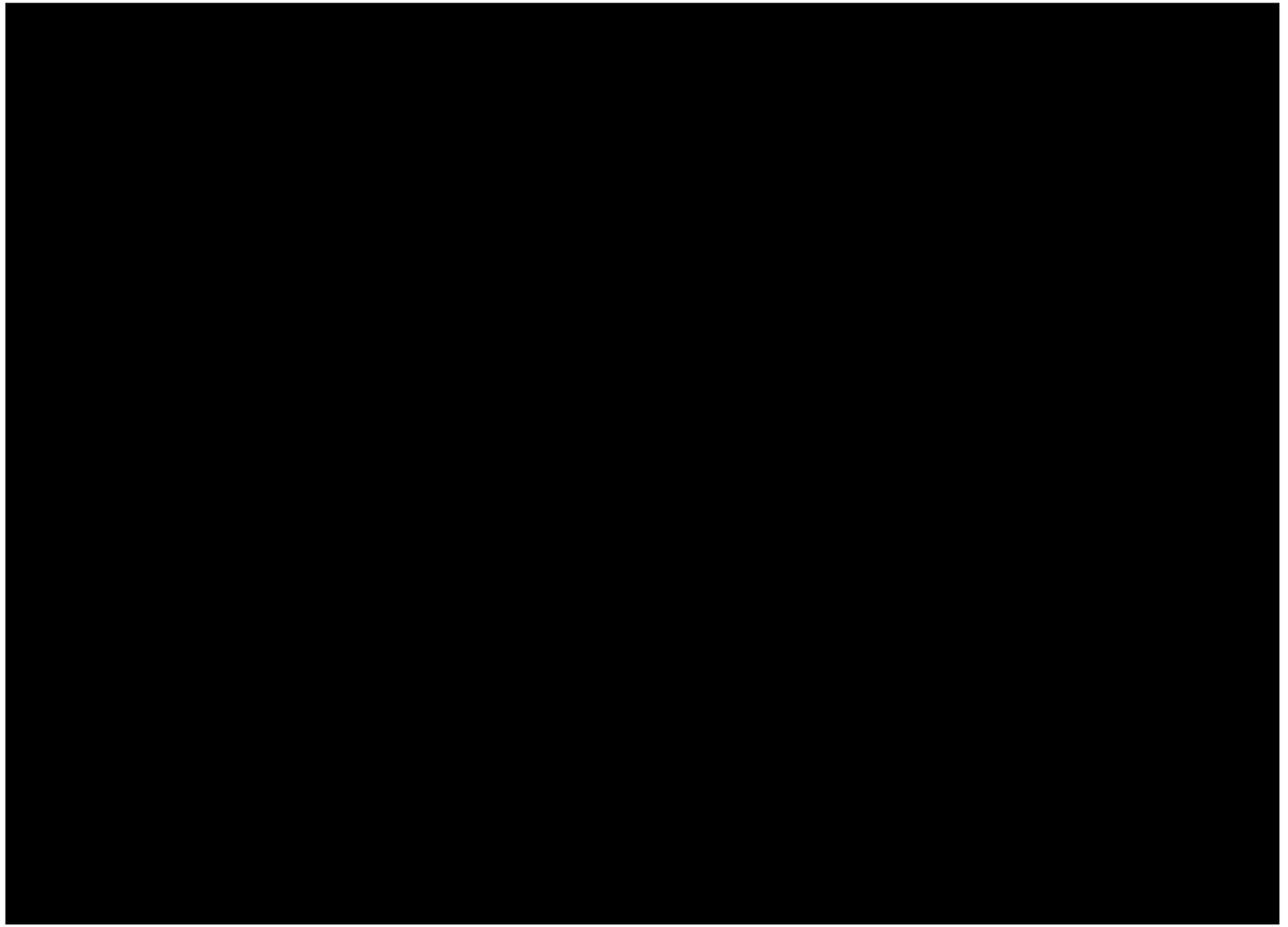
- Environment defines a set of states, actions, and rewards. Models is trained to understand what actions to take, at what states, to optimize rewards.
- Example: given stock prices (states) and let bots decide each day to buy, sell, or hold a particular stock (actions), bots will make decisions to optimize rewards (profit).

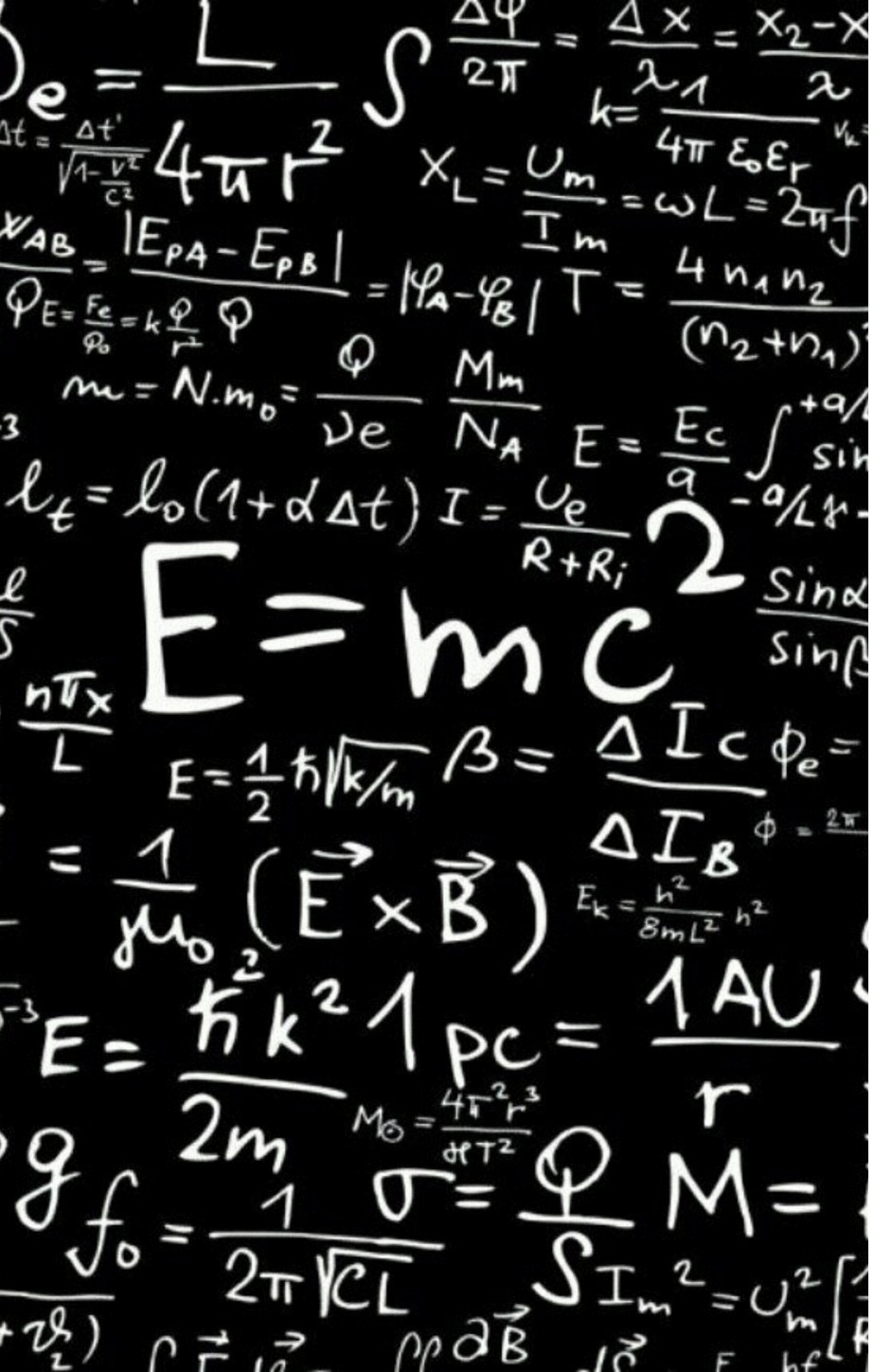
ATARI GAME

ATARI GAME



RUNNING BOT





MACHINE LEARNING BASICS I

HOUSE PRICE PREDICTION EXAMPLE

THE MOST BASIC EXAMPLE

An agent has been selling 300 houses in the last years and want to be able to predict the price of a house by just knowing the size of the house.

i	Size (m^2)	Price (Mbaht)
1	50	1.4
2	128	2.6
3	24	0.8
4	78	1.2
i	…	…

THE MOST BASIC EXAMPLE

In general,

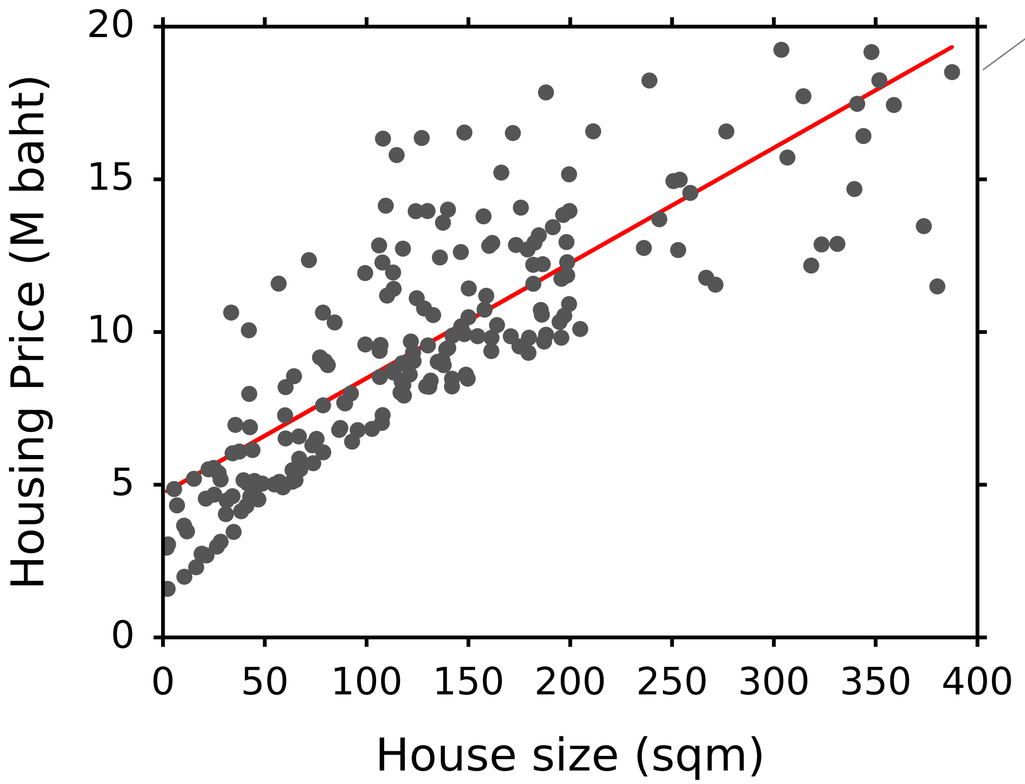
x: feature (input)

y: target (output)

i: sample index

	x		y	
i	Size (m^2)		Price (Mbaht)	
1	50	$= x_1$	1.4	$= y_1$
2	128	$= x_2$	2.6	$= y_2$
3	24	$= x_3$	0.8	$= y_3$
4	78	$= x_4$	1.2	$= y_4$
i	...	$= x_i$	...	$= y_i$

