What is Machine Learning?

Nate Derbinsky

Associate Teaching Professor

Director of Teaching Faculty

Khoury College of Computer Sciences

My Path to Khoury @ Northeastern

|	1998-2009	**BitX Solutions, Inc.** Founder & President • {.gov .edu .org .com} x {desktop web mobile}
	2002-2006	**NC State.** BS Computer Science • TA, DBMS
	2006-2012	**U of Michigan.** MS/PhD Comp Sci and Eng • TA, AI+DBMS
	2012-2014	**Disney Research.** Postdoctoral Associate • Machine Learning, Optimization, Robotics
	2014-2017	**Wentworth.** Assistant Professor • 3-3, Research/Service Learning

What is Machine Learning?
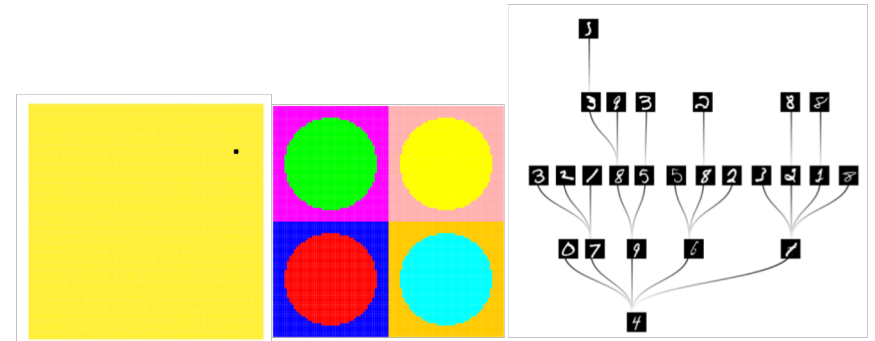
Research Interests

Cognitive Systems
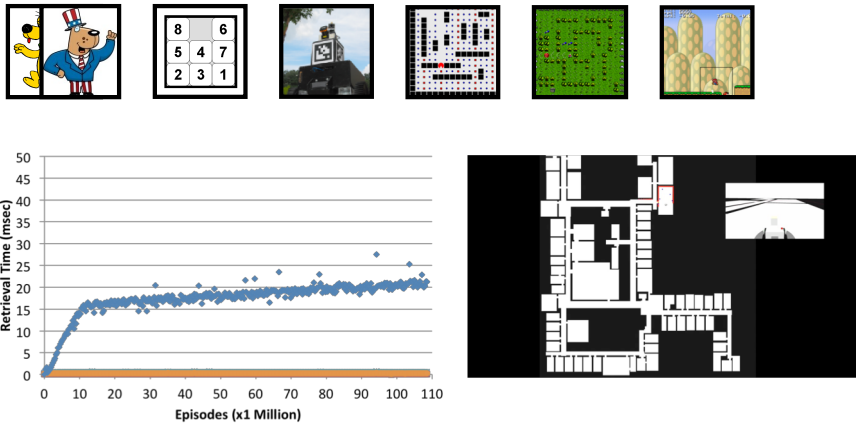
Scalable Optimization
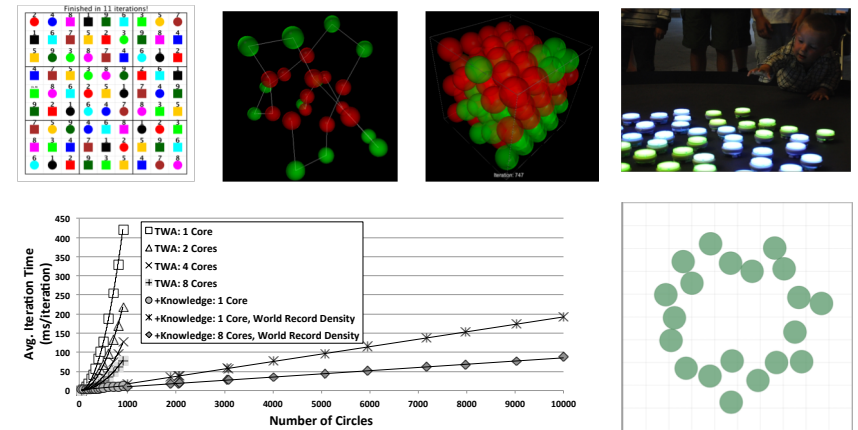
AI Applications/Education

Online ML

Teaching

K-12/ICT-D

UG/Grad

- CS/DS 1/2
  - OOP, Foundations

- Databases, Web

- AI, Machine Learning

- HTMAA
  - RPi, Arduino

What is Machine Learning?

DS2000: Programming with Data
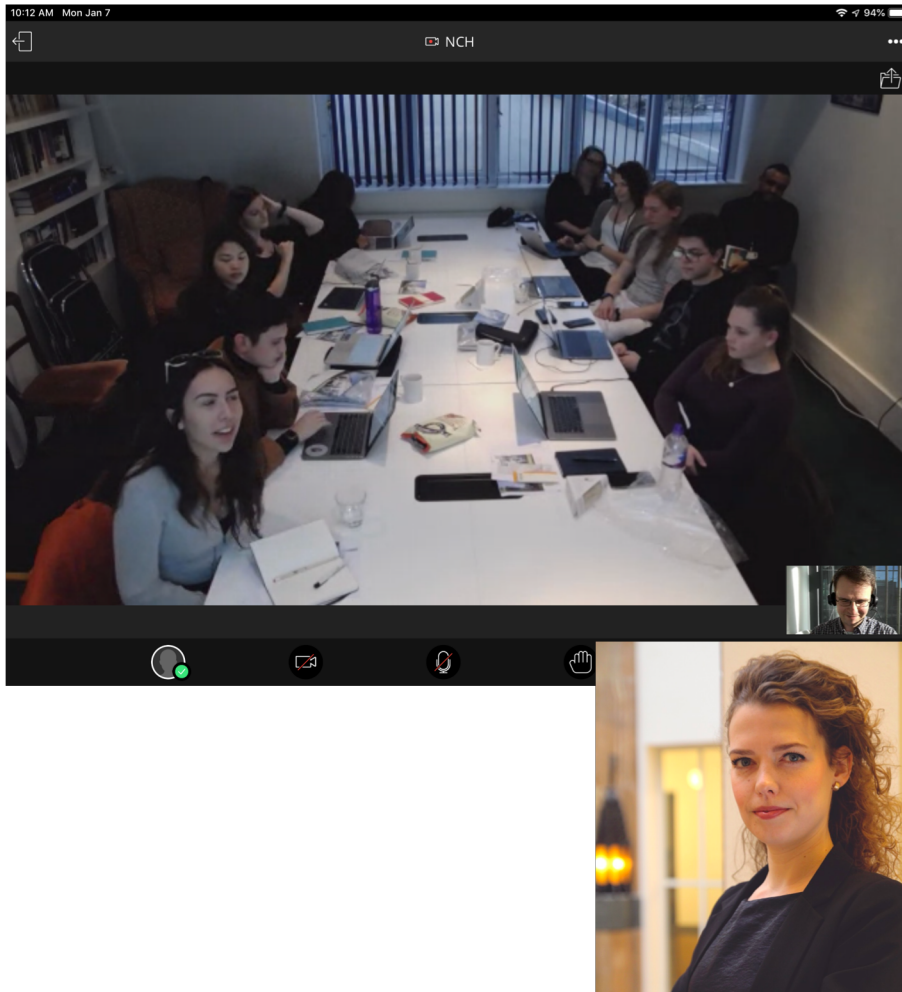
Topics

- Python
  - Get data in, process, visualize, use others' code
- Datasets
  - Files, APIs
- 2-2 split of lecture/Practicum
  - Students choose
  - Taught by discipline experts

If you only take one programming class ever, what should it be?!
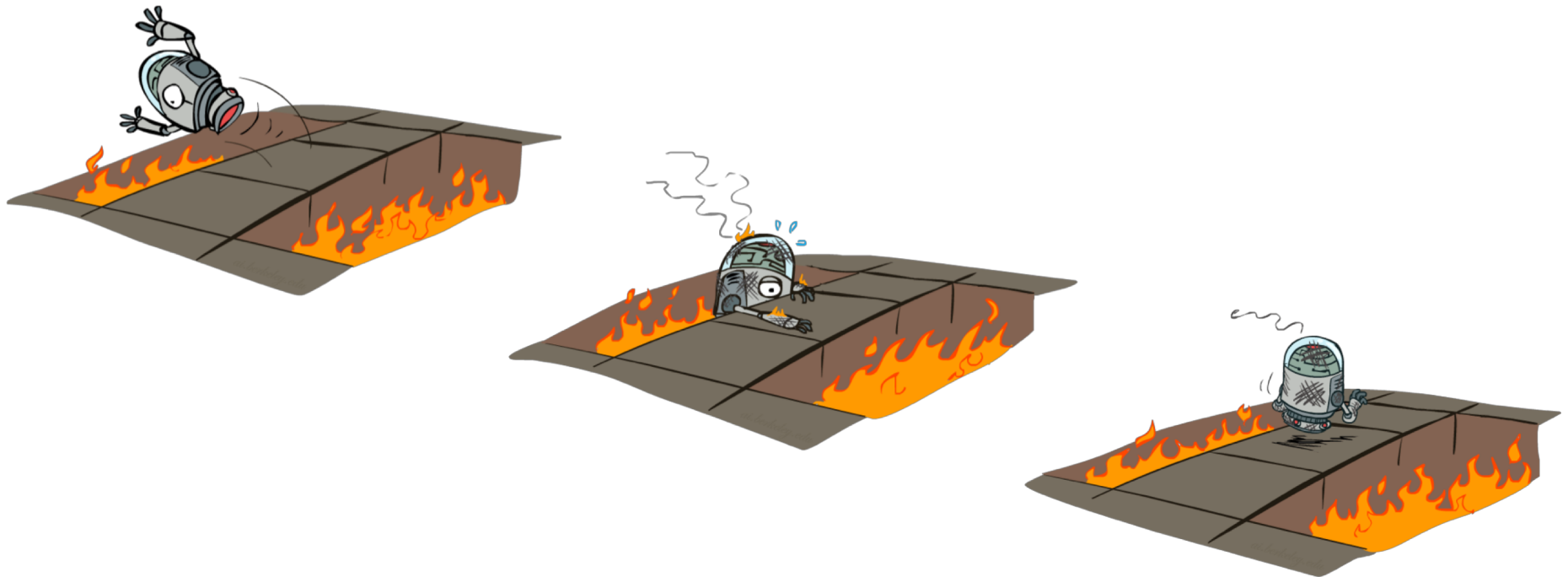
What is Machine Learning?

Agenda

- What is Machine Learning?

- Key Terminology/Tasks

- Challenges/Issues

What is Machine Learning?

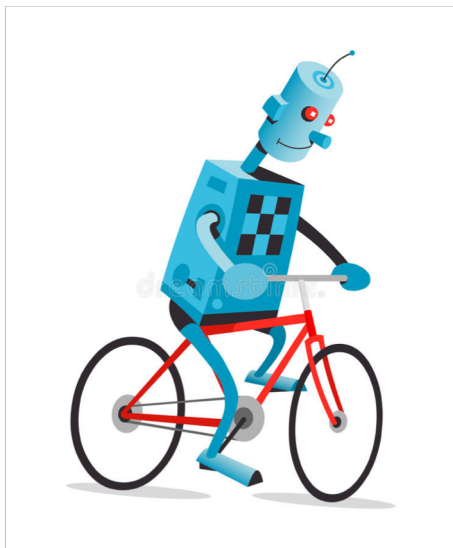# What is Machine Learning?

Computer programs that can <u>improve performance with experience</u>

But Wait…

Why Learn?

Many complex tasks are <u>hard to describe</u>, but <u>easy to learn</u> from experience

Why Now?

<u>Data</u> sources and powerful <u>computing</u> are increasingly <u>cheap and plentiful</u>

What is Machine Learning?

