Metis

INTRO TO MACHINE LEARNING

Paul Burkard
09/26/16

I. WHAT IS MACHINE LEARNING?

from Wikipedia:

“Machine learning, a branch of artificial intelligence, is about the construction and study of systems that can learn from data.”

source: http://en.wikipedia.org/wiki/Machine_learning

from Wikipedia:

“Machine learning, a branch of artificial intelligence, is about the construction and study of systems that can learn from data.”

“The core of machine learning deals with representation and generalization…”

source: http://en.wikipedia.org/wiki/Machine_learning

from Wikipedia:

“Machine learning, a branch of artificial intelligence, is about the construction and study of systems that can learn from data.”

“The core of machine learning deals with representation and generalization…”

- representation – extracting structure from data

source: http://en.wikipedia.org/wiki/Machine_learning

from Wikipedia:

"Machine learning, a branch of artificial intelligence, is about the construction and study of systems that can learn from data."

"The core of machine learning deals with representation and generalization…"

- representation – extracting structure from data
- generalization – making predictions from data

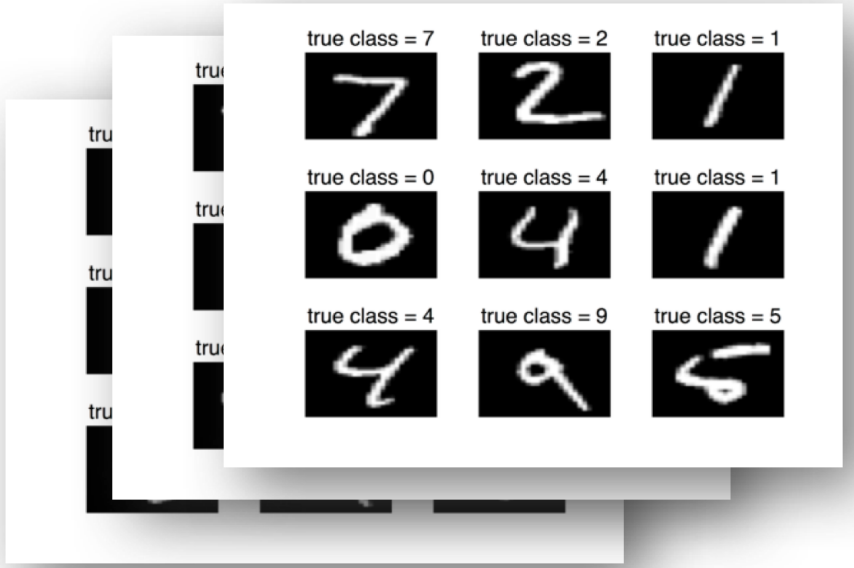
source: http://en.wikipedia.org/wiki/Machine_learning

Learning is not about memorizing and being able to recall, it is about generalizing the conclusions to previously unseen examples

II. MACHINE LEARNING PROBLEMS

Supervised learning: the goal is to learn mapping from given inputs x to outputs y, given a **labeled** set of input-output pairs