WHAT IS A MODEL



$$h(x) = \theta_0 + \theta_1 x$$

parameters

of the model

$$\theta_0 = 4.717$$

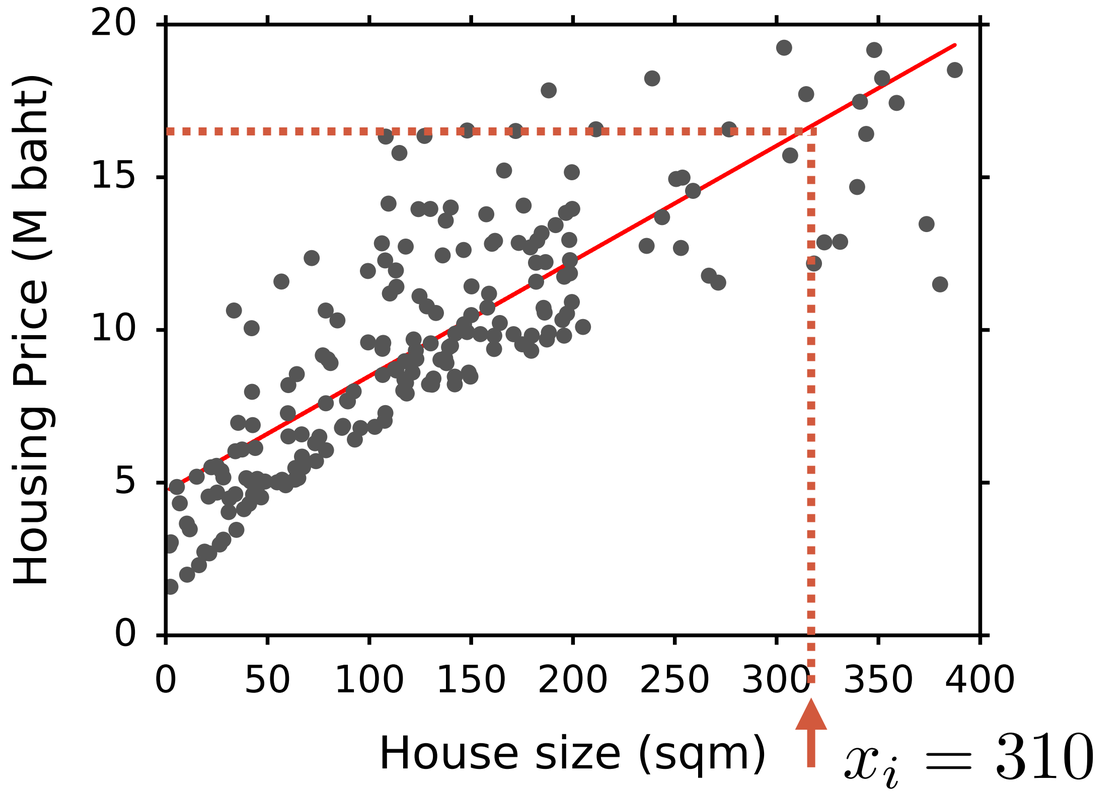
$$\theta_1 = 0.038$$

A model is a function that takes an input and yields the values we want to predict.

WHAT IS A MODEL

After we have a model, we can predict y value from any x value

Model: $h(x_i) = 4.717 + 0.038 * (x_i)$



Notice that:

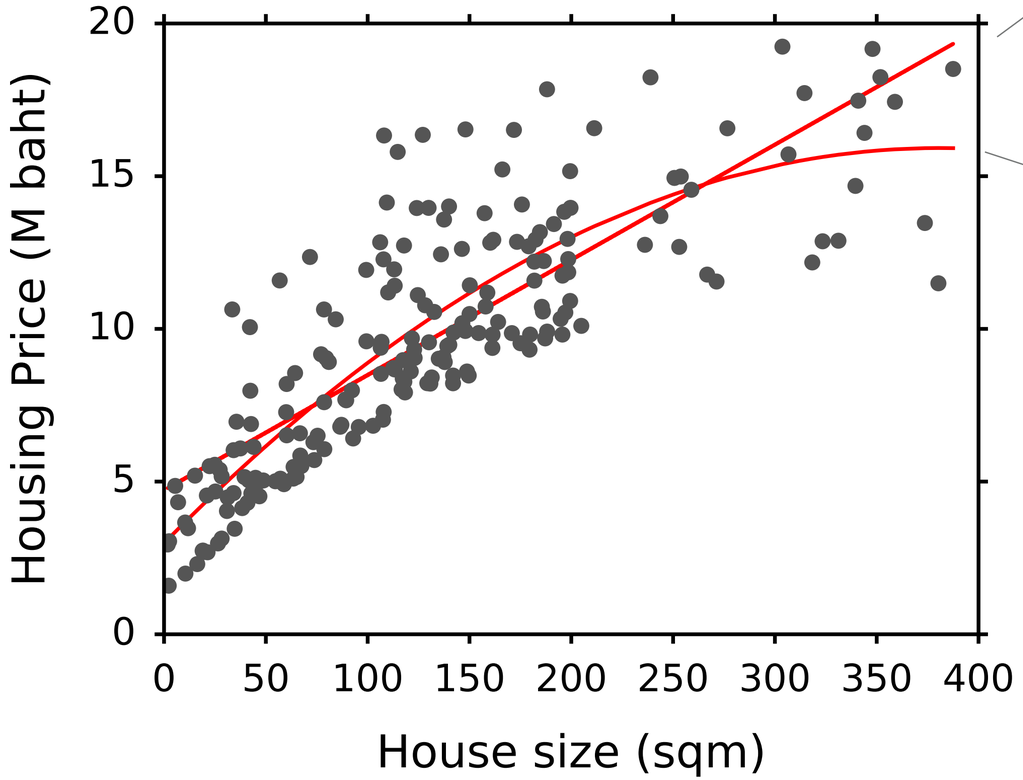
h(x) and y are not the same

h(x): value we predict

y: the actual value

WHAT IS A MODEL

For different models you have different functions, with perhaps different number of parameters. Different models can be fitted to the same data set.



$$h(x) = \theta_0 + \theta_1 x$$

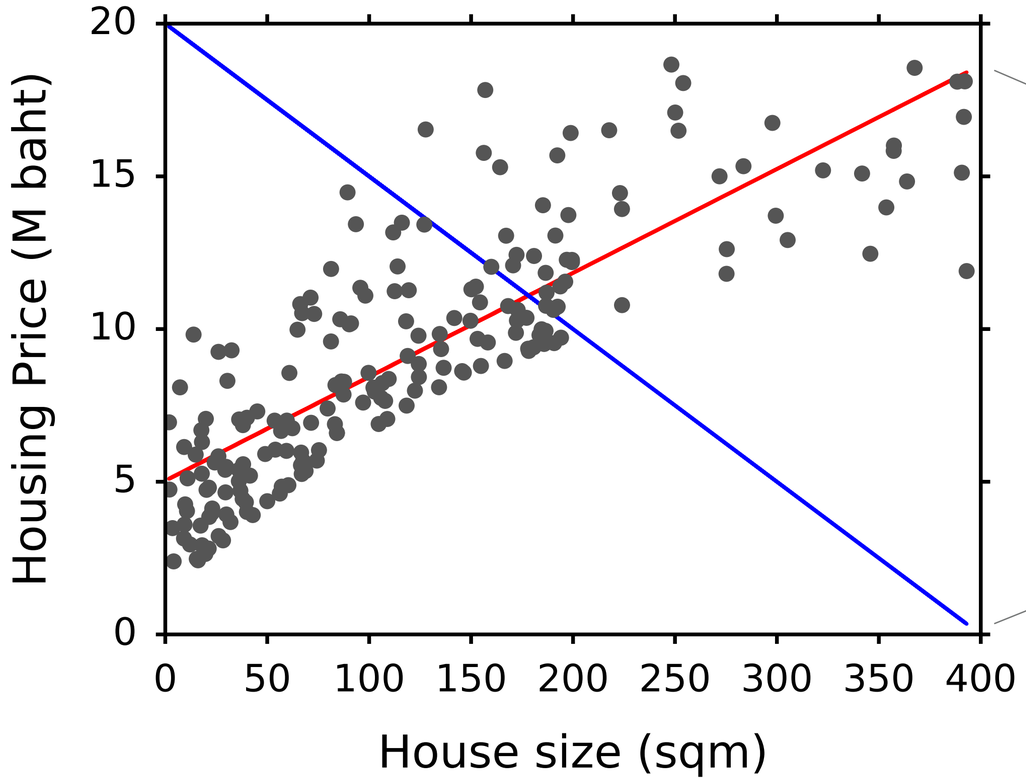
$$h(x) = \theta_0 + \theta_1 x + \theta_2 x^2$$

Model is a function of both features (x) and parameter (theta).

$$h(x; \theta)$$

GOOD AND BAD MODELS

Model: $h(x) = \theta_0 + \theta_1 x$



$\theta_0 = 4.717$

$\theta_1 = 0.038$

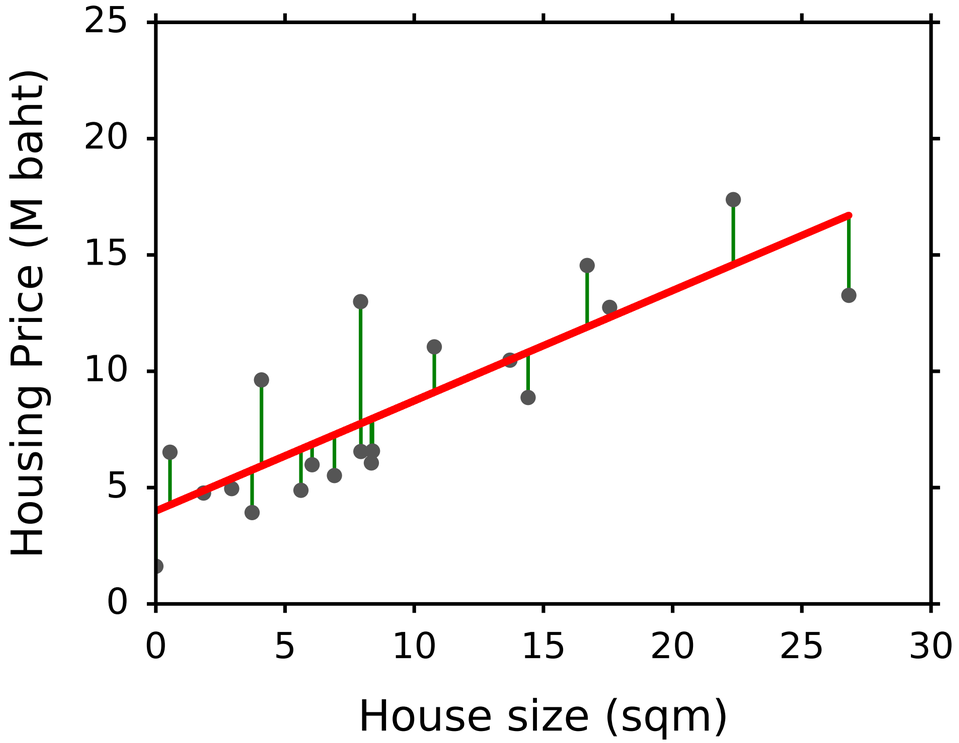
$\theta_0 = 20$

$\theta_1 = -0.05$

COST FUNCTION

Cost function (or loss function) is a measure of whether a model is a good fit to the data.