Natural Language Processing (NLP)

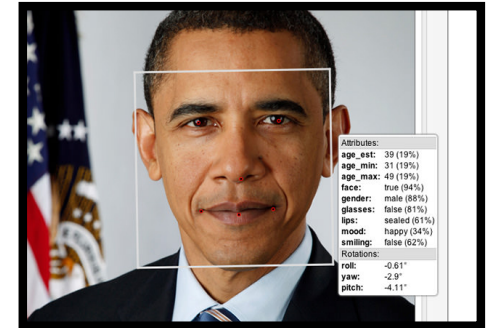Modern NLP algorithms are typically based on statistical ML

Applications

– Summarization

– Machine Translation

– Speech Processing
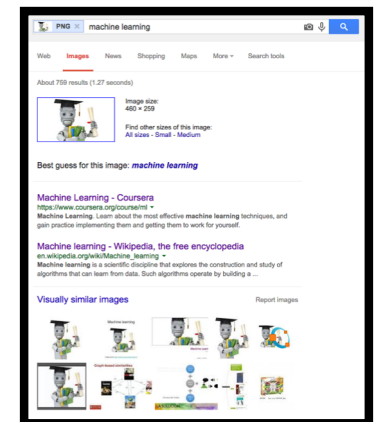
– Sentiment Analysis

…

Computer Vision

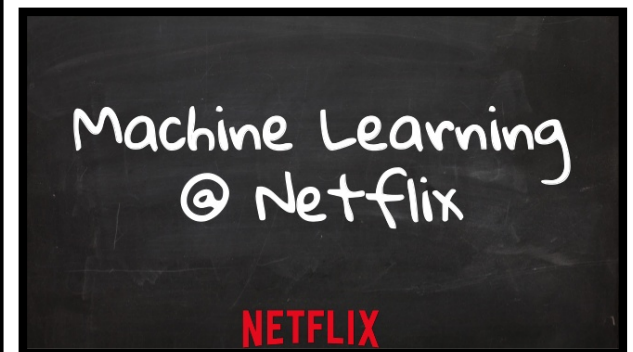Methods for acquiring, processing, analyzing, and understanding images

Applications

– Image search

– Facial recognition

– Object tracking

– Image restoration

…

Games, Robotics, Medicine, Ads, ...
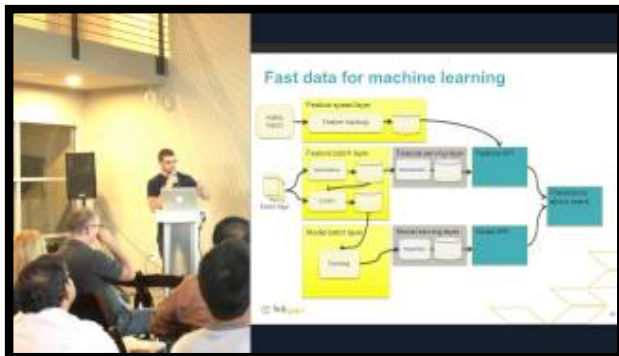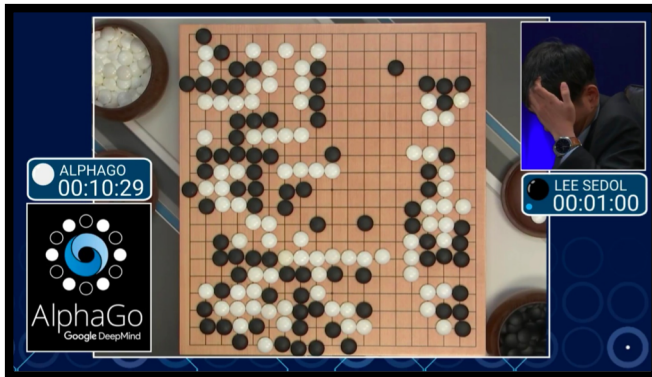
What is Machine Learning?

Fusing Disciplines
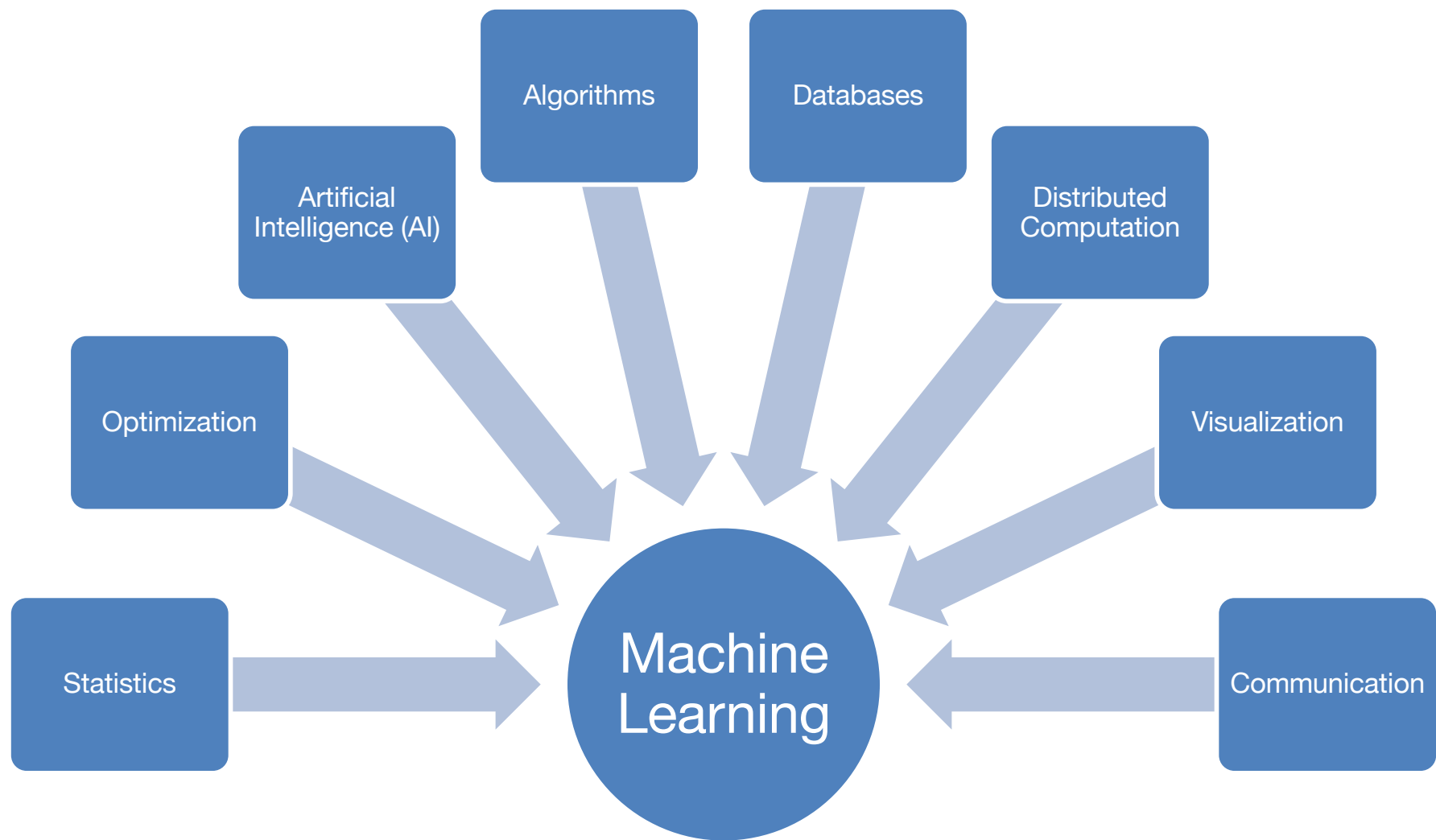
Machine Learning Pipeline

Collect Data ⬌ EDA/Preprocess ⬌ Learn! ⬌ Post Process ⬌ Deploy

API
Sensors
DBMS query

Association
Classification
Clustering
Outlier detection

Pipelining
Scaling

Integration
Normalization
Feature selection
Binning
Dimension red.

Evaluation
Interpretation
Visualization
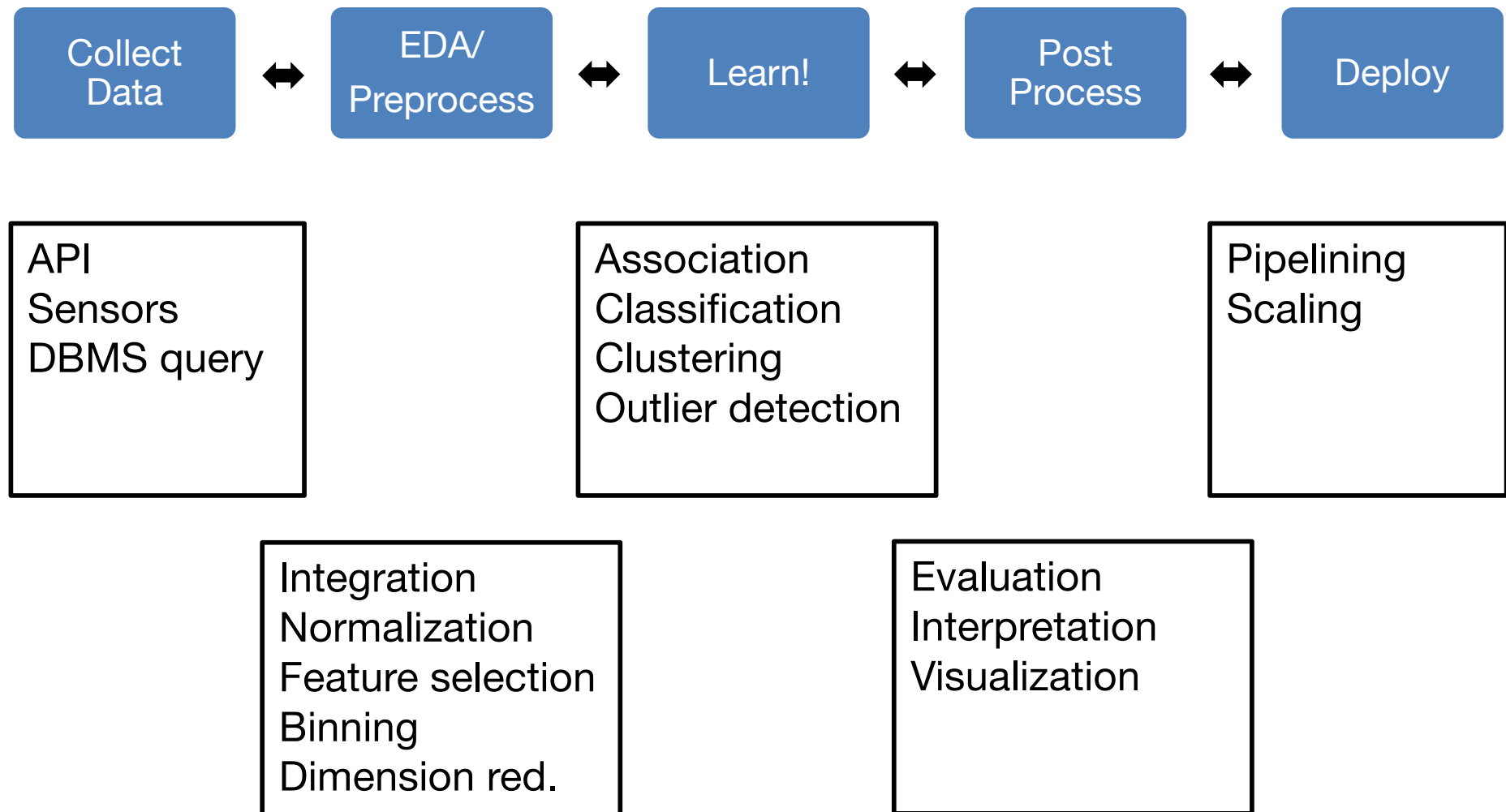
Jobs!