OCR

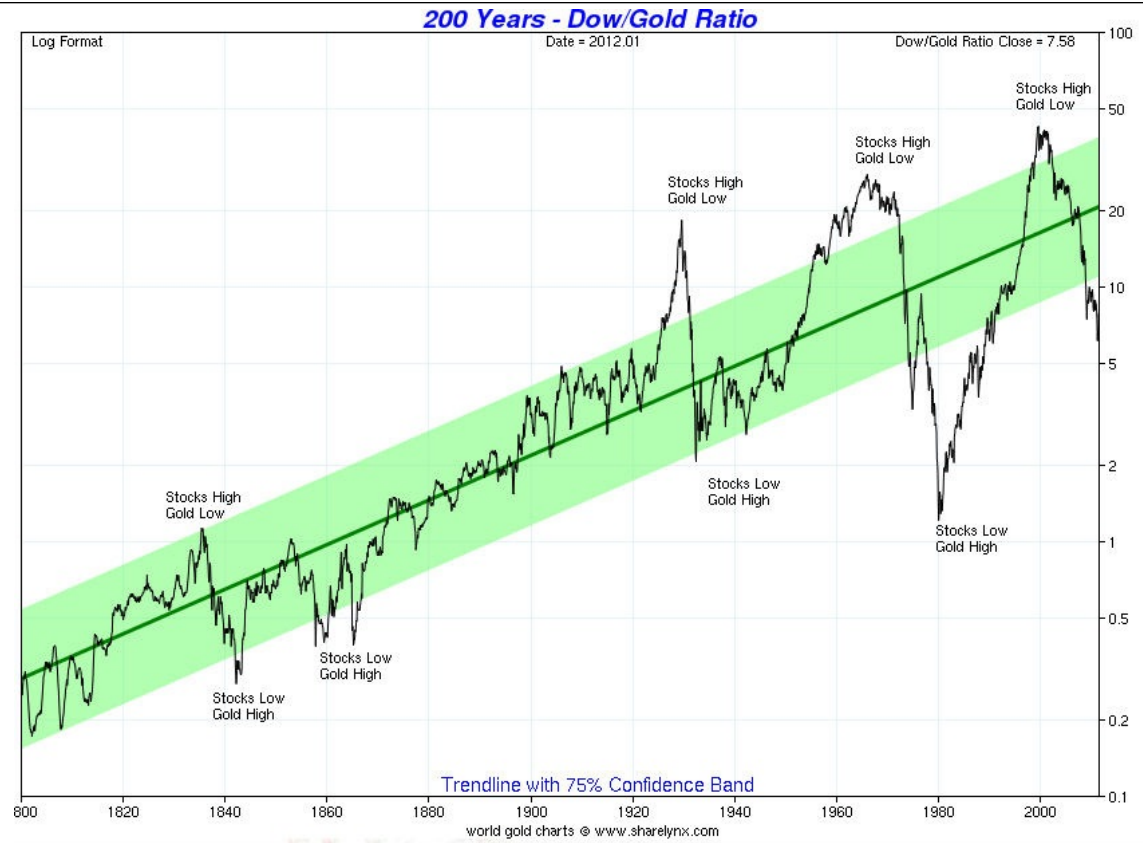

true class = 7
true class = 2
true class = 1
true class = 0
true class = 4
true class = 1
true class = 4
true class = 9
true class = 5

	Client 1	*Client 2*	*Client 3*
Age	*23*	*30*	*19*
Gender	*M*	*F*	*M*
Annual salary	*$30,000*	*$45,000*	*$15,000*
Years in residence	*3 years*	*1 year*	*3 month*
Years in job	*1 year*	*1 year*	*1 month*
Current debt	*$5,000*	*$1,000*	*$10,000*
Paid off credit	*Yes*	*Yes*	*No*

	Client 1	Client 2	Client 3
Age	23	30	19
Gender	M	F	M
Annual salary	$30,000	$45,000	$15,000
Years in residence	3 years	1 year	3 month
Years in job	1 year	1 year	1 month
Current debt	$5,000	$1,000	$10,000
Paid off credit	Yes	Yes	No