For example, a famous cost function is called ‘squared error’ function that takes the difference between what you predict and the actual data value and square it.



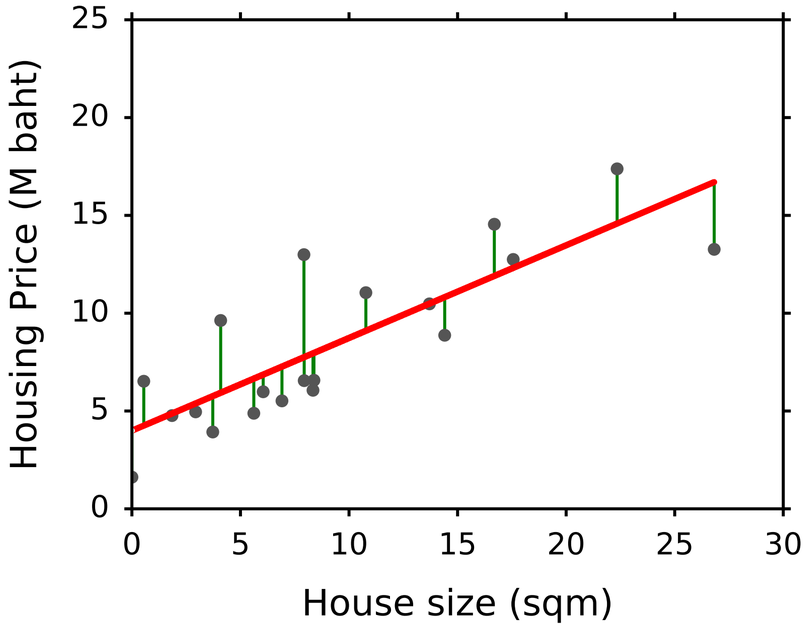
$$\sum_i (h(x_i) - y_i)^2$$

LOWERING COST FUNCTION

Model: $h(x) = \theta_0 + \theta_1 x$

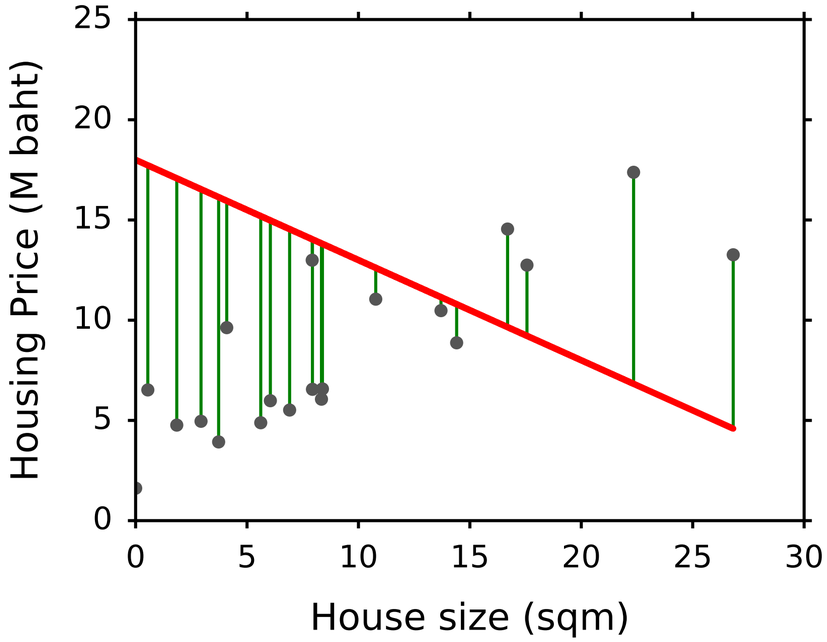
$\theta_0 = 4.717$

$\theta_1 = 0.038$



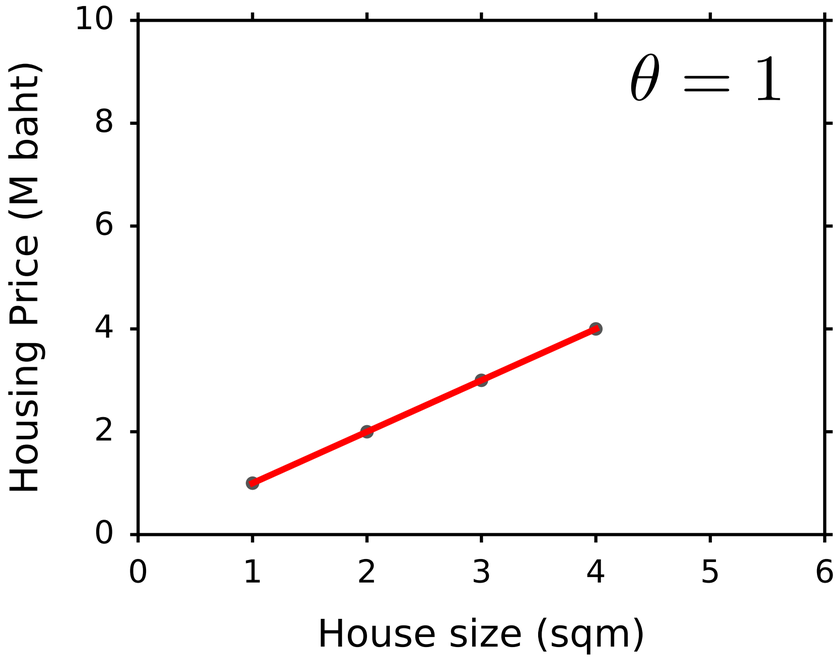
$\theta_0 = 20$

$\theta_1 = -0.05$

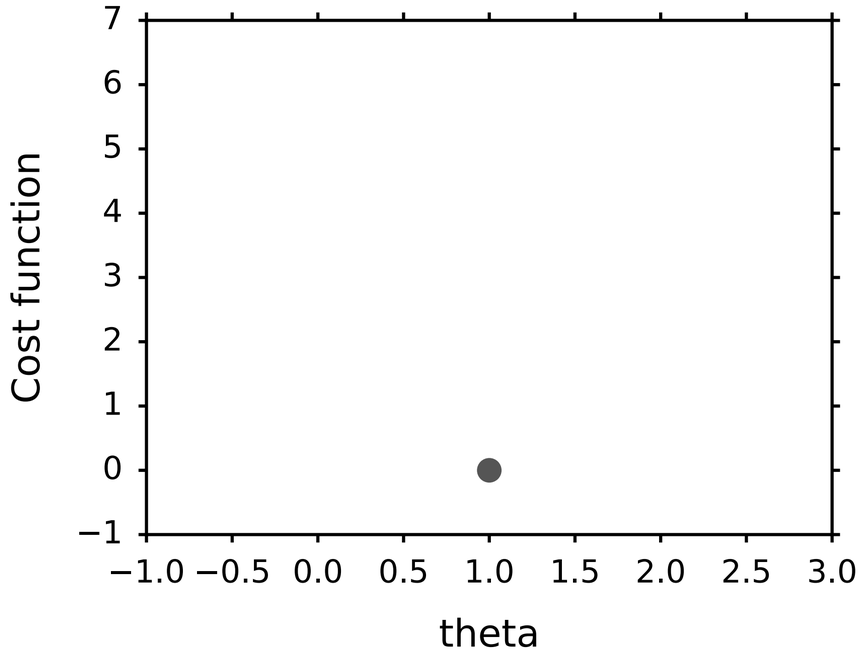


MINIMIZING COST FUNCTION

Model: $h(x) = \theta x$

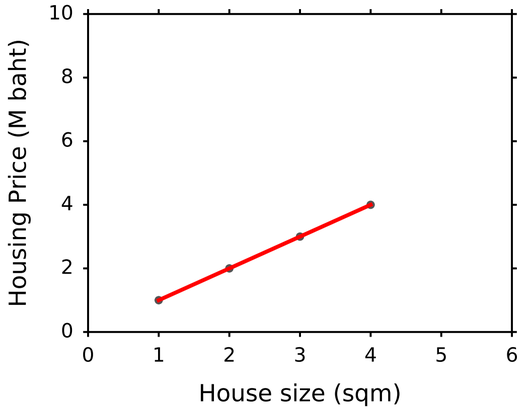


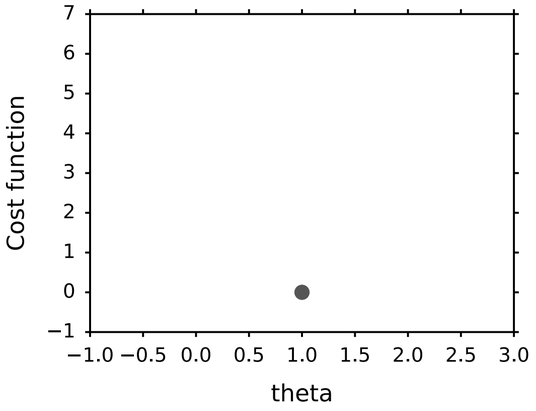
$Cost(\theta)$