Position	Salary
Data Scientist	£45,000
Software Engineer	£41,700

The New York Times

TECHNOLOGY

A.I. Researchers Are Making More Than $1 Million, Even at a Nonprofit

By CADE METZ   APRIL 19, 2018

Tech Giants Are Paying Huge Salaries for Scarce A.I. Talent

Nearly all big tech compa... intelligence project, and tl... experts millions of dollar...

查看简体中文版 | 查看繁體中...

By CADE METZ   C...

Typical A.I. specialists, including both Ph.D.s fresh out of school and people with less education and just a few years of experience, can be paid from $300,000 to $500,000 a year or more in salary and company stock, according to nine people who work for major tech companies or have entertained job offers from them. All of them requested anonymity because they did not want to damage their professional prospects.

"A data scientist is someone who knows more statistics than a computer scientist and more computer science than a statistician."
– Josh Blumenstock (UW)

"Data Scientist = statistician + programmer + coach + storyteller + artist"
– Shlomo Aragmon (Ill. Inst. of Tech)

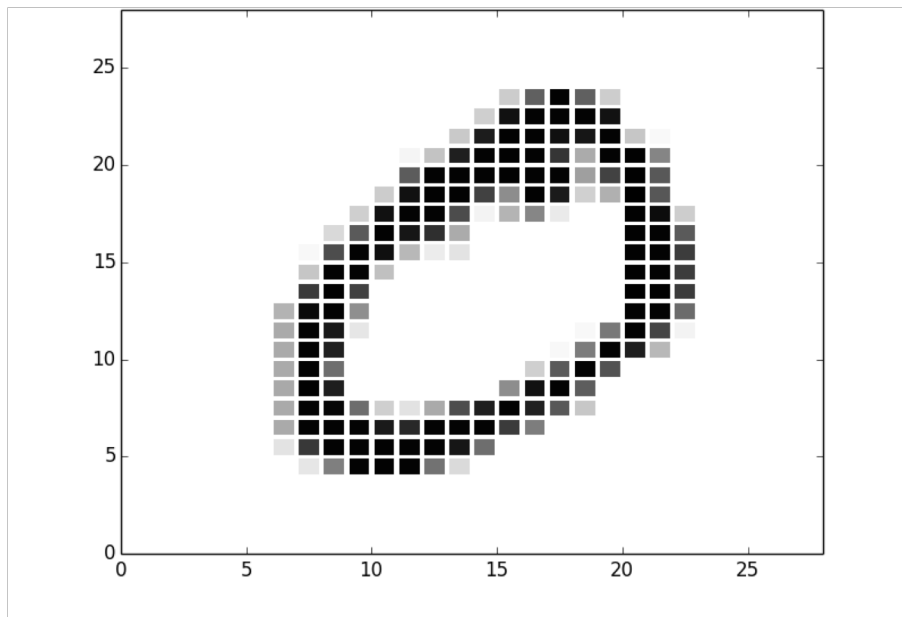"Software Is Eating the World, but AI Is Going to Eat Software"
– Jensen Huang (CEO, NVIDIA)

*indeed.co.uk, London Avg as of March 4, 2019

What is Machine Learning?

Common Terminology

example, instance

Unit of input
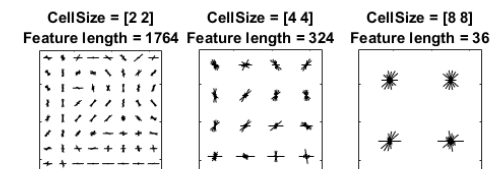
Composed of *features*
(or *attributes*)

- In this case, we could represent each digit via raw pixels: 28x28=784-pixel *vector* of greyscale values [0-255]
  - *Dimensionality*: number of features per instance (|vector|)

- But other *data representations* are possible, and might be advantageous

- In general, the problem of *feature selection* is challenging

What is Machine Learning?

Instances/Features = Table

Features

Outlook	Temperature	Humidity	Windy	Play
sunny	85	85	false	no
sunny	80	90	true	no
overcast	83	86	false	yes
rainy	70	96	false	yes
rainy	68	80	false	yes
rainy	65	70	true	no
overcast	64	65	true	yes
sunny	72	95	false	no
sunny	69	70	false	yes
rainy	75	80	false	yes
sunny	75	70	true	yes
overcast	72	90	true	yes
overcast	81	75	false	yes
rainy	71	91	true	no

Instance

What is Machine Learning?

"Target" Feature

When trying to predict a particular feature given the others

target, ***label***, ***class***, ***concept, dependent***

| Outlook	Temperature	Humidity	Windy	Play
sunny	85	85	false	no
sunny	80	90	true	no
overcast	83	86	false	yes
rainy	70	96	false	yes
rainy	68	80	false	yes
rainy	65	70	true	no
overcast	64	65	true	yes
sunny	72	95	false	no
sunny	69	70	false	yes
rainy	75	80	false	yes
sunny	75	70	true	yes
overcast	72	90	true	yes
overcast	81	75	false	yes
rainy	71	91	true	no

What is Machine Learning?

Machine Learning Tasks

- ## *Supervised*
  - Given a dataset of input-output pairs, learn a function that maps future (novel) inputs to appropriate outputs

- ## *Unsupervised*
  - Given a dataset and a hypothesis, find interesting patterns/parameters

- ## *Reinforcement*
  - Learn an optional action *policy* over time; given an environment that provides states, affords actions, and provides feedback as numerical *reward*, maximize the *expected* future reward

What is Machine Learning?

Supervised Learning (1)
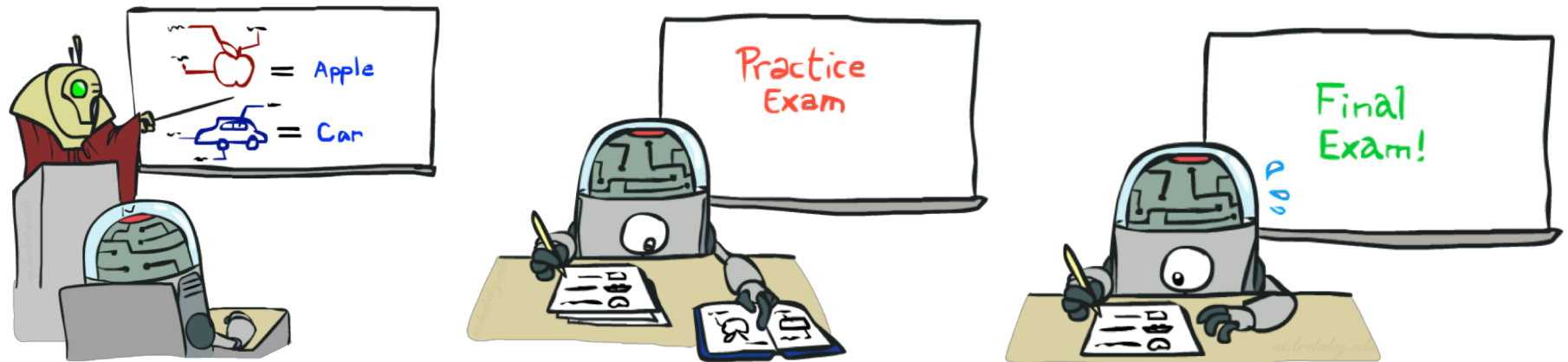
Supervised Learning (2)

Training Set   **_Testing Set_**

α   β   β   γ   ...   ?
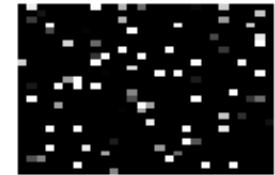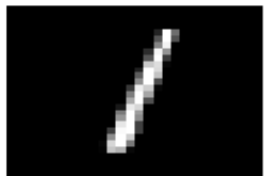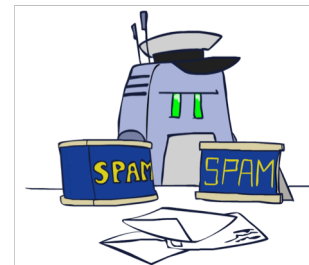

What is Machine Learning?

Supervised Tasks (1)

Classification

- Discrete target

- Binary vs. multi-class

SepalLength	SepalWidth	PetalLength	PetalWidth	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa

What is Machine Learning?

Supervised Tasks (2)

Regression

- Continuous target

| mpg	cylinders	displacement	horsepower	weight	acceleration	model year	origin	car name
18	8	307	130	3504	12	70	1	chevrolet chevelle malibu
15	8	350	165	3693	11.5	70	1	buick skylark 320
18	8	318	150	3436	11	70	1	plymouth satellite
16	8	304	150	3433	12	70	1	amc rebel sst
17	8	302	140	3449	10.5	70	1	ford torino
15	8	429	198	4341	10	70	1	ford galaxie 500
14	8	454	220	4354	9	70	1	chevrolet impala
14	8	440	215	4312	8.5	70	1	plymouth fury iii
14	8	455	225	4425	10	70	1	pontiac catalina
15	8	390	190	3850	8.5	70	1	amc ambassador dpl
15	8	383	170	3563	10	70	1	dodge challenger se
14	8	340	160	3609	8	70	1	plymouth 'cuda 340
15	8	400	150	3761	9.5	70	1	chevrolet monte carlo
14	8	455	225	3086	10	70	1	buick estate wagon (sw)
24	4	113	95	2372	15	70	3	toyota corona mark ii
22	6	198	95	2833	15.5	70	1	plymouth duster
18	6	199	97	2774	15.5	70	1	amc hornet
21	6	200	85	2587	16	70	1	ford maverick
27	4	97	88	2130	14.5	70	3	datsun pl510
26	4	97	46	1835	20.5	70	2	volkswagen 1131 deluxe sedan
25	4	110	87	2672	17.5	70	2	peugeot 504
24	4	107	90	2430	14.5	70	2	audi 100 ls
25	4	104	95	2375	17.5	70	2	saab 99e
26	4	121	113	2234	12.5	70	2	bmw 2002

What is Machine Learning?

Under/Over-fitting

Underfitting: the model does not capture the important relationship(s)

Overfitting: the model describes noise instead of the underlying relationship

Approaches
- ***Regularization***
- Robust evaluation
  – Cross validation
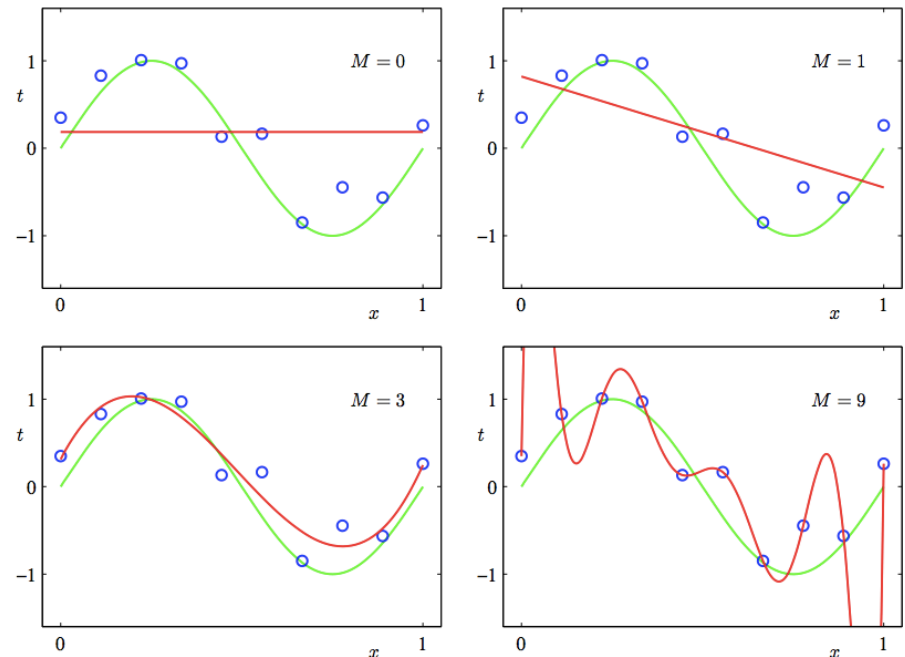
Validation Set

- One approach in an ML-application pipeline is to use a ***validation*** dataset (could be a ***holdout*** from the training set)

- Each model is built using just training; the validation dataset is then used to compare performance and/or select model parameters

- But still, the final performance is only measured via an independent test set

What is Machine Learning?