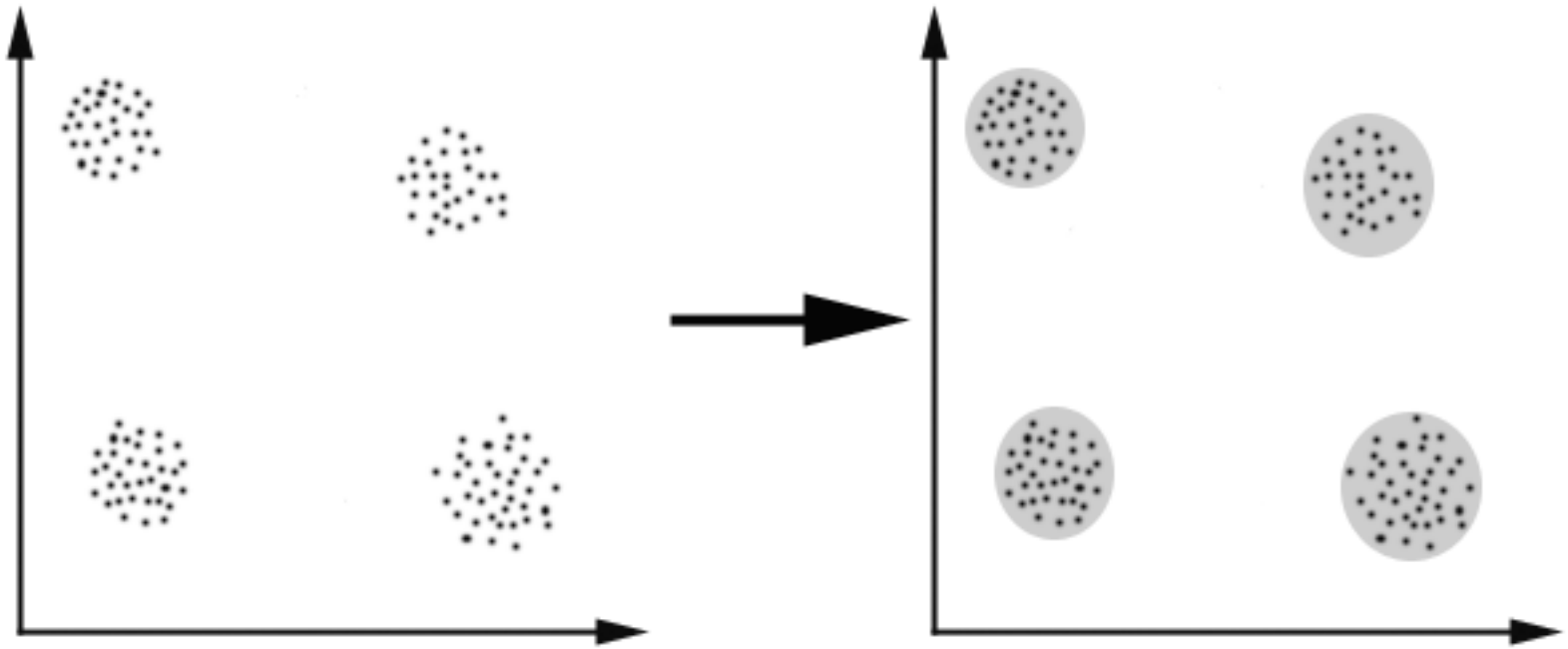
	Applicant
Age	25
Gender	M
Annual salary	$25,000
Years in residence	1 year
Years in job	2 year3
Current debt	$15,000
Credit decision/ score	???