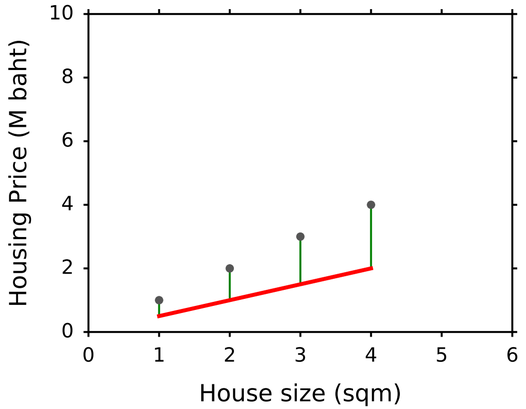
MINIMIZING COST FUNCTION

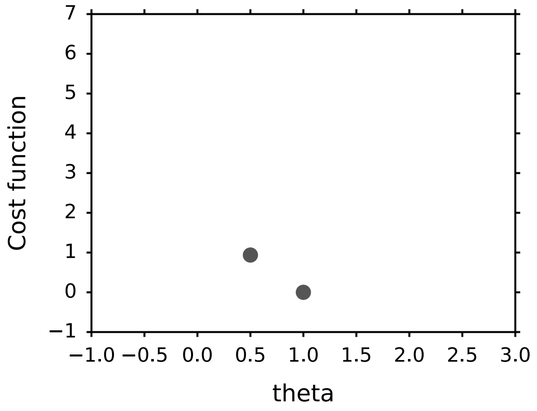
$\theta = 1$



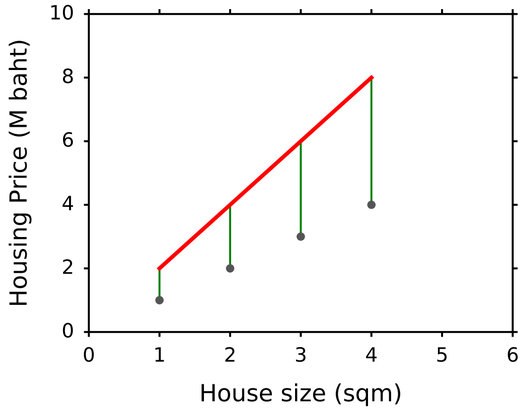


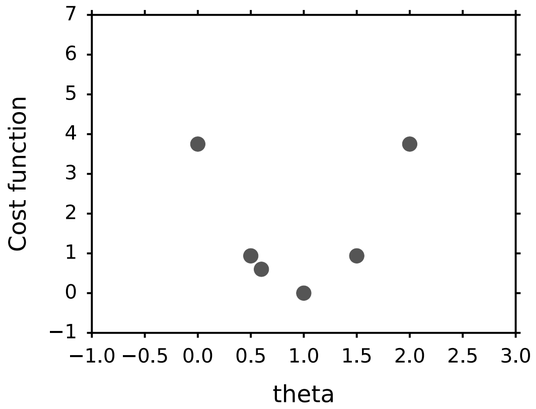
$\theta = 0.5$



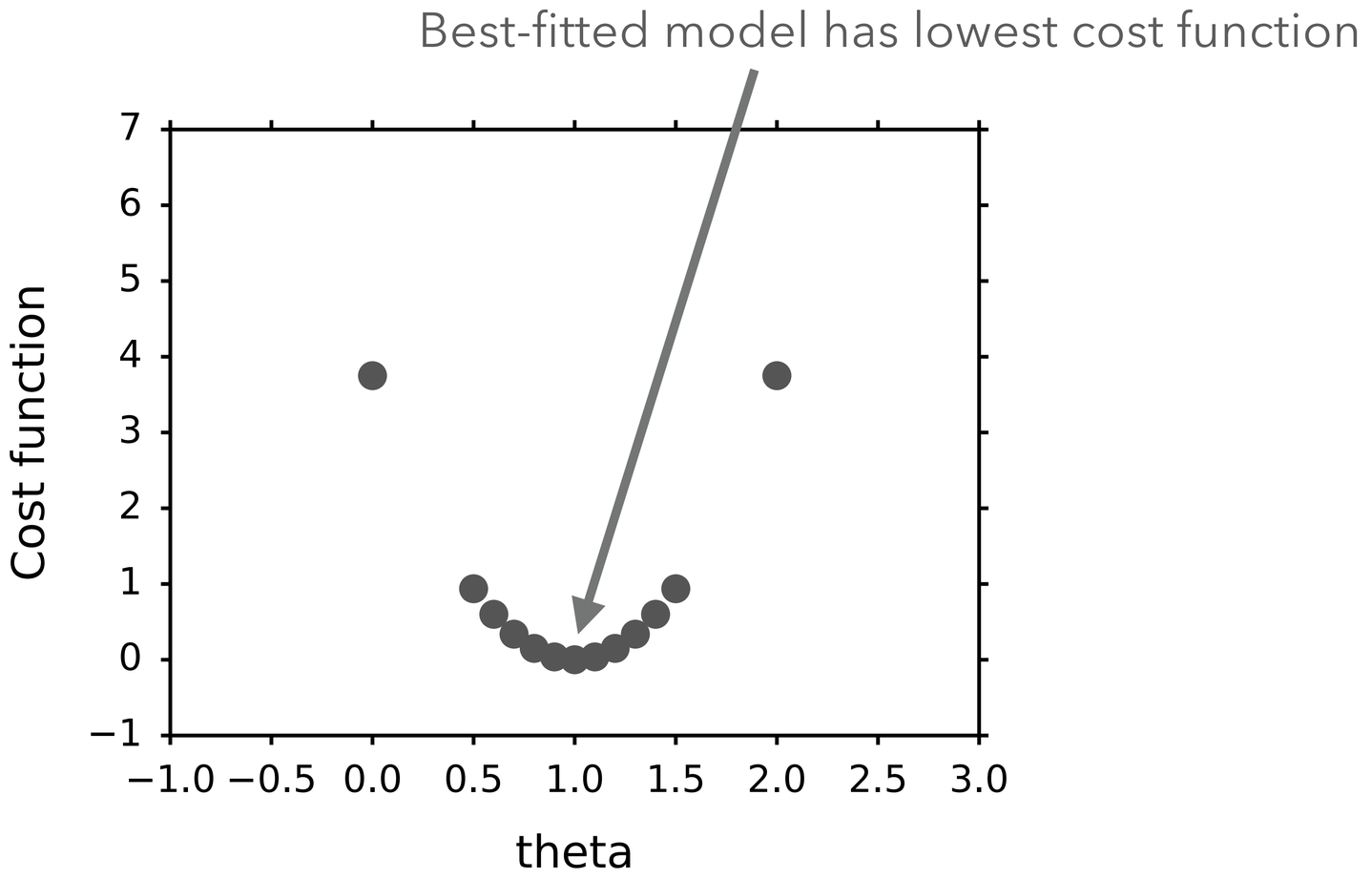


$\theta = 2$





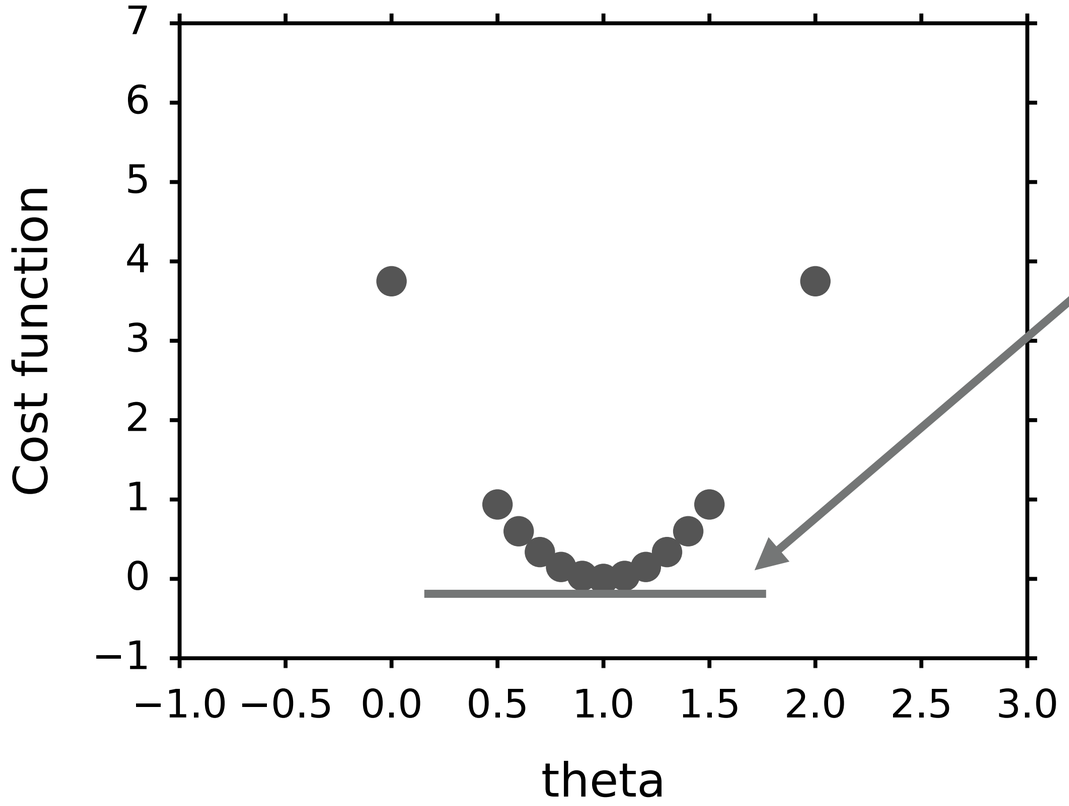
MINIMIZING COST FUNCTION



Training machine learning models = minimizing cost function

Often accomplished by using ‘optimization algorithms’

MINIMIZING COST FUNCTION



$$\frac{\partial Cost}{\partial \theta} = 0$$

Some optimization algorithm is so simple,

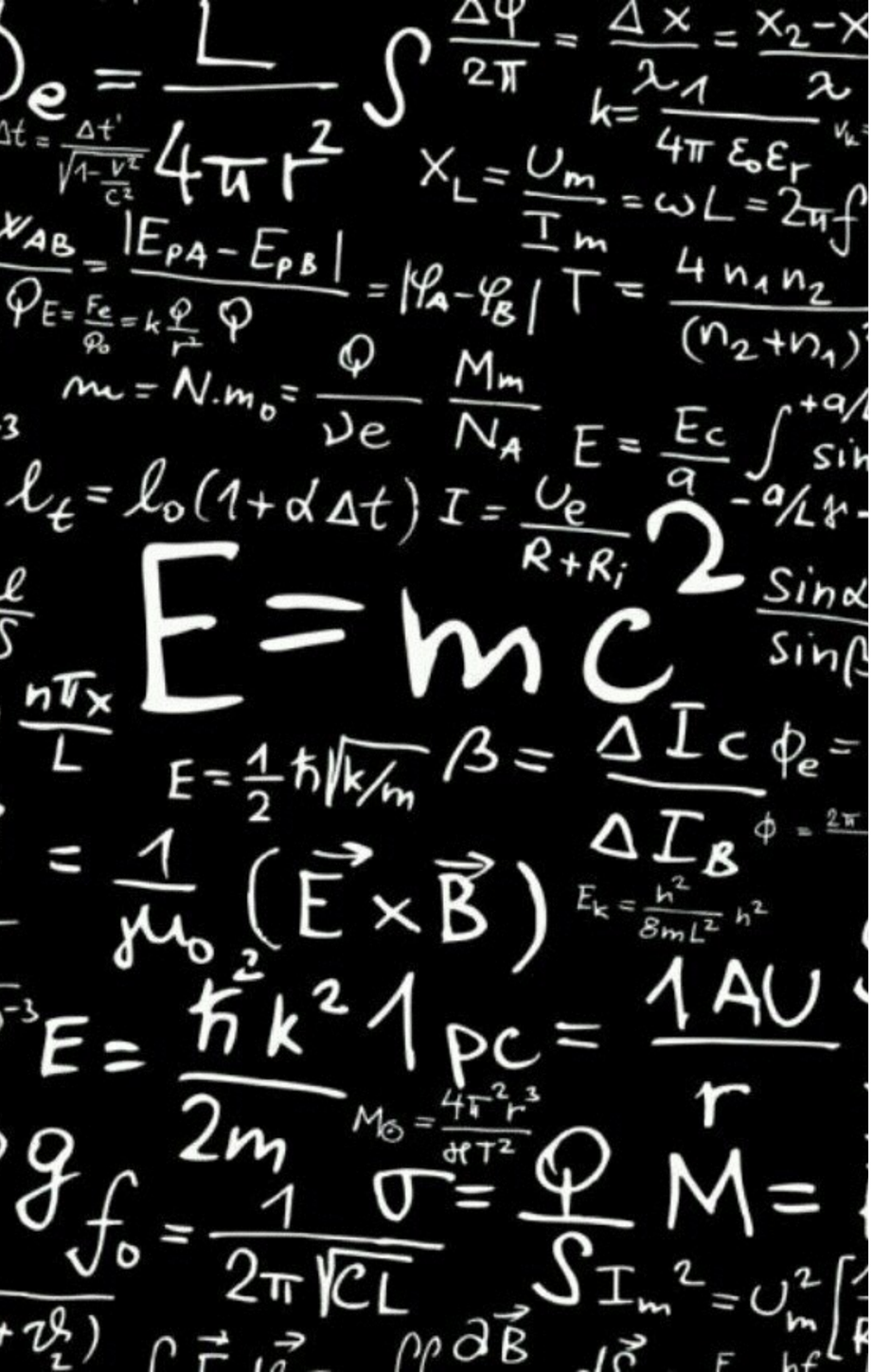
you just solve an equation and done.