More Training Data = Better

- In general, the greater the amount of training data, the better we expect the learning algorithm to perform
  - But we also want reasonable amounts of validation/testing data!

- So how do we not delude ourselves, achieve high performance, *and* a reasonable expectation of future performance?

What is Machine Learning?

k-Fold Cross Validation
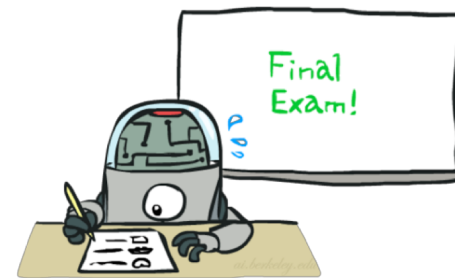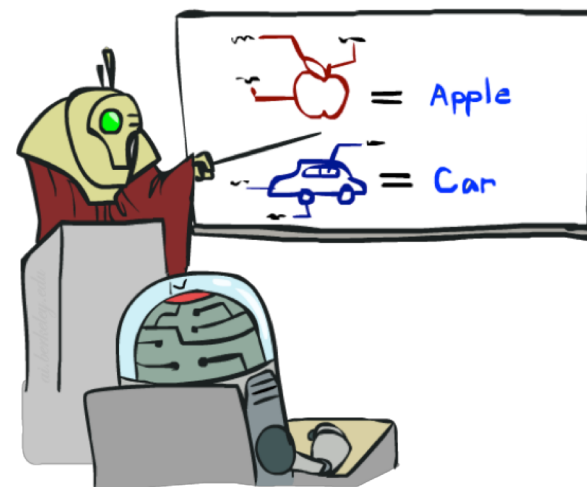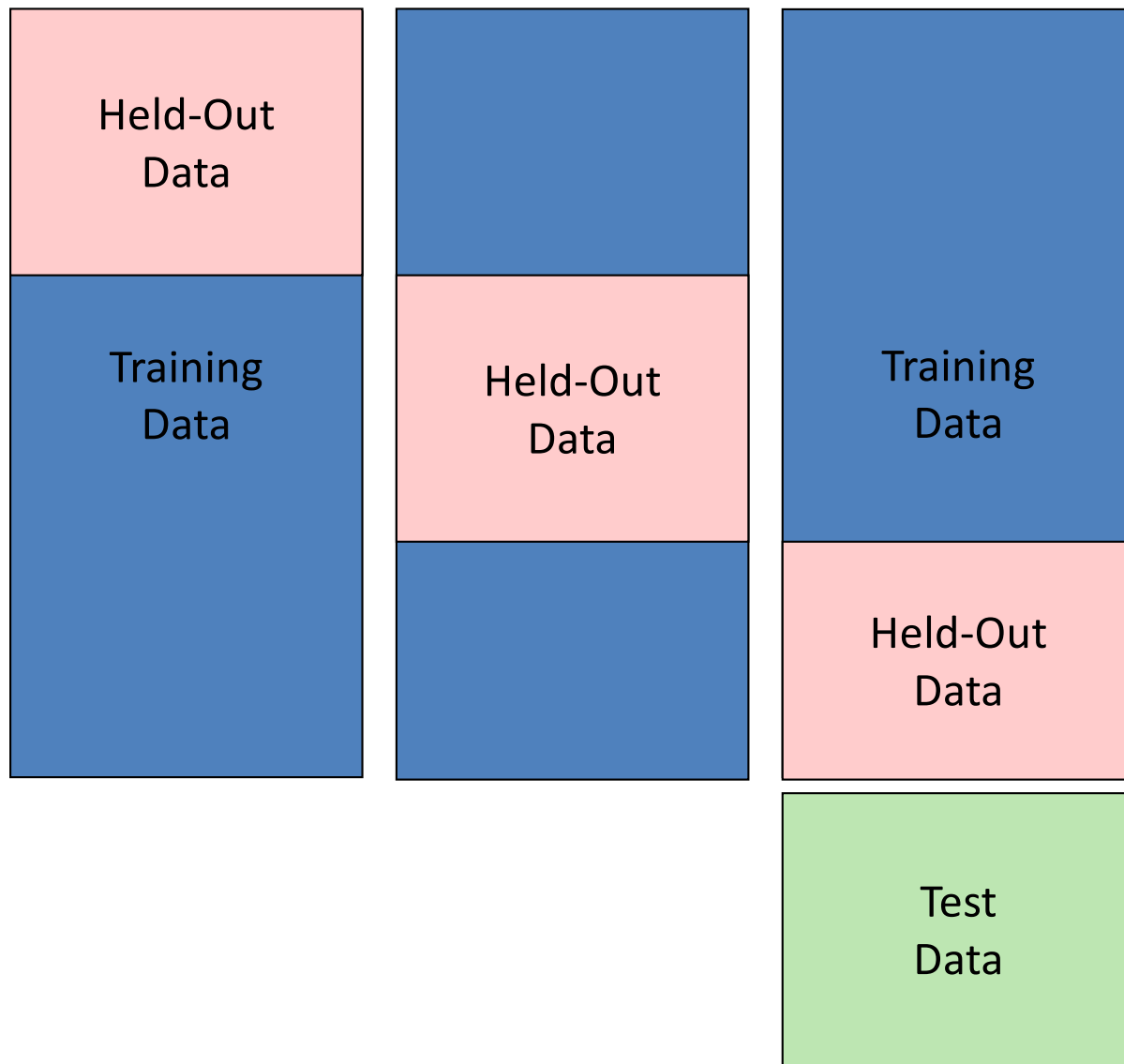
Common Algorithms

- Instance-based
  - **Nearest Neighbor (kNN)**

- Tree-based
  - ID3, C4.5, Random Forests

- Optimization-based
  - **Linear regression**, logistic regression, support vector machines (SVM)

- Probabilistic
  - Naïve Bayes, HMM

- Artificial Neural Networks
  - Backpropagation
  - Deep learning

What is Machine Learning?

kNN

Training

- Store all examples

Testing

- Find the nearest k neighbors to target
  - Via distance function
- Vote on class

Non-parametric algorithm
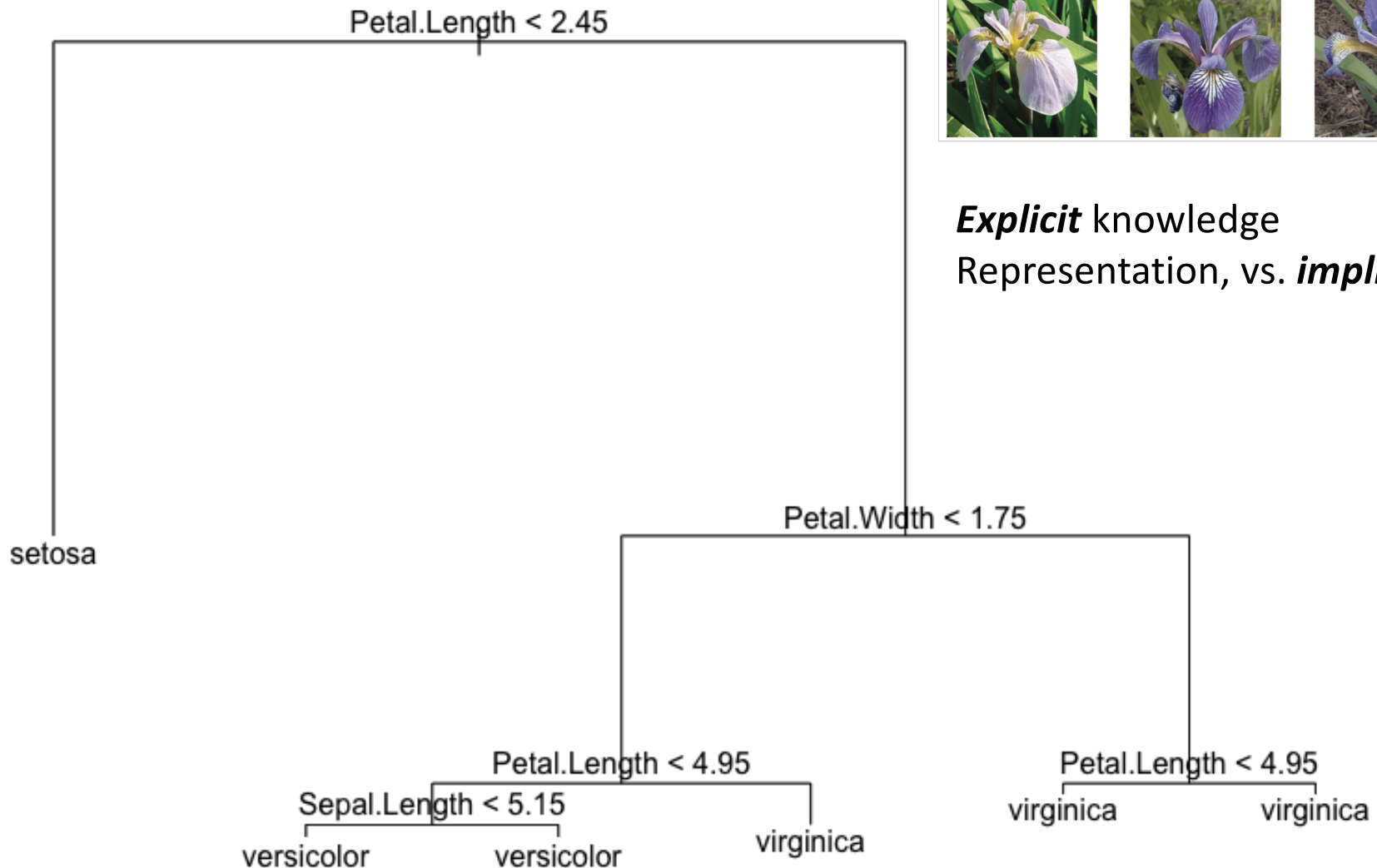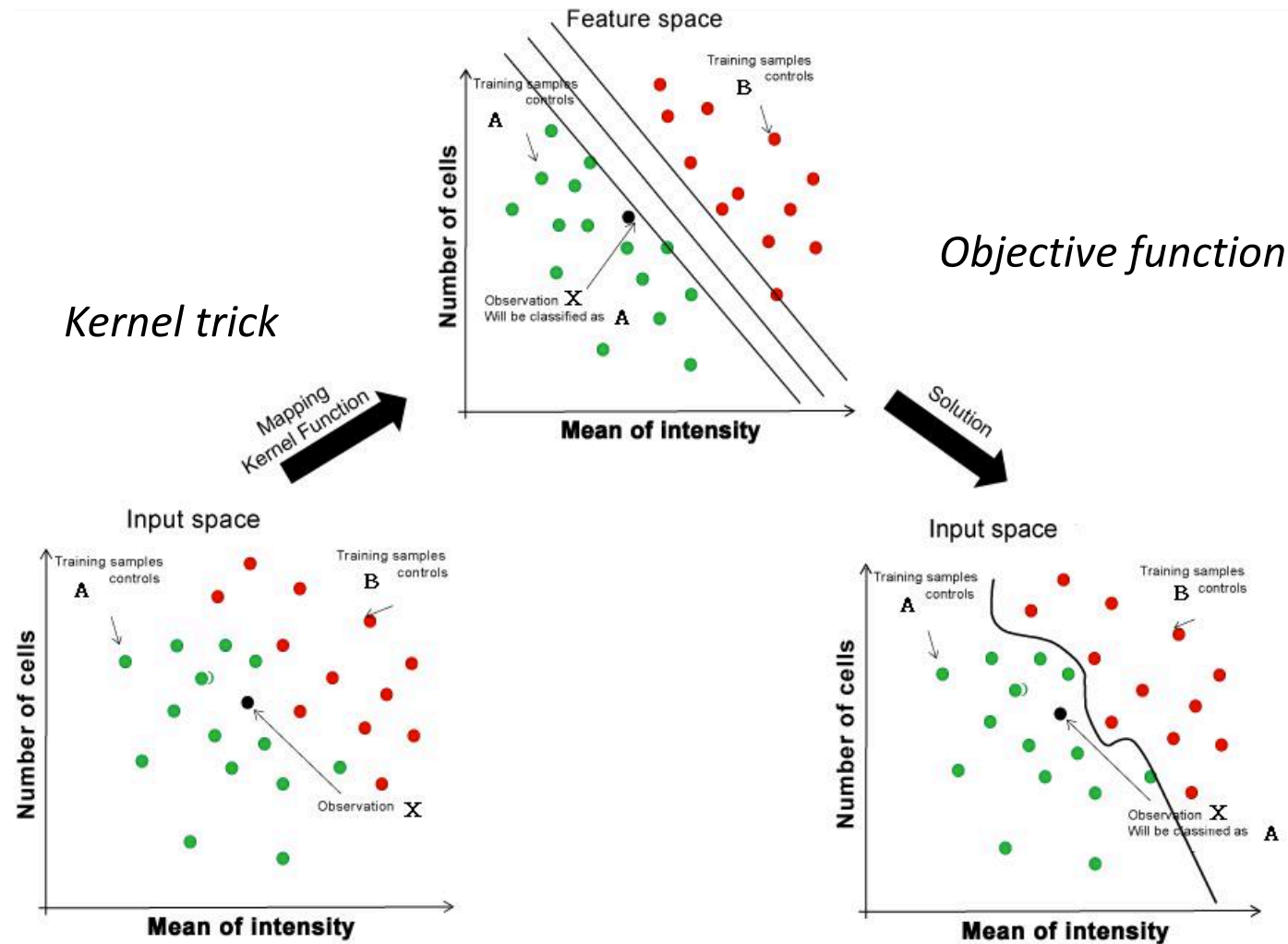(i.e. grows with |examples|!)
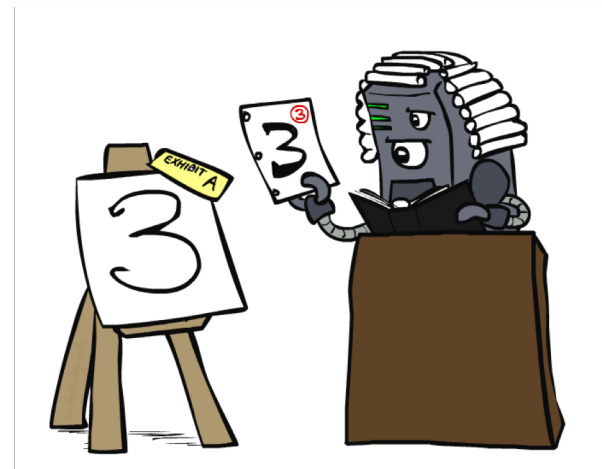
2D Multiclass Classification