SPAM

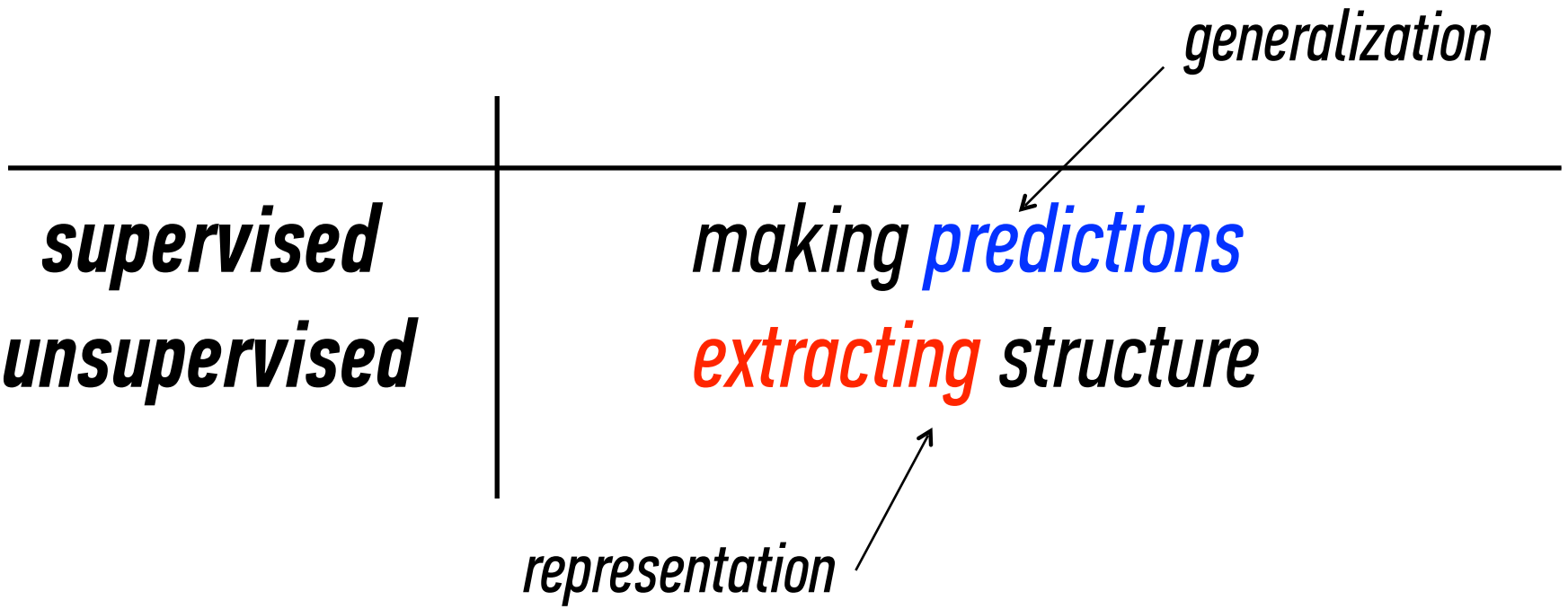

200 Years - Dow/Gold Ratio
Log Format
Date = 2012.01
Dow/Gold Ratio Close = 7.58
Stocks High
Gold Low
Stocks Low
Gold High
Stocks High
Gold Low
Stocks Low
Gold High
Stocks High
Gold Low
Stocks Low
Gold High
Stocks High
Gold Low
Stocks Low
Gold High
Trendline with 75% Confidence Band
100
50
20
10
5
2
1
0.5
0.2
0.1
800
1820
1840
1860
1880
1900
1920
1940
1960
1980
2000
world gold charts © www.sharelynx.com

Unsupervised learning: the goal is to learn interesting patterns and structure in data given only inputs