MACHINE LEARNING BASICS

- To train the model, we collect **samples**, which are data points that the model need to learn.
- In sample data, we have **targets** (the quantity we want to predict) and **features** (the data we use to predict targets).
- A **model** is a set of equations that map **features** to **predictions**.
- All models have **parameters** which we can adjust to fit the dataset.

MACHINE LEARNING BASICS

- Model's **predictions** (often called h(x)) are not the same as **targets** (y).
- **Cost function** is an equation that measures the difference between h(x) and y.
- To fit the model to the data, we adjust **parameters** in a way that **minimize cost function.**



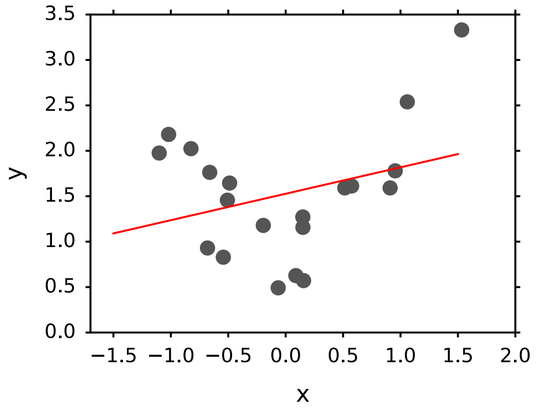
OVERFITTING

OVERFITTING

In ML, we strive to minimize cost function, but sometimes, we have the opposite problem. Cost function is too low!

underfit

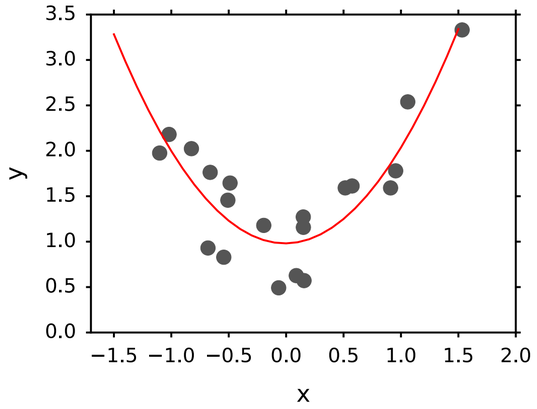
cost = 0.21



$$h(x) = \theta_0 + \theta_1 x$$

just right

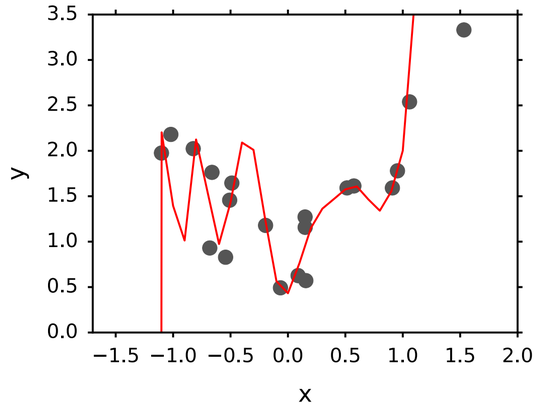
cost = 0.05



$$h(x) = \theta_0 + \theta_1 x + \theta_2 x^2$$

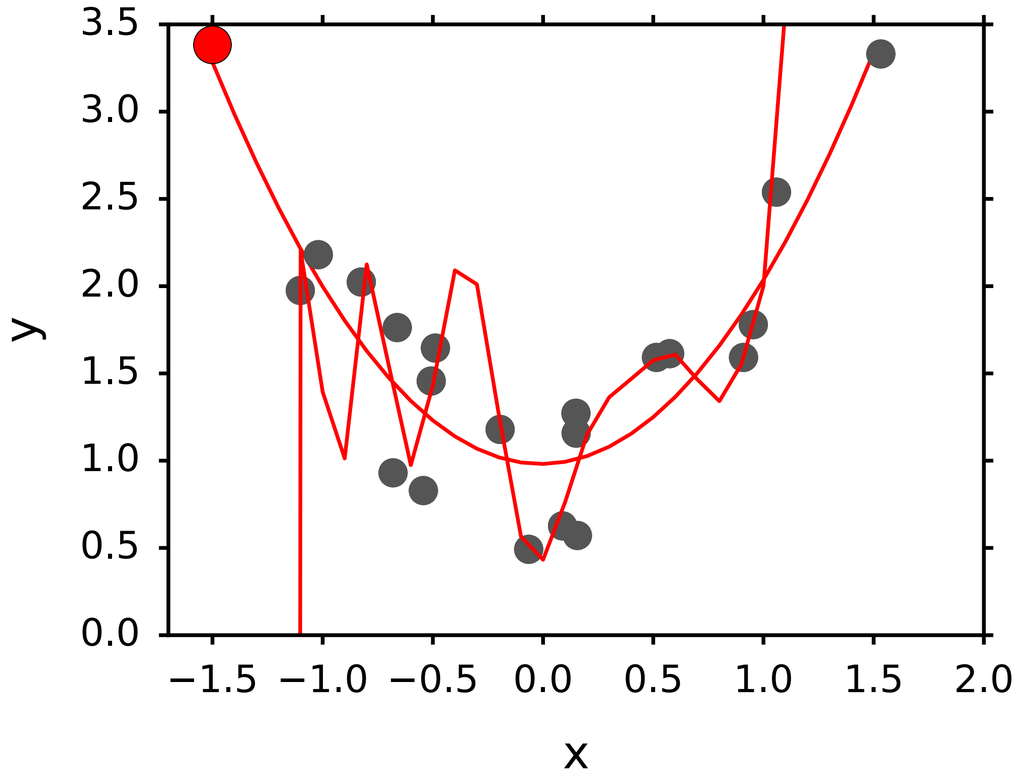
overfit

cost = 0.02



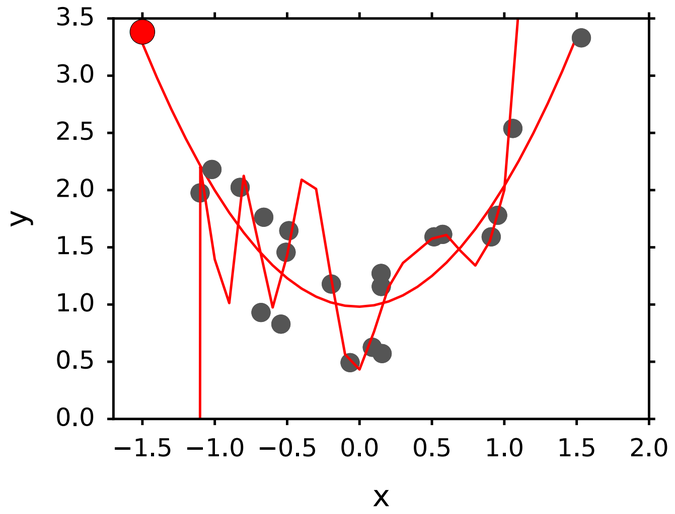
$$h(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \ldots + \theta_{15} x^{15}$$

OVERFITTED MODELS CANNOT GENERALIZE



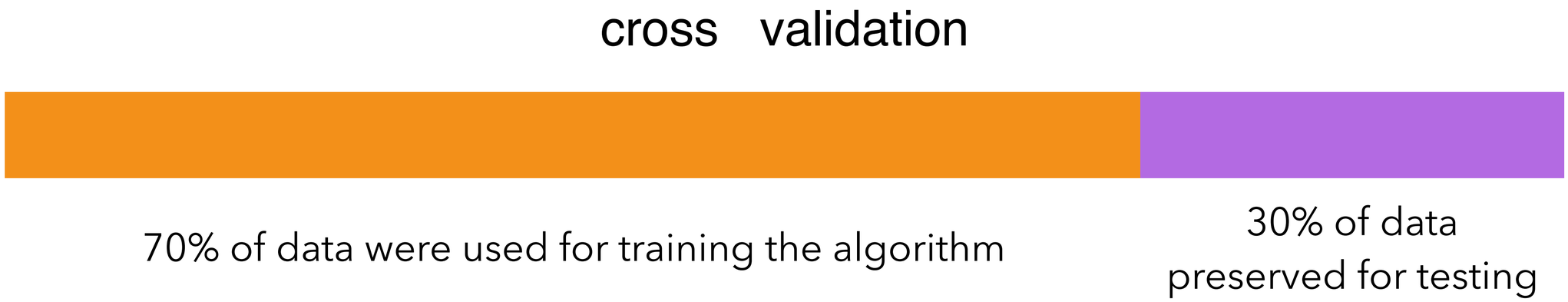
FACTS ABOUT OVERFITTING

- The more parameters you have, the more likely your model will overfit.
- The more data you have (compared to parameters) the less likely your model will overfit, because noise will be more likely to average out.