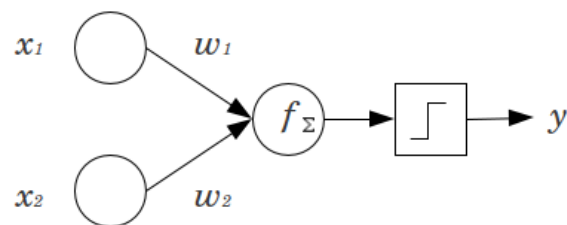## Boundary Tree

1-NN via Linear Scan

What is Machine Learning?

Decision Trees/Forests

Petal.Length < 2.45

Explicit knowledge Representation, vs. *implicit*

setosa

Petal.Width < 1.75

Petal.Length < 4.95

Sepal.Length < 5.15

versicolor   versicolor   virginica

Petal.Length < 4.95

virginica   virginica

What is Machine Learning?
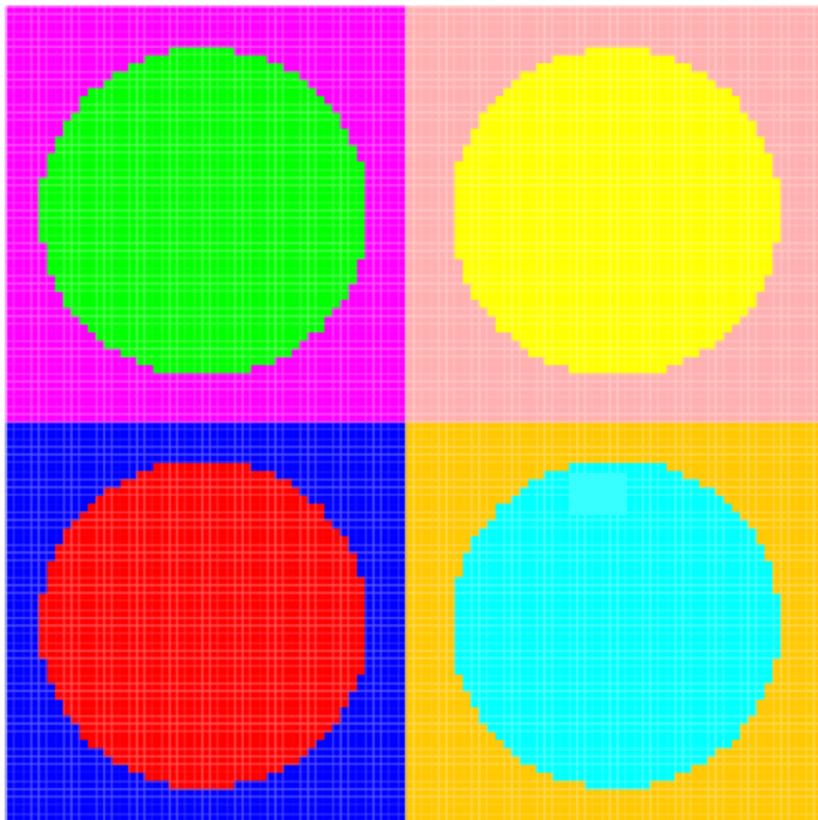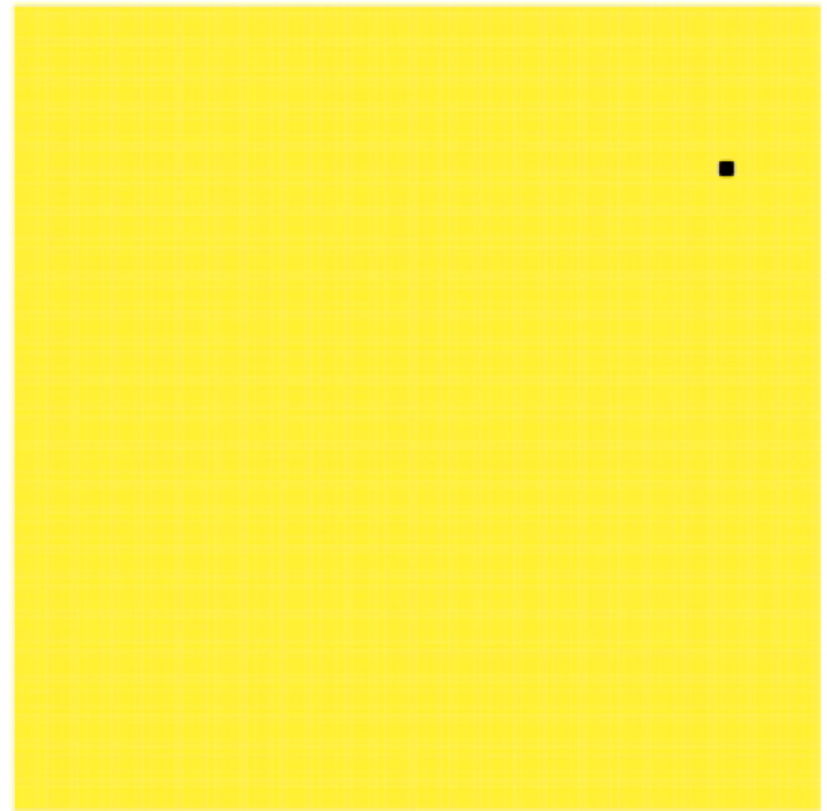
Support Vector Machine (SVM)

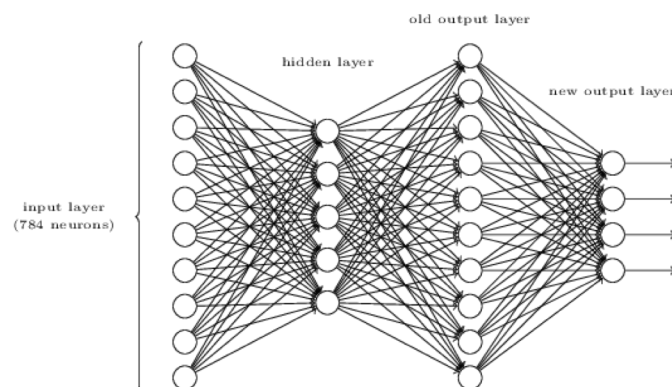*Kernel trick*

Objective function

Artificial Neural Networks (ANN)

Perceptron
Linear classifier
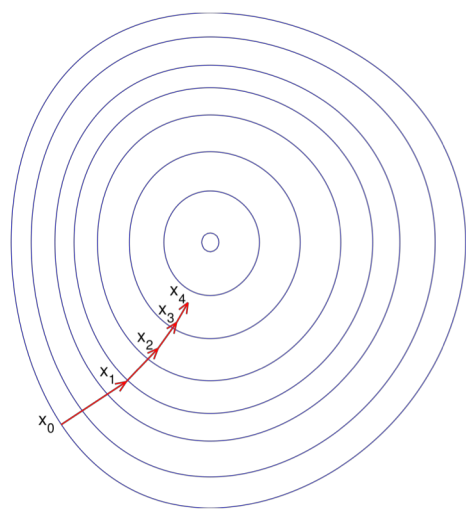
Backpropagation
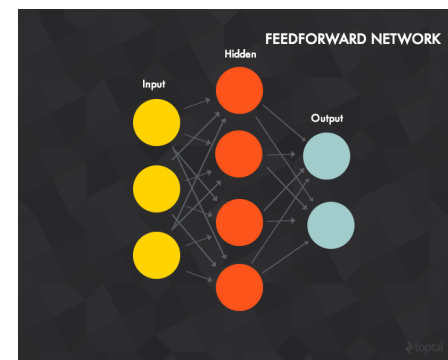
Feedforward vs. Recurrant

Gradient descent
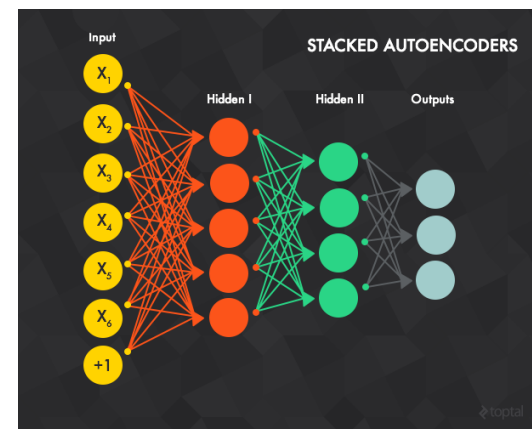
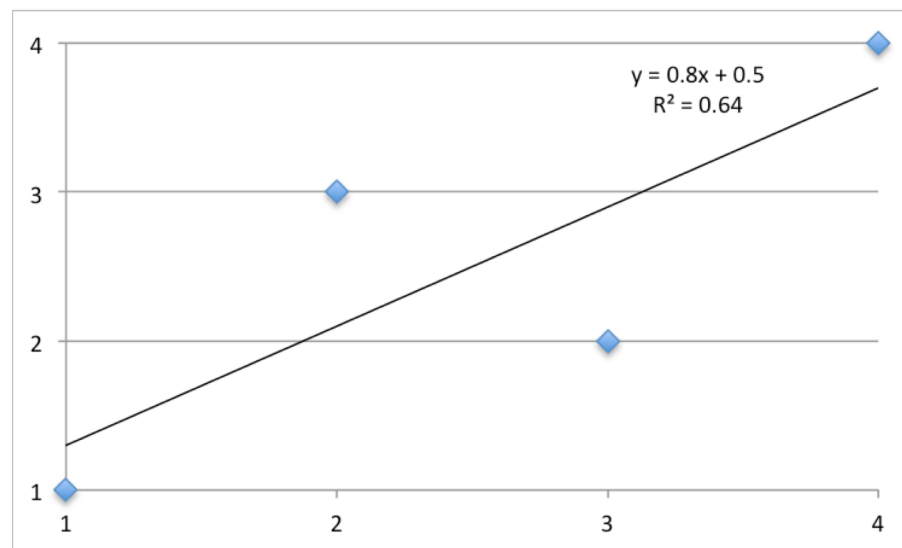Deep Architectures
Vanishing Gradient

What is Machine Learning?