no label information given at all

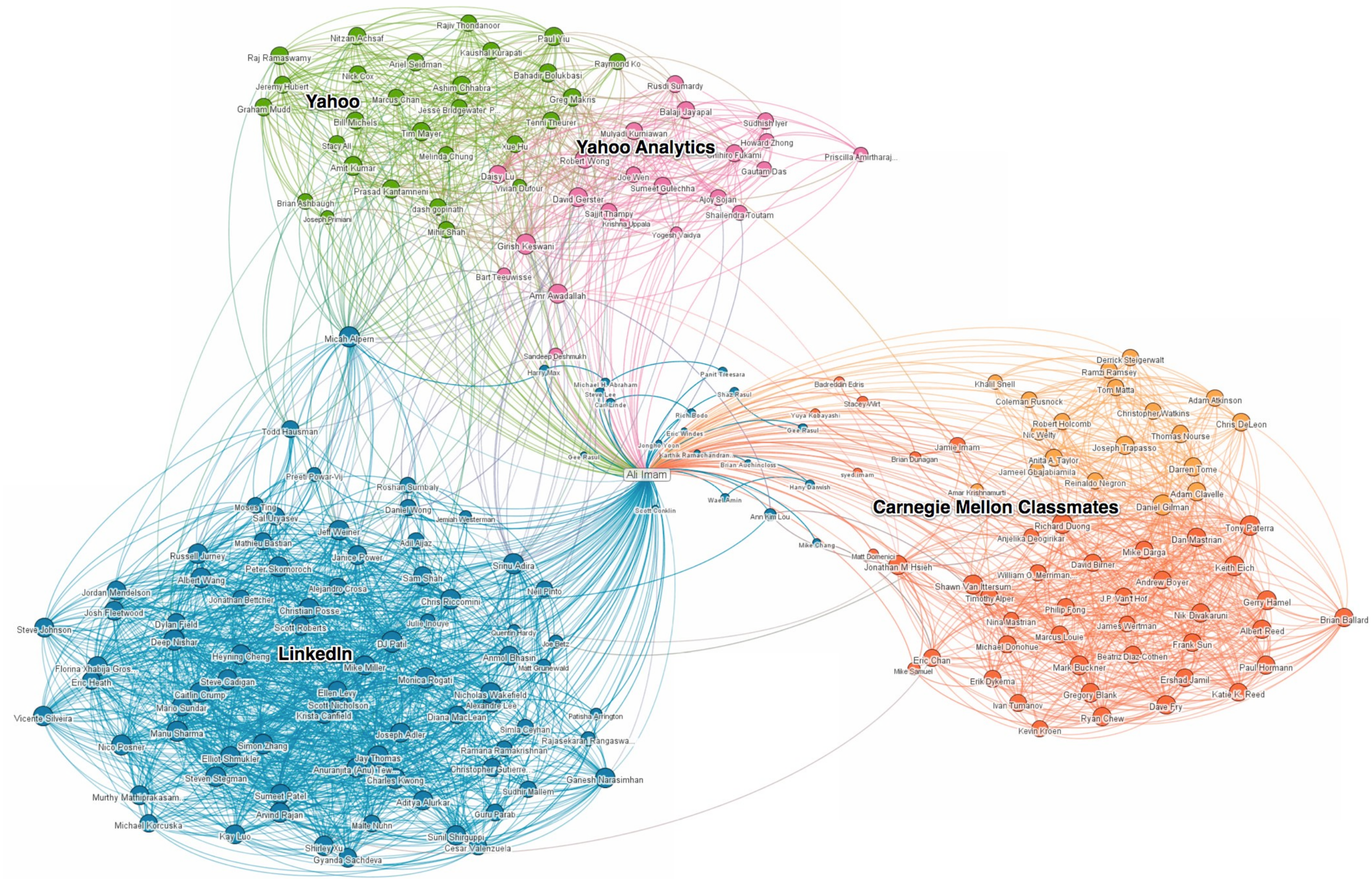


supervised	*making predictions*
unsupervised	*extracting structure*


generalization
supervised
unsupervised
making predictions
extracting structure
representation

EXERCISE:

supervised or unsupervised?