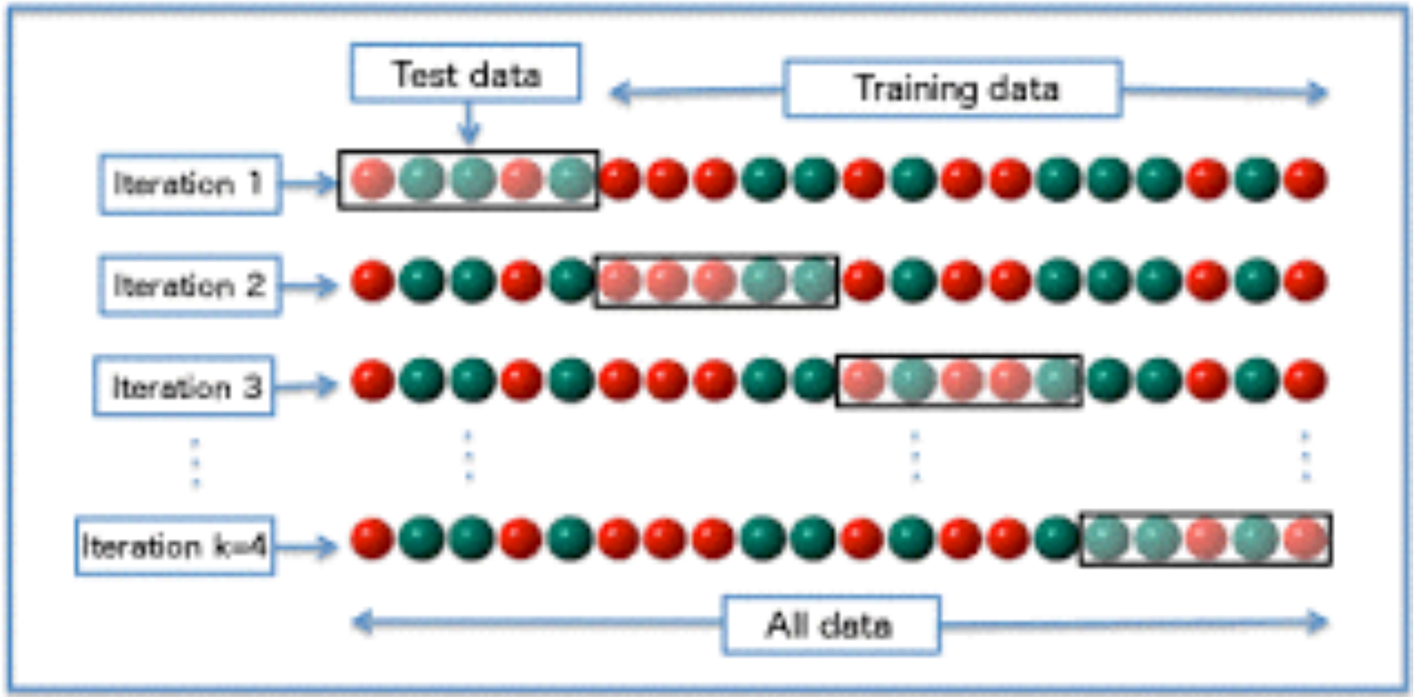
AVOID OVERFITTING

- Cross-validate your models: train models with one set of data (training set) and test them with another set of data (test set)



AVOID OVERFITTING

- K-Fold cross validation: divide data into K chunks.
- Fit the model K times, each time use one chunk as test data, the rest as training data.



AVOID OVERFITTING

- Get more data
- Reduce the number of features (select only a few features that are most powerful)
- Reduce the number of parameters (using an algorithm to remove parameters that are not necessary)
- Constrain the parameter set in the optimization process (regularization)

REGULARIZATION

- Regularization is a mathematical way to reduce overfitting automatically, by limiting the influence of each feature.

$$\frac{1}{m}\sum_{i}(h(x_i)-y_i)^2+\lambda\sqrt{\sum_{j}\|\theta_j\|^2}$$

The normal cost function

norm of parameter vector

- To minimize this cost function you have to minimize both first and second terms
- Minimize the second term means less overfitting.

REGULARIZATION

$$\frac{1}{m}\sum_{i}(h(x_i)-y_i)^2+\lambda\sqrt{\sum_{j}\|\theta_j\|^2}$$

The normal cost function

norm of parameter vector

- Lambda is called ‘regularization parameter’
- Because we adjust lambda to adjust the weight of the two cost function terms
- High lambda: severely limit parameter size
- Low lambda: allow parameters to scale up more freely, give more importance to lowering error

OVERFITTING TAKE-HOME MESSAGE

OVERFITTING IS VERY DANGEROUS.

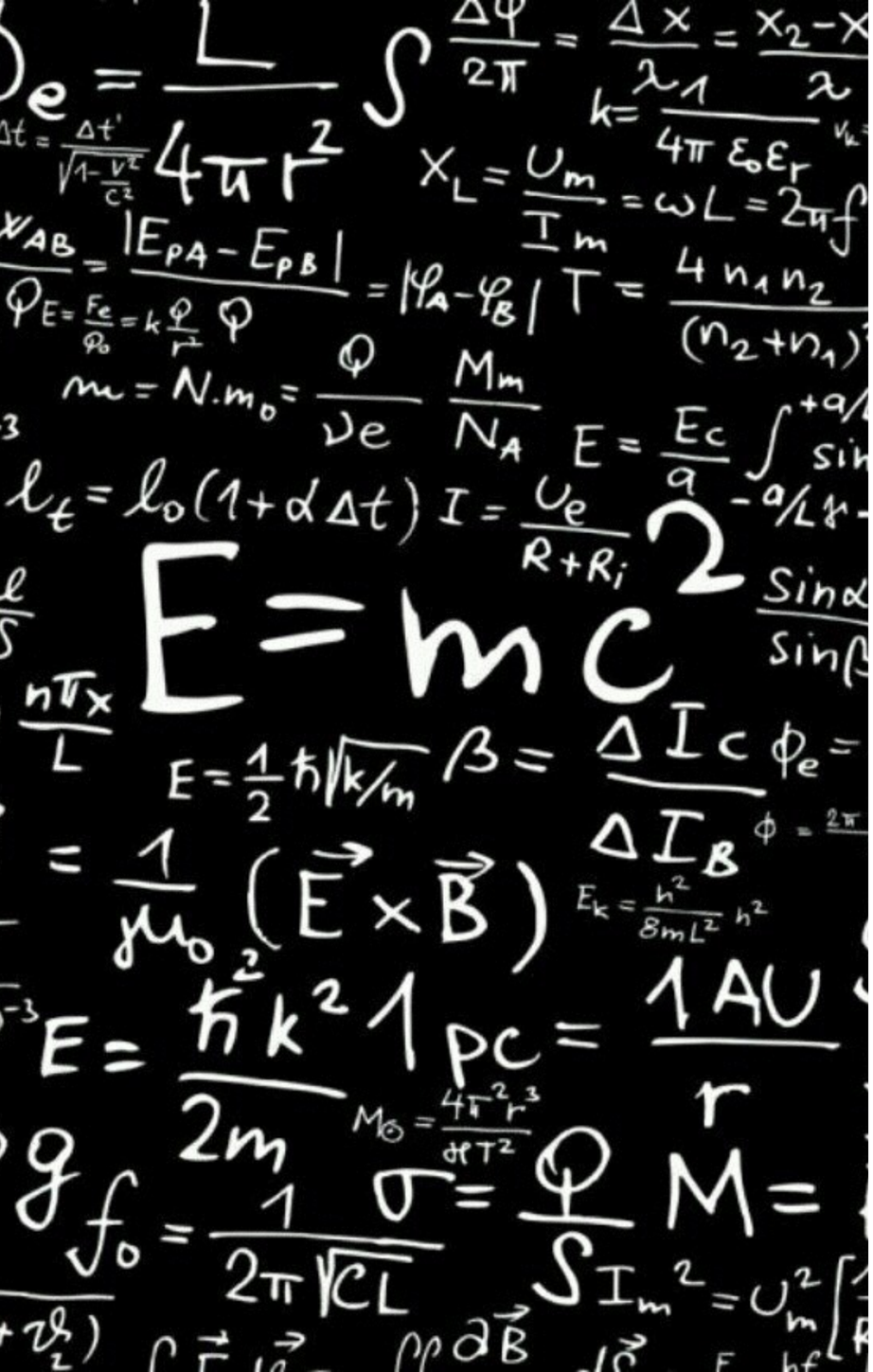
WE ARE NOT DONE, WE WILL LEARN MORE ABOUT HOW TO COMBAT THIS PROBLEM LATER.

Challenges for ML students

- Turning business problems into ML problem
- Framing the ML problem (input, output, model)
- Having an intuition about what features to look for
- Fixing the model, when straightforward library fits yield terrible results
- Improving on an already good model
- Studying new models they have not seen before and actually understand them

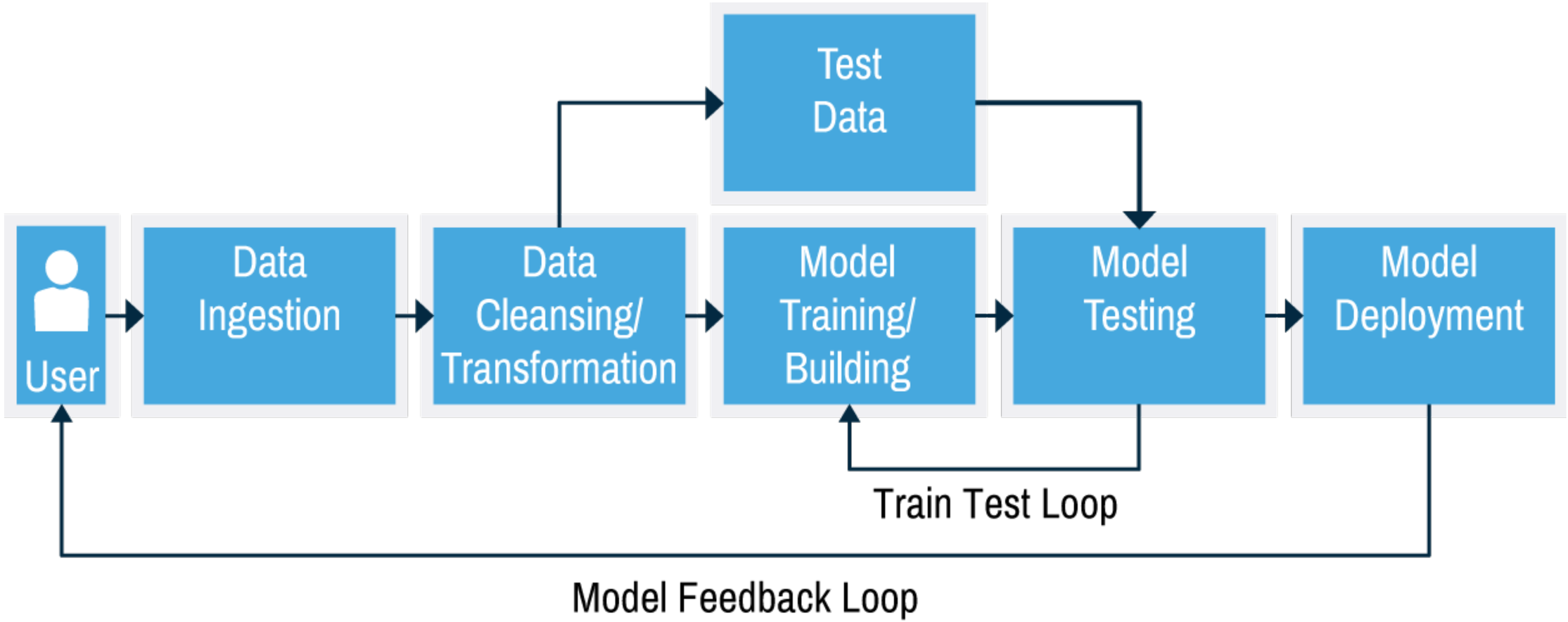
Mastering Machine Learning

- People from math, stats, science background:
  - Coding cannot be taught (like bicycle riding) it solely depends on how many hours you put in.
  - Read documentations. It makes everything much easier.
- People from computer-science background:
  - Don’t just trust your libraries to do all the work. Try to gain mathematical intuitions. It helps the innovation process.
  - Your algorithm course is not useless, computational power still turns out to be an important constraints in most things you do.