Example: Linear Regression

Input

x	y
1	1
2	3
3	2
4	4

Output

$y = 0.8x + 0.5$
$R^2 = 0.64$
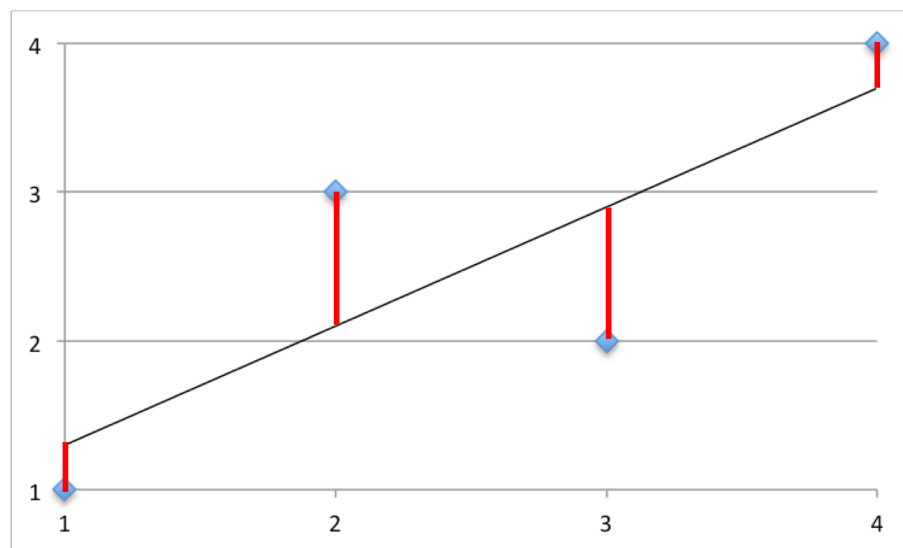
Linear Regression as Optimization

- Why this line?
  - Minimizer **error**

- In 2D, the algorithm tries to find a slope and intercept that yields the smallest sum of the square of the error (SSE)

$$\underset{m,b}{\arg\min} \sum_{i=1}^{N} e_i^2 = (y_i - (mx_i + b))^2$$

What is Machine Learning?

Machine Learning via Optimization

1. Define an error function

2. Find model parameters that minimize the error function given the data

   – Sometimes closed-form solution (e.g. linear)

   – Sometimes [iterative] solution [with guarantees] (e.g. convex)

   – Most of the time will require approximation

     • Iteration (limited by number, delta)

     • Softening constraints

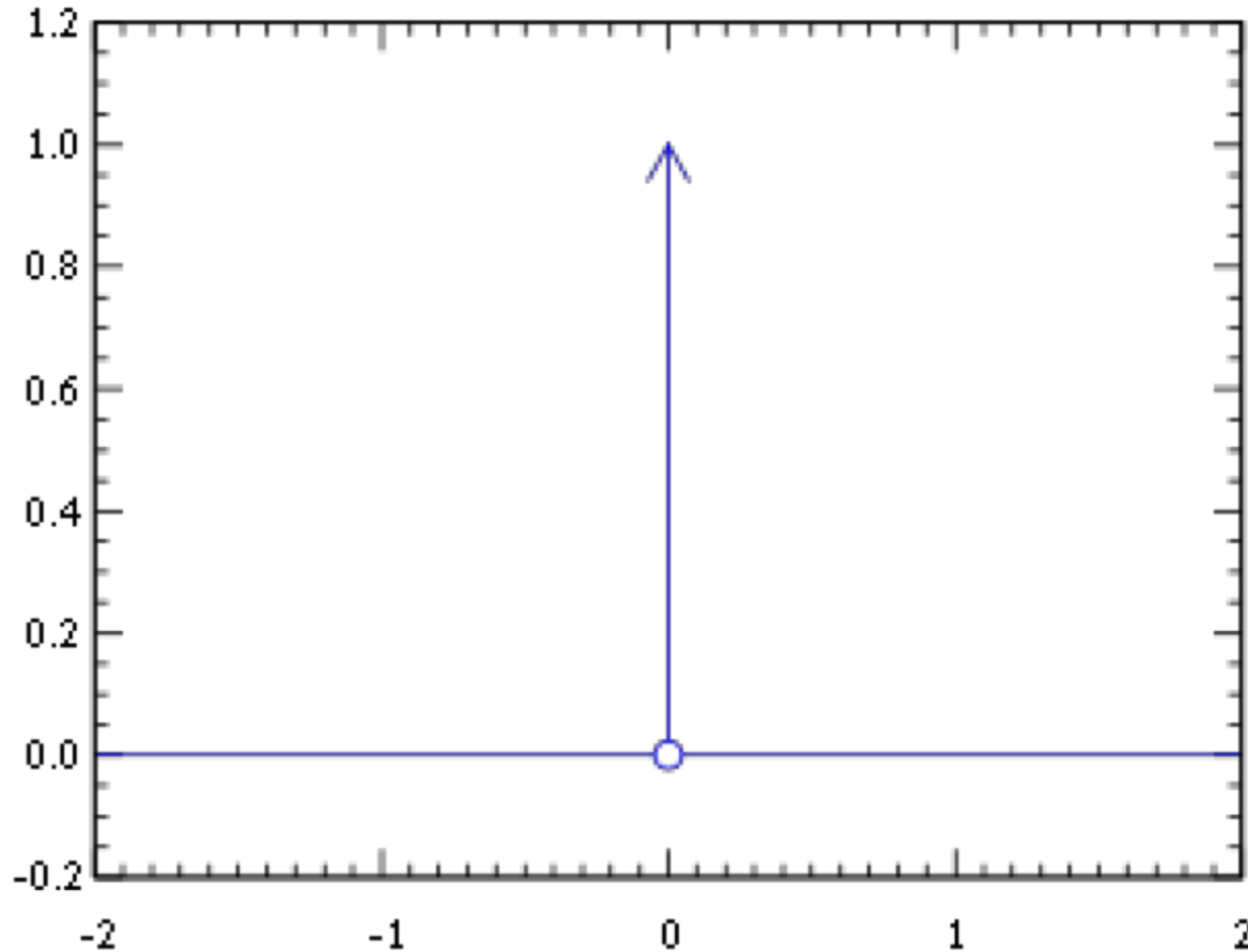   • Post-processing

     …

What is Machine Learning?

Optimization is Hard in General
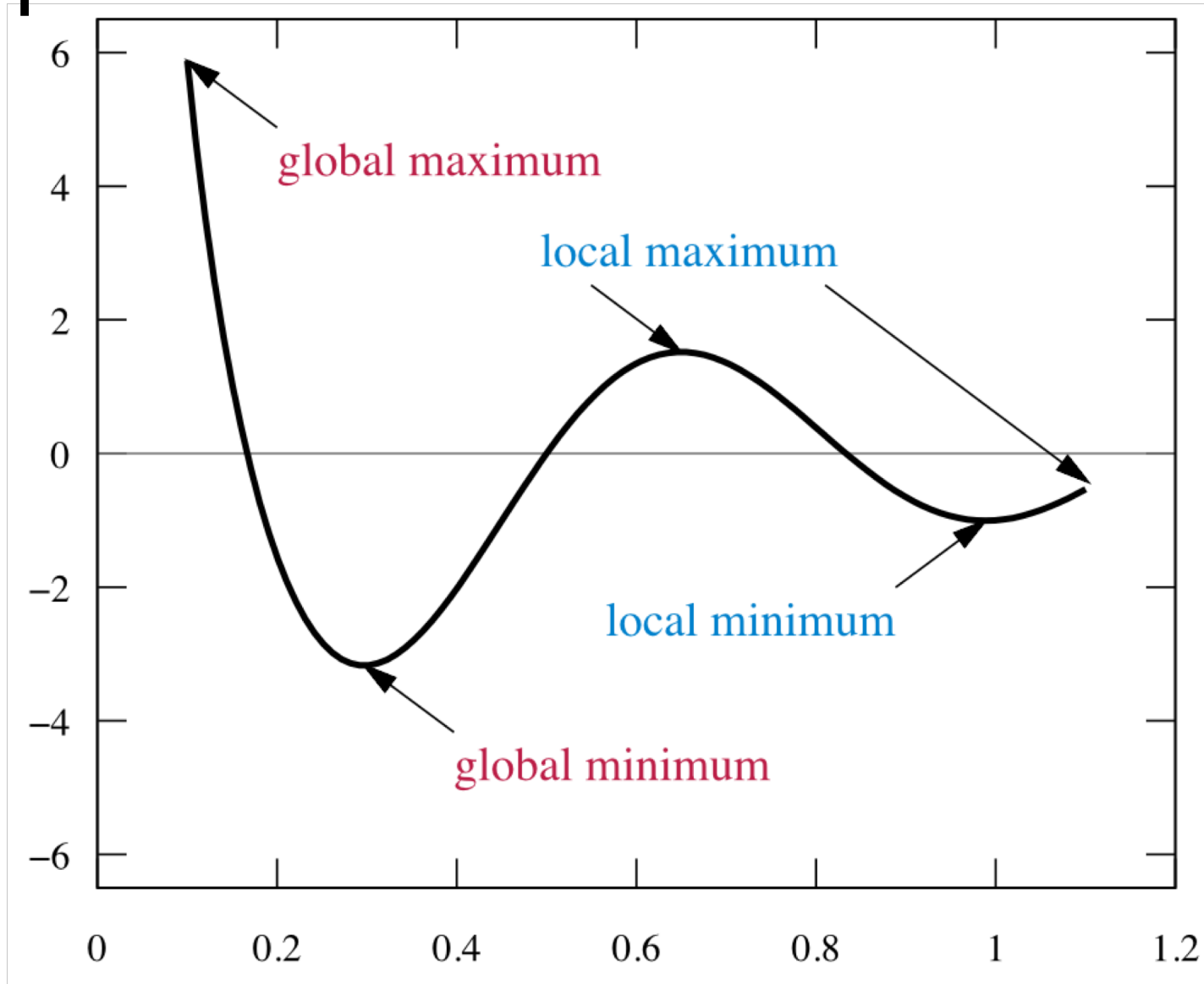
Consider Many [Cursed] Dimensions