Yahoo
Yahoo Analytics
LinkedIn
Carnegie Mellon Classmates
Ali Imam
Rajiv Thondanoor
Nitzan Achsaf
Paul Yiu
Raj Ramaswamy
Kaushal Kurapati
Ariel Seidman
Raymond Ko
Nick Cox
Bahadir Bolukbasi
Jeremy Hubert
Ashim Chhabra
Rusdi Sumardy
Graham Mudd
Marcus Chan
Greg Makris
Jesse Bridgewater
Balaji Jayapal
Bill Michels
Tenni Theurer
Sudhish Iyer
Tim Mayer
Mulyadi Kurniawan
Howard Zhong
Stacy Ali
Xue Hu
Melinda Chung
Robert Wong
Chihiro Fukami
Priscilla Amirtharaj...
Amit Kumar
Daisy Lu
Joe Wen
Gautam Das
Prasad Kantamneni
Vivian Dufour
Sumeet Gulechha
Brian Ashbaugh
David Gerster
Ajoy Sojan
dash gopinath
Sajjit Thampy
Shailendra Toutam
Joseph Primiani
Krishna Uppala
Mihir Shah
Yogesh Vaidya
Girish Keswani
Bart Teeuwisse
Amr Awadallah
Micah Alpern
Sandeep Deshmukh
Harry Max
Michael H. Abraham
Steve Lee
Carl Linde
Panit Treesara
Shaz Rasul
Rich Bodo
Gee Rasul
Eric Windes
Jongho Yoon
Karthik Ramachandran
Brian Auchincloss
Hany Darwish
Wael Amin
Ann Kim Lou
Mike Chang
Scott Conklin
Todd Hausman
Preeti Powar-Vij
Roshan Sumbaly
Moses Ting
Daniel Wong
Sal Uryasev
Jemiah Westerman
Jeff Weiner
Mathieu Bastian
Adil Aijaz
Russell Jurney
Janice Power
Srinu Adira
Peter Skomoroch
Sam Shah
Albert Wang
Alejandro Crosa
Neil Pinto
Jordan Mendelson
Jonathan Bettcher
Chris Riccomini
Josh Fleetwood
Christian Posse
Julie Inouye
Dylan Field
Scott Roberts
Quentin Hardy
Steve Johnson
Deep Nishar
DJ Patil
Joe Betz
Heyning Cheng
Anmol Bhasin
Florina Xhabija Gros...
Mike Miller
Matt Grunewald
Eric Heath
Monica Rogati
Steve Cadigan
Caitlin Crump
Ellen Levy
Nicholas Wakefield
Mario Sundar
Scott Nicholson
Alexandre Lee
Vicente Silveira
Krista Canfield
Diana MacLean
Patisha Arrington
Manu Sharma
Simla Ceyhan
Joseph Adler
Simon Zhang
Rajasekaran Rangaswa...
Nico Posner
Elliot Shmukler
Jay Thomas
Ramana Ramakrishnan
Anuranjita (Anu) Tew...
Christopher Gutierre...
Steven Stegman
Charles Kwong
Ganesh Narasimhan
Sudhir Mallem
Murthy Mathiprakasam...
Sumeet Patel
Aditya Alurkar
Arvind Rajan
Guru Parab
Michael Korcuska
Malte Nuhn
Kay Luo
Sunil Shirguppi
Shirley Xu
Cesar Valenzuela
Gyanda Sachdeva
Derrick Steigerwalt
Ramzi Ramsey
Badreddin Edris
Khalil Snell
Tom Matta
Stacey Wirt
Coleman Rusnock
Adam Atkinson
Yuya Kobayashi
Christopher Watkins
Robert Holcomb
Chris DeLeon
Nic Welty
Thomas Nourse
Jamie Imam
Joseph Trapasso
Brian Dunagan
Anita A. Taylor
Darren Tome
syed imam
Jameel Gbajabiamila
Reinaldo Negron
Amar Krishnamurti
Adam Clavelle
Daniel Gilman
Richard Duong
Tony Paterra
Anjelika Deogirikar
Dan Mastrian
Mike Darga
Matt Domenici
Jonathan M Hsieh
David Birner
Keith Eich
William O. Merriman
Andrew Boyer
Shawn Van Ittersum
Timothy Alper
J.P. Van't Hof
Gerry Hamel
Philip Fong
Nik Divakaruni
Brian Ballard
Nina Mastrian
James Wertman
Albert Reed
Marcus Louie
Frank Sun
Michael Donohue
Beatriz Diaz-Cothen
Eric Chan
Mike Samuel
Mark Buckner
Paul Hormann
Erik Dykema
Ershad Jamil
Gregory Blank
Katie K. Reed
Ivan Tumanov
Dave Fry
Ryan Chew
Kevin Kroen