ENABLING MACHINE LEARNING

Machine Learning Workflow



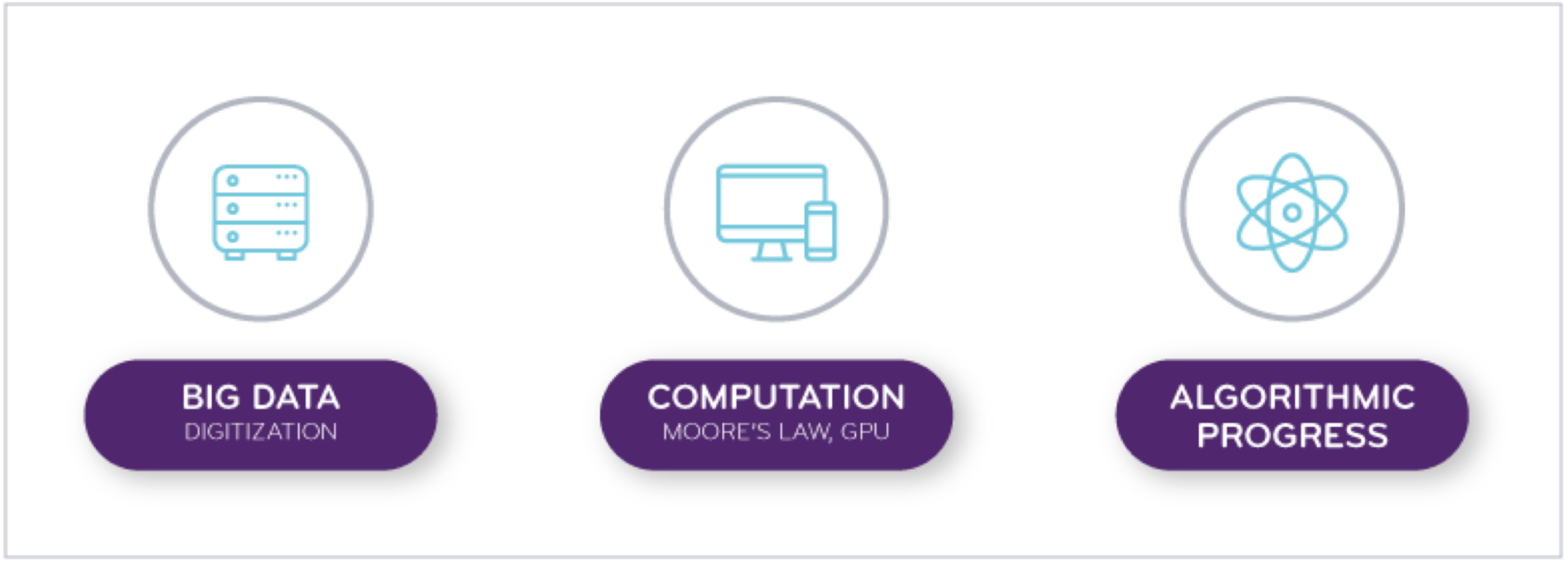
credit: mapr.com

ENABLING MACHINE LEARNING

IN THIS COURSE WE WILL LEARN ABOUT SO MANY ALGORITHMS

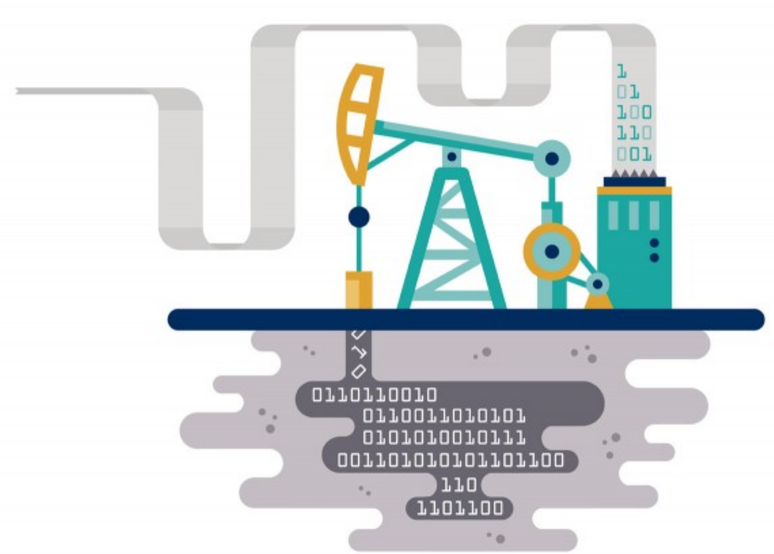
BUT WHEN YOU START DOING ML, YOU WILL REALIZE THAT ALGORITHM IS NOT THE ONLY THING.

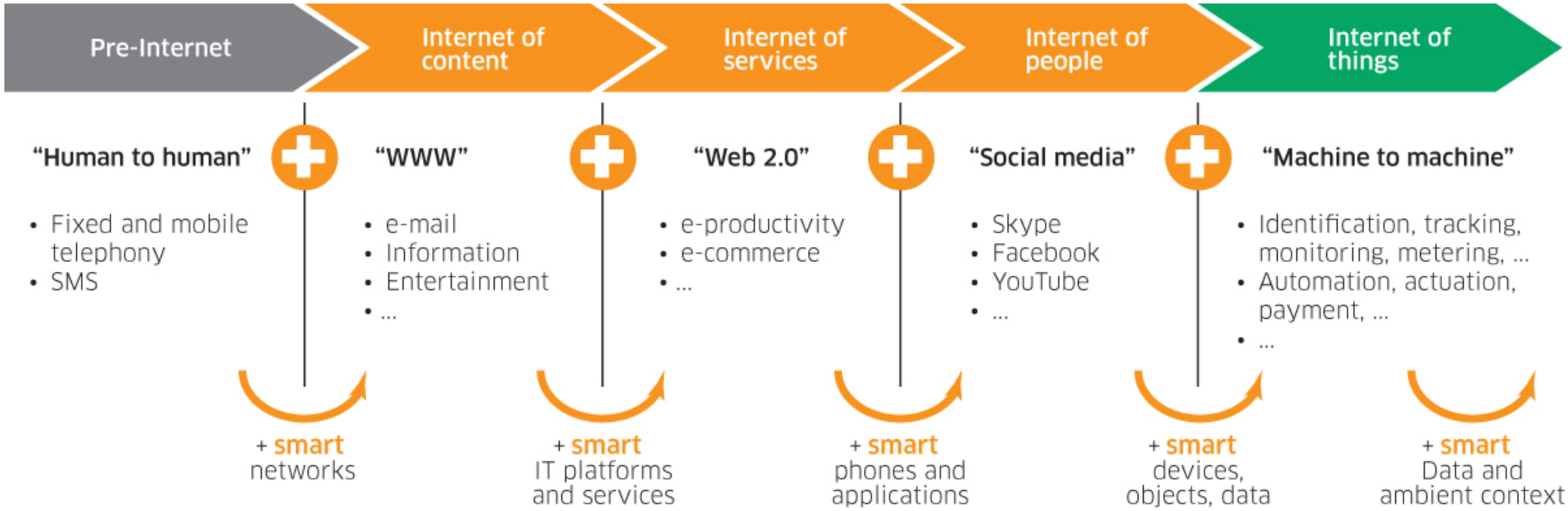
Enablers of Artificial Intelligence



Data is the New Oil

Drilling the “DATA” well for insights





More Data -> Better Models

MACHINE LEARNING
+ BIG DATA
= BETTER MODELS

DESCRIPTIVE ANALYTICS

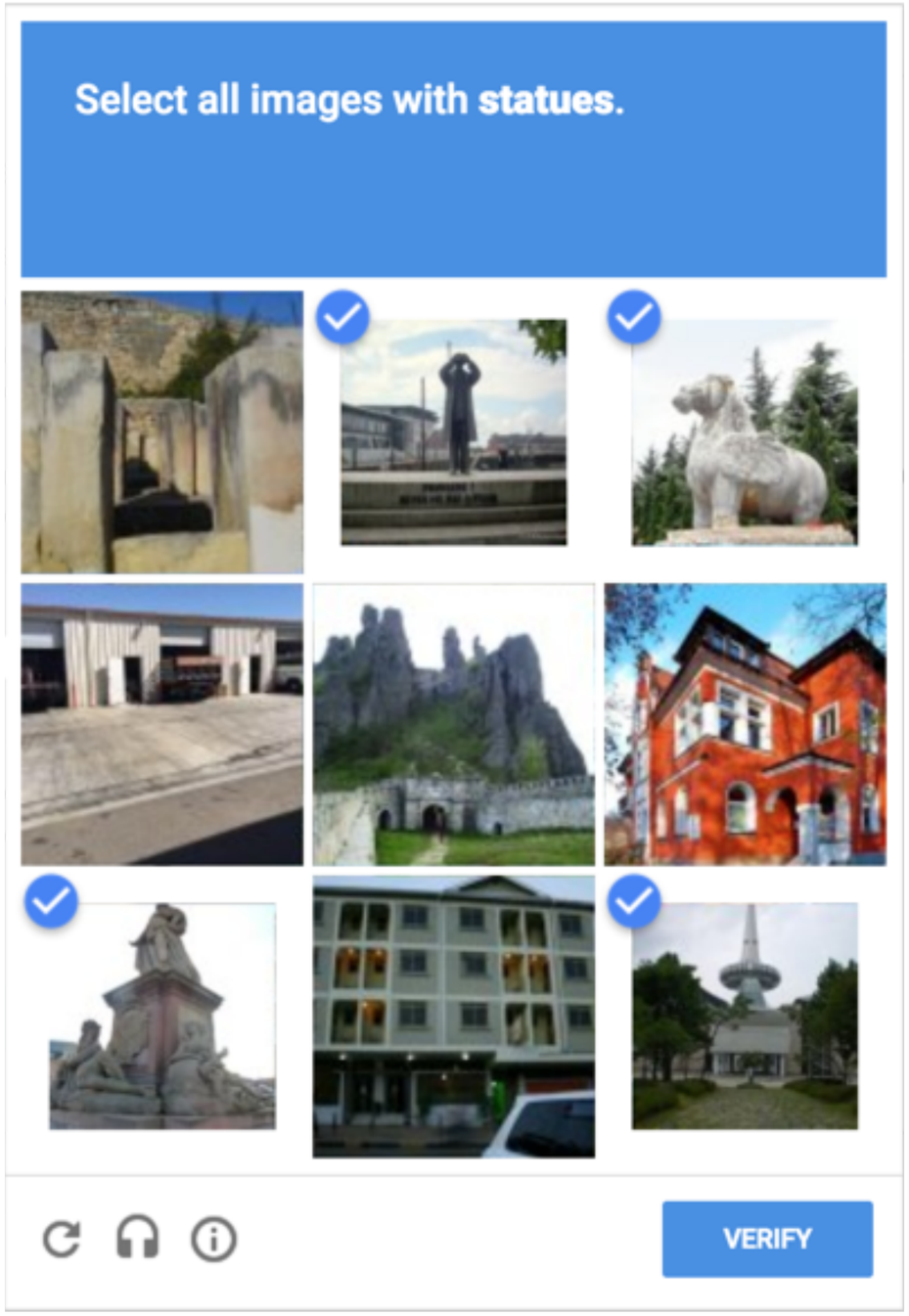
"What happened? What exactly is the problem?"