What is Machine Learning?
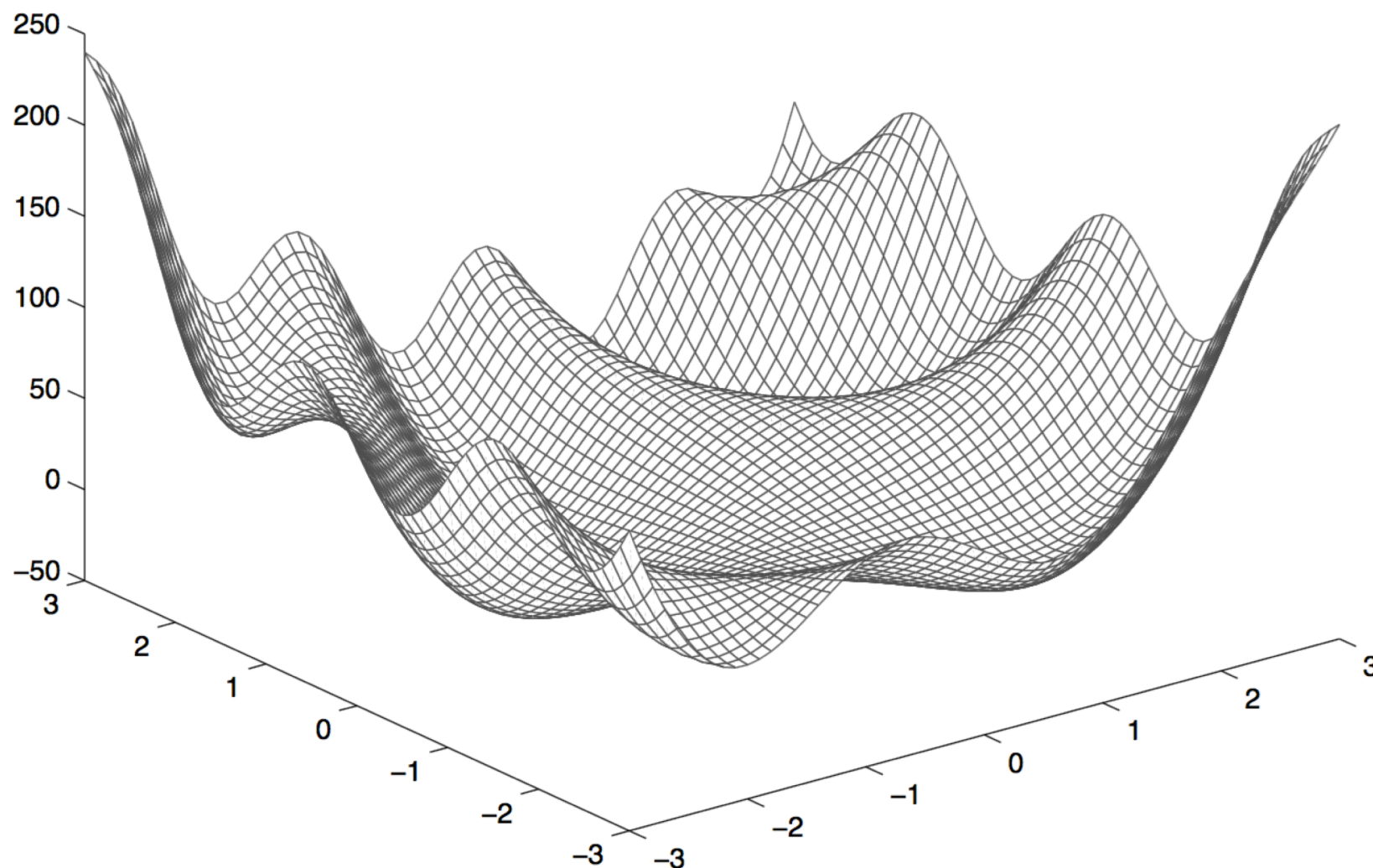
Consider Discontinuities

What is Machine Learning?

Linear Regression

Recipe

1. Define error function

2. Find parameter values that minimize error given the data
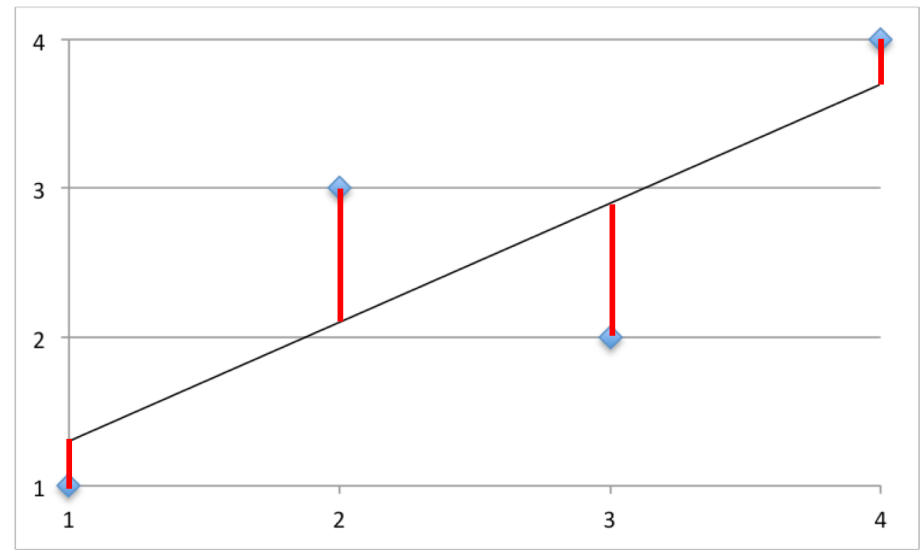

$$y = 0.8x + 0.5$$
$$R^2 = 0.64$$
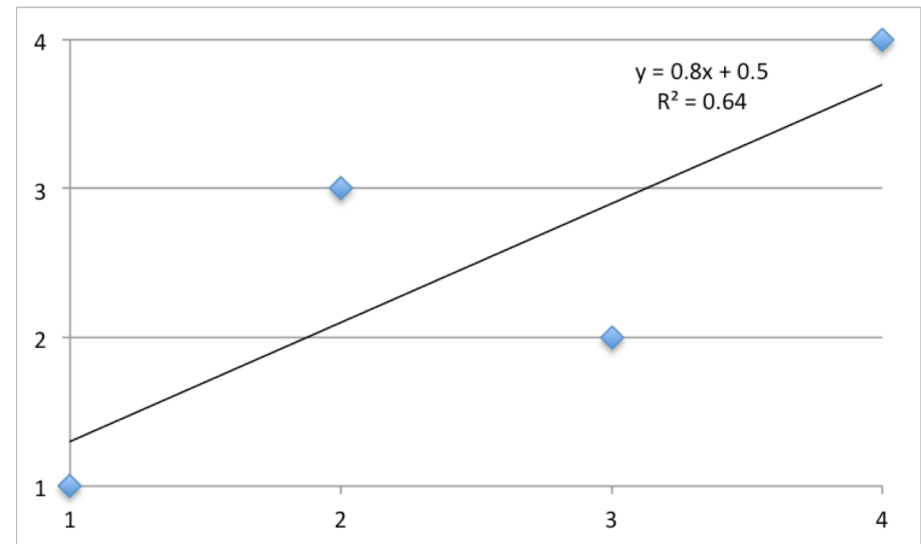
Error Function

Sum of Squared Error (SSE)
aka Residual Sum of Squares (RSS)

$$\mathrm{SSE}_{\mathrm{line}} = \sum_{i=1}^{N}(y_i - f(x_i)) = (y_i - (mx_i + b))^2$$

What is Machine Learning?

Algebra (1)

$$\mathrm{SSE}_{\mathrm{line}} = \sum_{i=1}^{N} (y_i - (mx_i + b))^2$$

$$= \sum_{i=1}^{N} y_i{}^2 - 2y_i(mx_i + b) + (mx_i + b)^2$$

$$= \sum_{i=1}^{N} y_i{}^2 - 2mx_iy_i - 2by_i + m^2x_i{}^2 + 2mbx_i + b^2$$

Algebra (2)

$$\sum_{i=1}^{N} a_i = N\overline{a}$$

so...

$$\text{SSE}_{\text{line}} = \sum_{i=1}^{N} y_i{}^2 - 2mx_iy_i - 2by_i + m^2x_i{}^2 + 2mbx_i + b^2$$

$$= N\overline{y^2} - 2Nm\overline{xy} - 2Nb\overline{y} + Nm^2\overline{x^2} + 2Nmb\overline{x} + Nb^2$$

What is Machine Learning?

Recall: Critical Points

- For a differentiable function of several variables, a critical point is a value in its domain where all partial derivatives are zero

- So to find the point at which error is minimized, we take partial derivatives of the error function w.r.t. the parameters, set these equal to 0, solve

What is Machine Learning?

Calculus (1)

$$\text{SSE}_{\text{line}} = N\overline{y^2} - 2Nm\overline{xy} - 2Nb\overline{y} + Nm^2\overline{x^2} + 2Nmb\overline{x} + Nb^2$$

$$\frac{\partial \text{SSE}_{\text{line}}}{\partial m} = -2N\overline{xy} + 2Nm\overline{x^2} + 2Nb\overline{x} = 0$$

$$\frac{\partial \text{SSE}_{\text{line}}}{\partial b} = -2N\overline{y} + 2Nm\overline{x} + 2Nb = 0$$

What is Machine Learning?

Algebra (3)

$$\frac{\partial \text{SSE}_{\text{line}}}{\partial b} = -2N\overline{y} + 2Nm\overline{x} + 2Nb = 0$$

$$0 = -2N\overline{y} + 2Nm\overline{x} + 2Nb$$

$$0 = -\overline{y} + m\overline{x} + b$$

$$\overline{y} = m\overline{x} + b \qquad\qquad (\overline{x}, \overline{y})$$

What is Machine Learning?

Algebra (4)

$$\frac{\partial \mathrm{SSE}_{\mathrm{line}}}{\partial m} = -2N\overline{xy} + 2Nm\overline{x^2} + 2Nb\overline{x} = 0$$

$$0 = -2N\overline{xy} + 2Nm\overline{x^2} + 2Nb\overline{x}$$

$$0 = -\overline{xy} + m\overline{x^2} + b\overline{x}$$

$$\overline{xy} = m\overline{x^2} + b\overline{x}$$

$$\frac{\overline{xy}}{\overline{x}} = m\frac{\overline{x^2}}{\overline{x}} + b$$

$$\left(\frac{\overline{x^2}}{\overline{x}}, \frac{\overline{xy}}{\overline{x}}\right)$$

What is Machine Learning?

And Finally...

$$(\overline{x}, \overline{y})$$

$$\left(\frac{\overline{x^2}}{\overline{x}}, \frac{\overline{xy}}{\overline{x}}\right)$$

$$m = \frac{\dfrac{\overline{xy}}{\overline{x}} - \overline{y}}{\dfrac{\overline{x^2}}{\overline{x}} - \overline{x}}$$

$$= \boxed{\frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - \overline{x}^2}}$$

$$\overline{y} = m\overline{x} + b$$

$$b = \boxed{\overline{y} - m\overline{x}}$$

What is Machine Learning?

The Multi-Dimensional Case

We begin with an analogous representation

$$Y = XB + e$$

where…

- **Y** is N x 1
- **X** is N x (k+1); extra 1 to multiply intercept
- **B** is (k+1) x 1; first intercept, then coefficients
- **e** is N x 1

What is Machine Learning?

k-Dimensional Linear Regression

$$
\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdot & \cdot & \cdot & x_{1k} \\ 1 & x_{21} & x_{22} & \cdot & \cdot & \cdot & x_{2k} \\ \cdot & \cdot & & & & & \\ \cdot & \cdot & & & & & \\ \cdot & \cdot & & & & & \\ 1 & x_{N1} & x_{N2} & \cdot & \cdot & \cdot & x_{Nk} \end{bmatrix} \begin{bmatrix} b \\ m_1 \\ m_2 \\ \cdot \\ \cdot \\ \cdot \\ m_k \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ e_N \end{bmatrix}
$$

What is Machine Learning?

Step 1: Error Function

- We will use the same error method as last time, which is SSE (i.e. square the difference between Y and XB)

$$\text{SSE} = e^{\mathsf{T}} e$$

$$= (Y - XB)^{\mathsf{T}}(Y - XB)$$

What is Machine Learning?