CHEQUE
APPLICATION
ORDER
CLAIM

	continuous	*categorical*
	quantitative	*qualitative*

continuous	*categorical*
Height of children	*Eye colors*
Weight of cars	*Courses at GA*
Speed of the train	*Highest degree*
Temperature	*Gender*
Stock price	*If an email is spam or not*

	continuous	*categorical*
	quantitative	*qualitative*

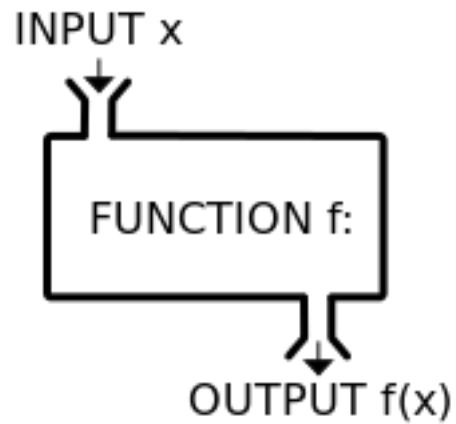
NOTE

The space where data live is called the feature space.

Each point in this space is called a record.

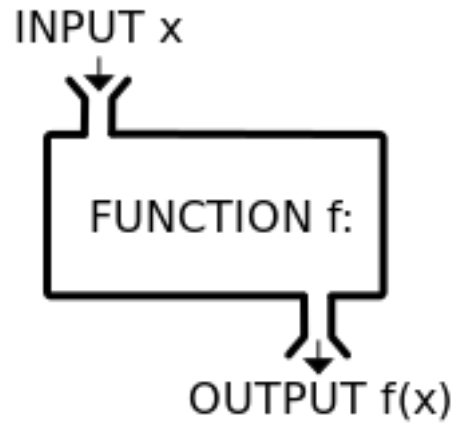
	continuous	*categorical*
supervised	*regression*	*classification*
unsupervised	*dim reduction*	*clustering*

– *At its core*, **Supervised Learning** *boils down to* **Function Approximation**

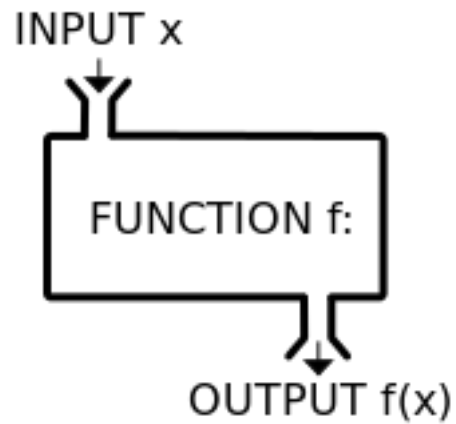


- *At its core,* **Supervised Learning** *boils down to* **Function Approximation**
  - *Given some* **inputs, x**, *there is some unknown* **function f(x)** *that perfectly* **maps** *our* **inputs** *to the* **outputs, y**:

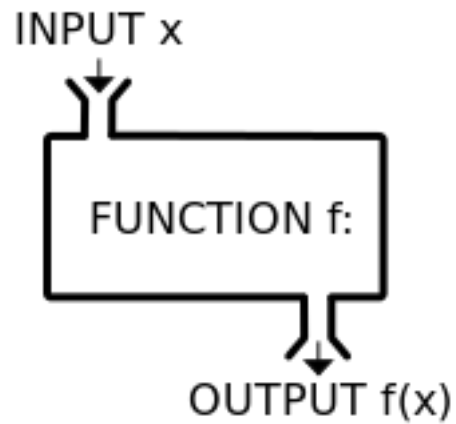
    y = f(x)