PREDICTIVE ANALYTICS

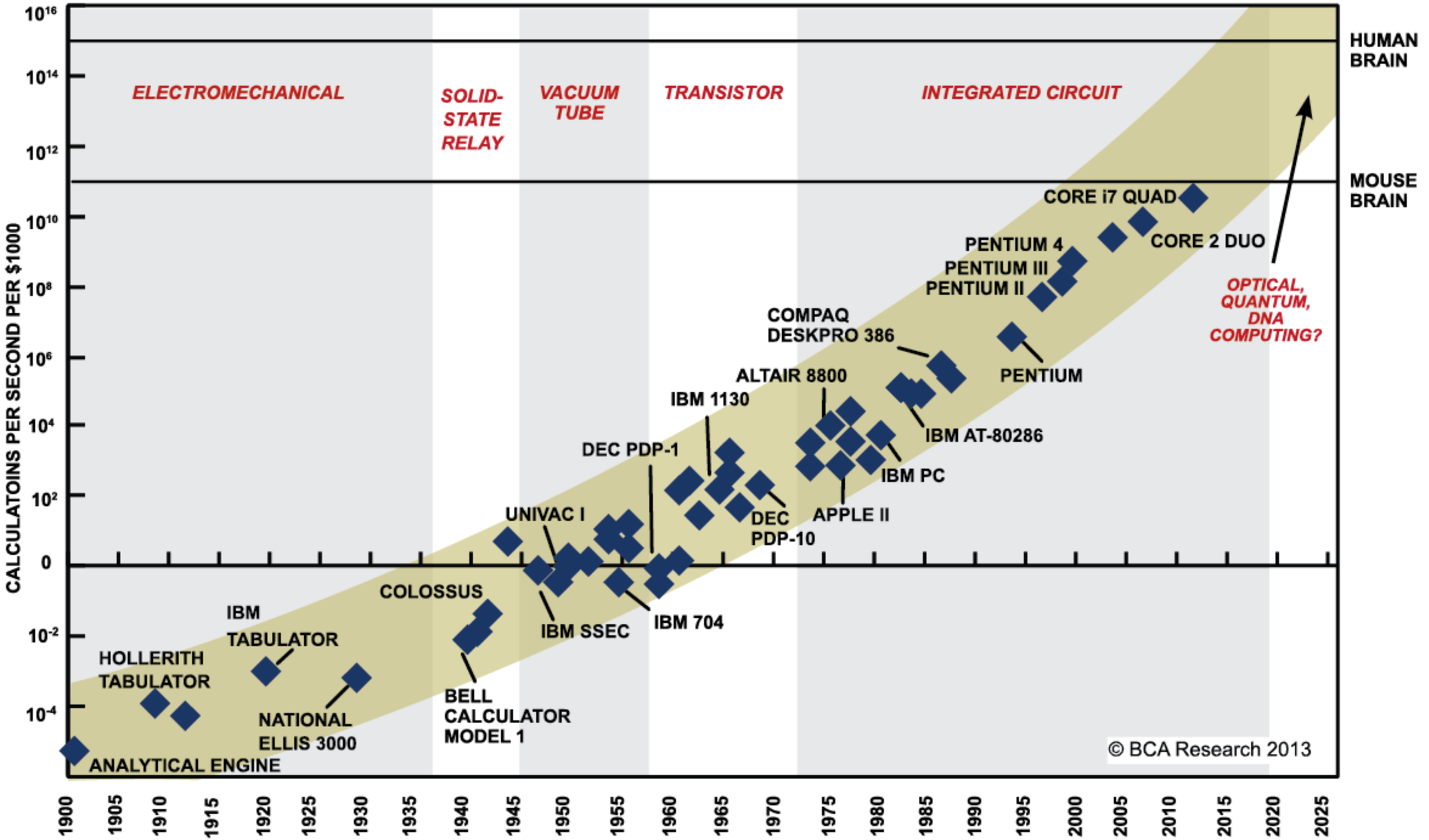
"What will happen?"

PRESCRIPTIVE ANALYTICS

"What's the best that can happen? How do we achieve it?"



- reCaptcha is a small applet that helps Google identify whether you are a robot or not when your request for service access is suspicious.
- Data are collected to train artificial intelligence.
- Aside from these examples, Google has been collecting data from your mobile phone about your locations, your search behaviors, your app usages to train AIs.



SOURCE: RAY KURZWEIL, "THE SINGULARITY IS NEAR: WHEN HUMANS TRANSCEND BIOLOGY", P.67, *THE VIKING PRESS*, 2006. DATAPOINTS BETWEEN 2000 AND 2012 REPRESENT BCA ESTIMATES.

CPU v.s. GPU

GPU speed increases exponentially in the last decade, while CPU is flat. Extremely parallel processing makes it perfect for training large neural network.

Credit: infogram.com, medium.com/@andriylazorenko

8

Number of cores in a high end desktop CPU

2688

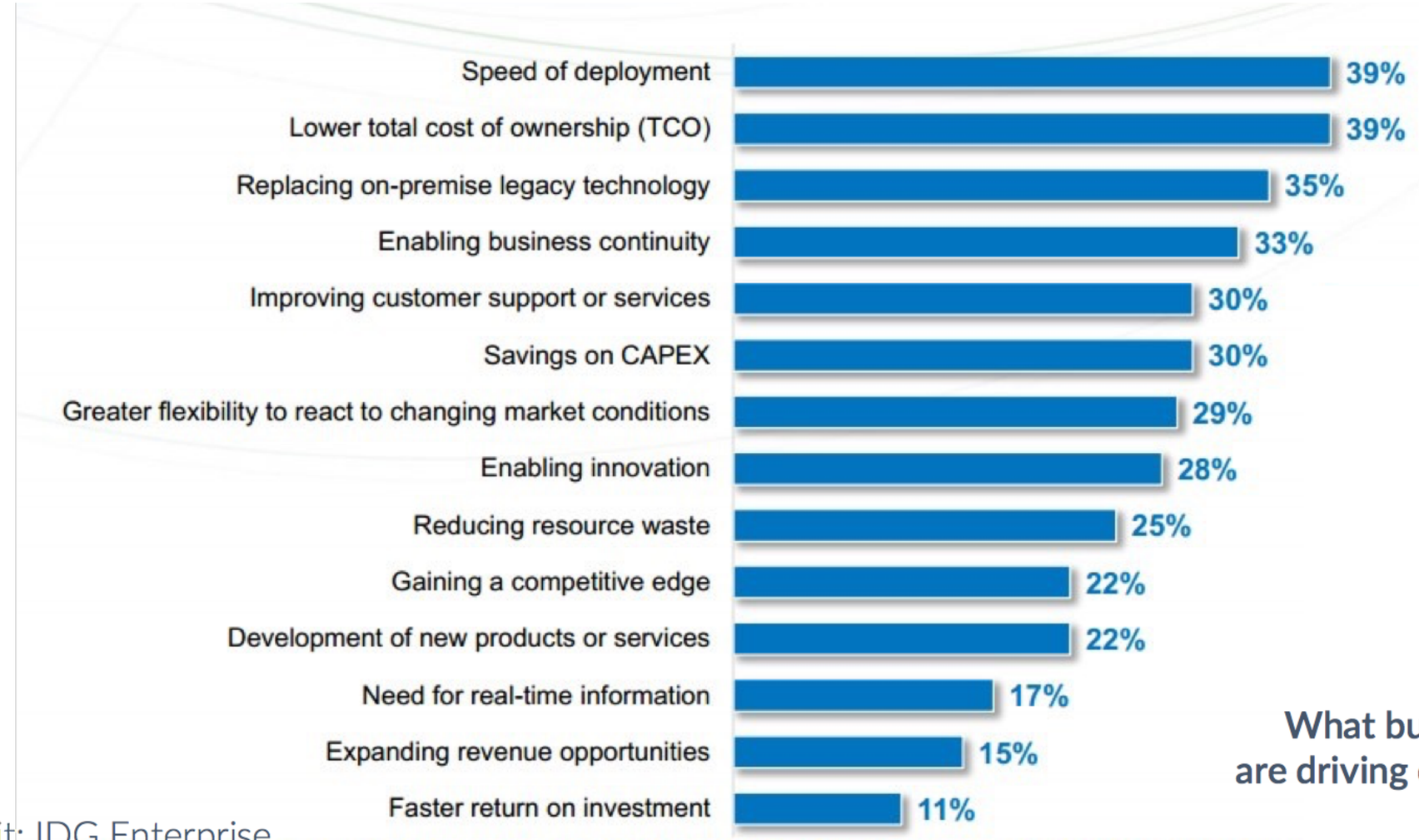
Number of cores in the nVidia GTX TITAN

Device	Speed of training, examples/sec
2 x AMD Opteron 6168	440
i7-7500U	415
GeForce 940MX	1190
GeForce 1070	6500

Cloud Computing

- Hybrid cloud adoption grew 3X, from 19% to 57% of organizations.
- 80% of all IT budgets will be committed to cloud solutions.
- 73% of companies are planning to move to a fully software-defined data center within 2 years.
- 49% of businesses are delaying cloud deployment due to a cybersecurity skills gap.

Cloud Computing



What business goals or objectives are driving cloud investments in your organization?

credit: IDG Enterprise

Cloud Computing

- Massive parallel computing is not a luxury in AI, it's necessity.
- Cloud allows scalability at a fraction of the cost.
- Zero lead time.
- World-class operational engineers.
- Latest technology at no R&D cost