- *At its core,* **Supervised Learning** *boils down to* **Function Approximation**
  - *Given some* **inputs, x**, *there is some unknown* **function f(x)** *that perfectly* **maps** *our* **inputs** *to the* **outputs, y**:
    y = f(x)
  - *Our task is just to approximate f(x)*



- *At its core,* **Supervised Learning** *boils down to* **Function Approximation**
  - *Given some* **inputs, x**, *there is some unknown* **function f(x)** *that perfectly* **maps** *our* **inputs** *to the* **outputs, y**:
    y = f(x)
  - *Our task is just to approximate f(x)*
  - *e.g. Linear Regression:* $f(x) = \alpha + \beta x$



TYPES OF LEARNING PROBLEMS

supervised	*labeled examples*
unsupervised	*no labeled examples*

- ***Model Inputs:*** *Features, Attributes, Predictors, Inputs, Independent Variables, Dimensions, probably more...*

- ***Model Inputs:*** *Features, Attributes, Predictors, Inputs, Independent Variables, Dimensions, probably more...*
- ***Model Outputs*** *(what we're trying to predict): Target, Response, Output, Dependent Variable, Labels*

- ***Model Inputs:*** *Features, Attributes, Predictors, Inputs, Independent Variables, Dimensions, probably more...*
- ***Model Outputs*** *(what we're trying to predict): Target, Response, Output, Dependent Variable, Labels*
- ***Row of Data*** *(Inputs + Outputs): Observation, Datapoint, Record, Row*

- ***Model Inputs:*** *Features, Attributes, Predictors, Inputs, Independent Variables, Dimensions, probably more...*
- ***Model Outputs*** *(what we're trying to predict): Target, Response, Output, Dependent Variable, Labels*
- ***Row of Data*** *(Inputs + Outputs): Observation, Datapoint, Record, Row*
- ***Labels****: The values on the target variables in Supervised Learning*

– **Supervised Learning** *becomes a task of* **guessing the function** *that* **maps features onto targets**

- **Supervised Learning** *becomes a task of* **guessing the function** *that* **maps features onto targets**
  - *Each observation will have a* **vector** *(value for each dimension aka feature)* **of features**

- **Supervised Learning** *becomes a task of* **guessing the function** *that* **maps features onto targets**
  - *Each observation will have a* **vector** *(value for each dimension aka feature)* **of features**
  - *Each observation (usually) has just* **one target variable**

- **Supervised Learning** *becomes a task of* **guessing the function** *that* **maps features onto targets**
  - *Each observation will have a* **vector** *(value for each dimension aka feature)* **of features**
  - *Each observation (usually) has just* **one target variable**
  - *The* **labels** *are the* **target variables**

- **Supervised Learning** *becomes a task of* **guessing the function** *that* **maps features onto targets**
  - *Each observation will have a* **vector** *(value for each dimension aka feature)* **of features**
  - *Each observation (usually) has just* **one target variable**
  - *The* **labels** *are the* **target variables**
- **Unsupervised Learning** *has* **no labels**

- **Supervised Learning** *becomes a task of* **guessing the function** *that* **maps features onto targets**
  - *Each observation will have a* **vector** *(value for each dimension aka feature)* **of features**
  - *Each observation (usually) has just* **one target variable**
  - *The* **labels** *are the* **target variables**
- **Unsupervised Learning** *has* **no labels**
  - *Thus its task is to determine inter feature relationships*

- **Function Approximation isn’t everything!**

- **Function Approximation isn't everything!**
  - *The biggest thing is finding the right features, aka right x*

- **Function Approximation isn't everything!**
  - *The biggest thing is finding the right features, aka right x*
  - *This is* **Feature Engineering**

- **Function Approximation isn't everything!**
  - *The biggest thing is finding the right features, aka right x*
  - *This is* **Feature Engineering**
  - *Data Science runs on Feature Engineering*