Matrix Algebra (1)

$$\begin{aligned}
\text{SSE} &= (Y - XB)^{\intercal}(Y - XB) \\
&= (Y^{\intercal} - B^{\intercal}X^{\intercal})(Y - XB) \\
&= Y^{\intercal}Y - Y^{\intercal}XB - B^{\intercal}X^{\intercal}Y + B^{\intercal}X^{\intercal}XB \\
&= Y^{\intercal}Y - 2Y^{\intercal}XB + B^{\intercal}X^{\intercal}XB
\end{aligned}$$

What is Machine Learning?

Matrix Calculus (1)

$$\text{SSE} = Y^{\mathsf{T}} Y - 2 Y^{\mathsf{T}} X B + B^{\mathsf{T}} X^{\mathsf{T}} X B$$

$$\frac{\partial \text{SSE}}{\partial B} = -2 X^{\mathsf{T}} Y + 2 X^{\mathsf{T}} X B$$

What is Machine Learning?

Matrix Algebra (2)

$$0 = -2X^\intercal Y + 2X^\intercal X B$$

$$-2X^\intercal X B = -2X^\intercal Y$$

$$X^\intercal X B = X^\intercal Y$$

$$B = \boxed{(X^\intercal X)^{-1} X^\intercal Y}$$

What is Machine Learning?

Unsupervised Learning

Find structure or patterns in data

<u>Tasks</u>

- Clustering

- Dimensionality reduction

- Density estimation

- Discovering graph structure

- Matrix completion

…

What is Machine Learning?

Common Algorithms

- k-Means Clustering

- Collaborative Filtering

- Principle Component Analysis (PCA)

k-Means Clustering (1)

- Pick K random points as cluster centers (means)

- Alternate:

  - Assign data instances to closest mean

  - Assign each mean to the average of its assigned points

- Stop when no points' assignments change
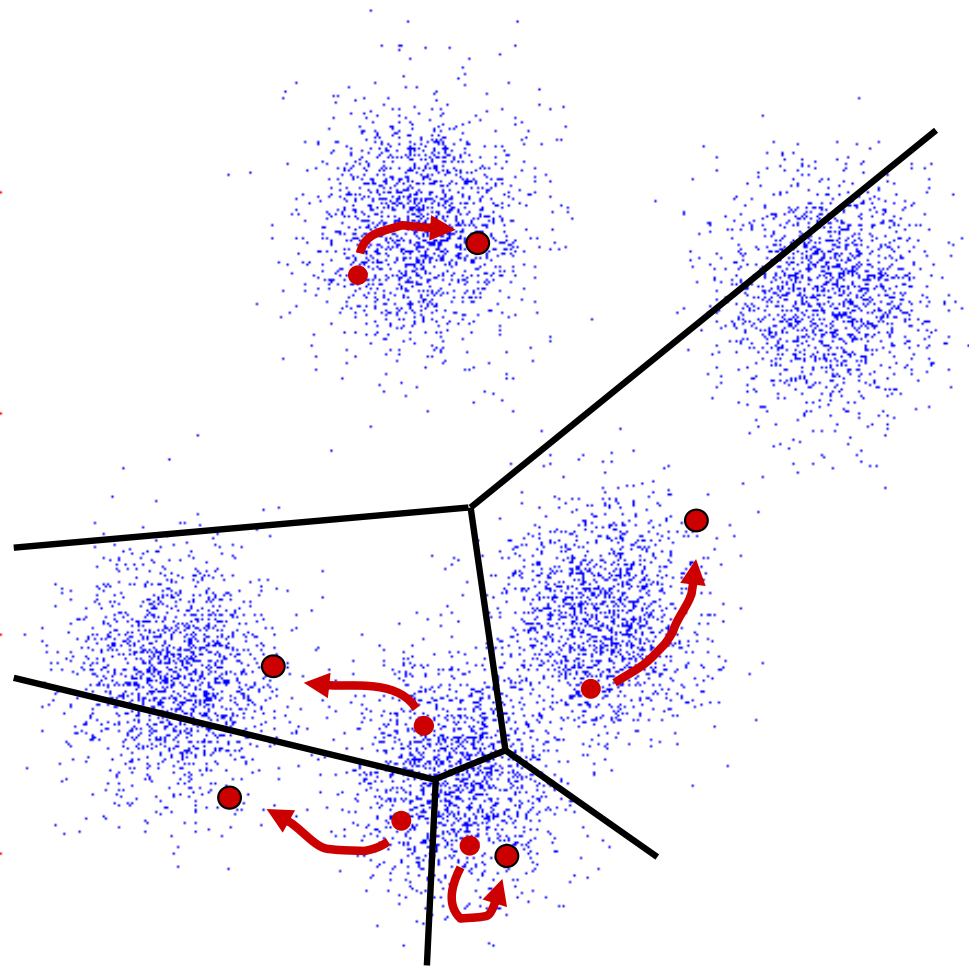
k-Means Clustering (2)
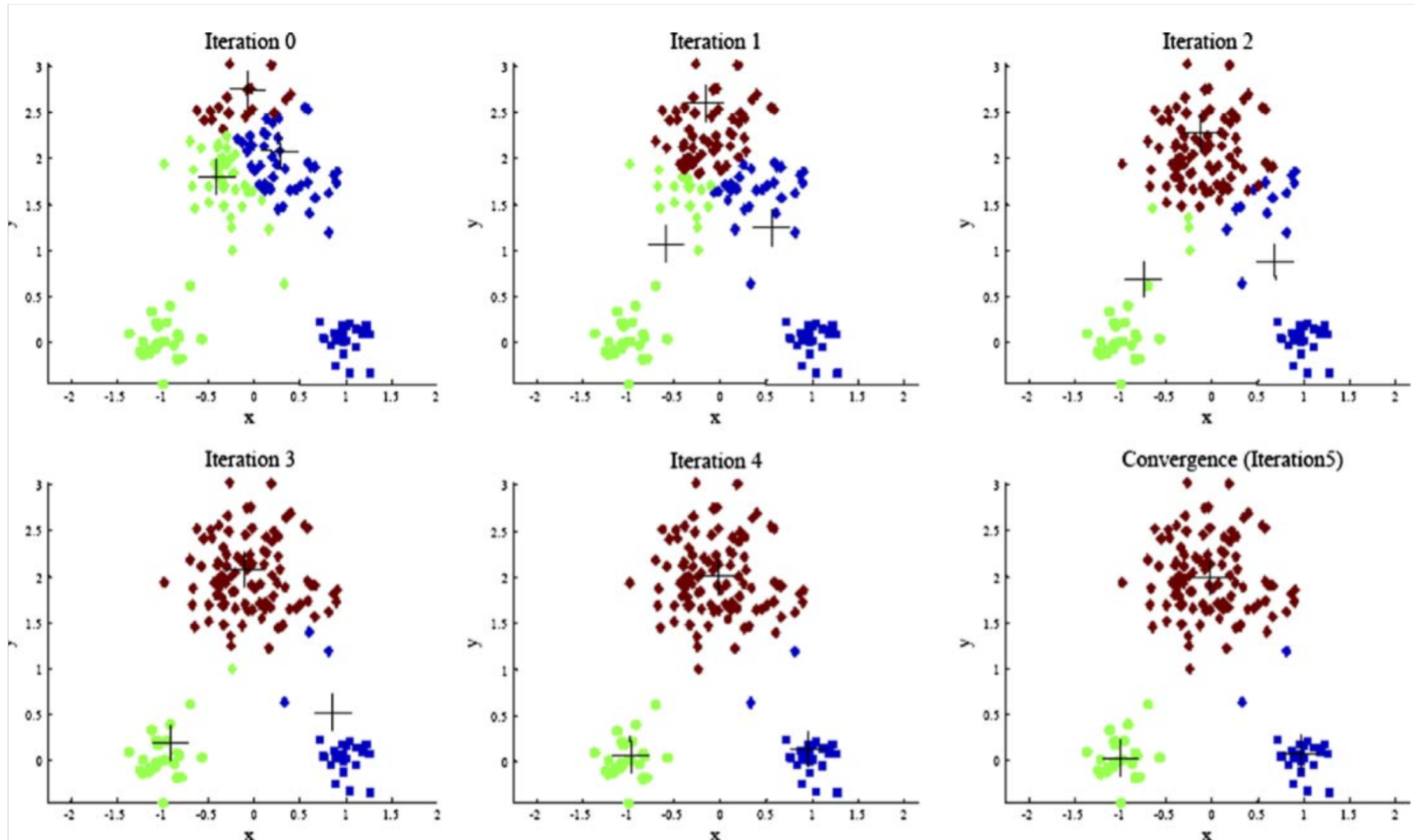
ML ala XKCD

What is Machine Learning?

What Makes for a "Good" Clustering?
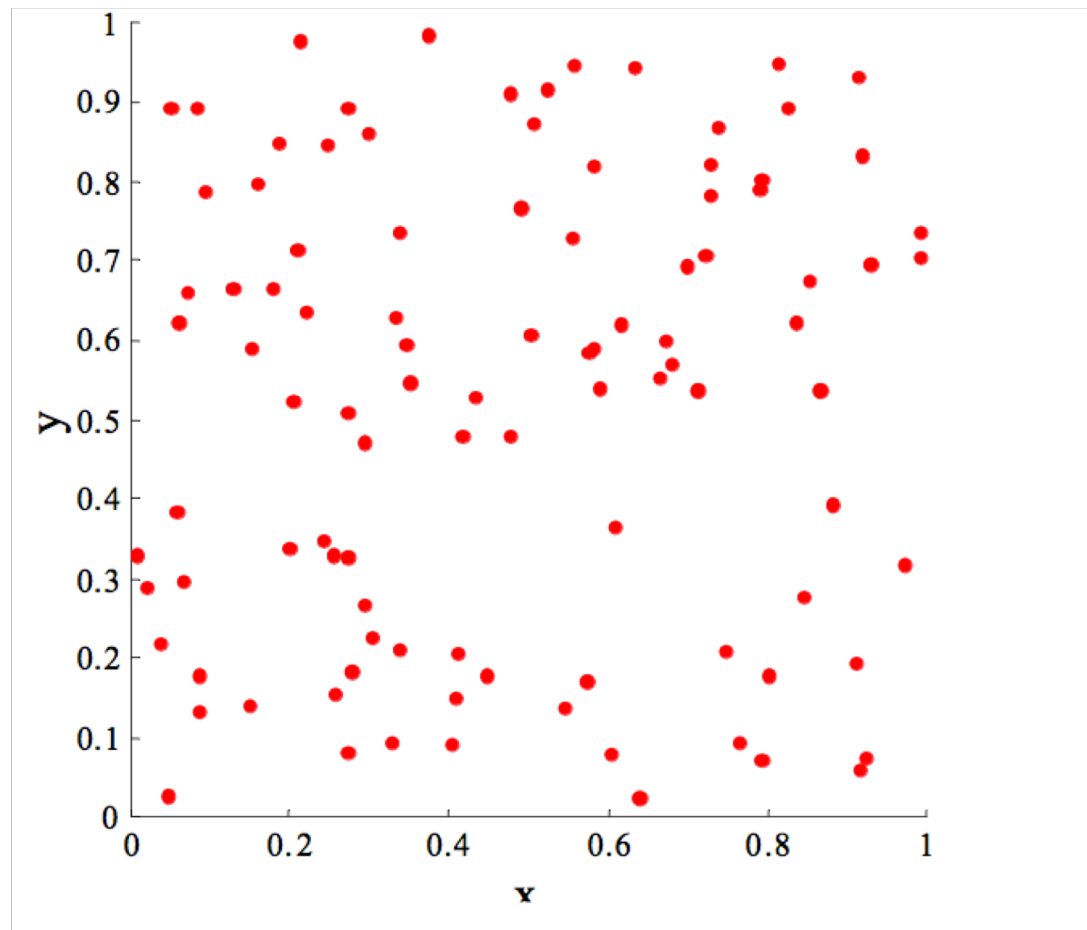
Did I Cluster *Well*?
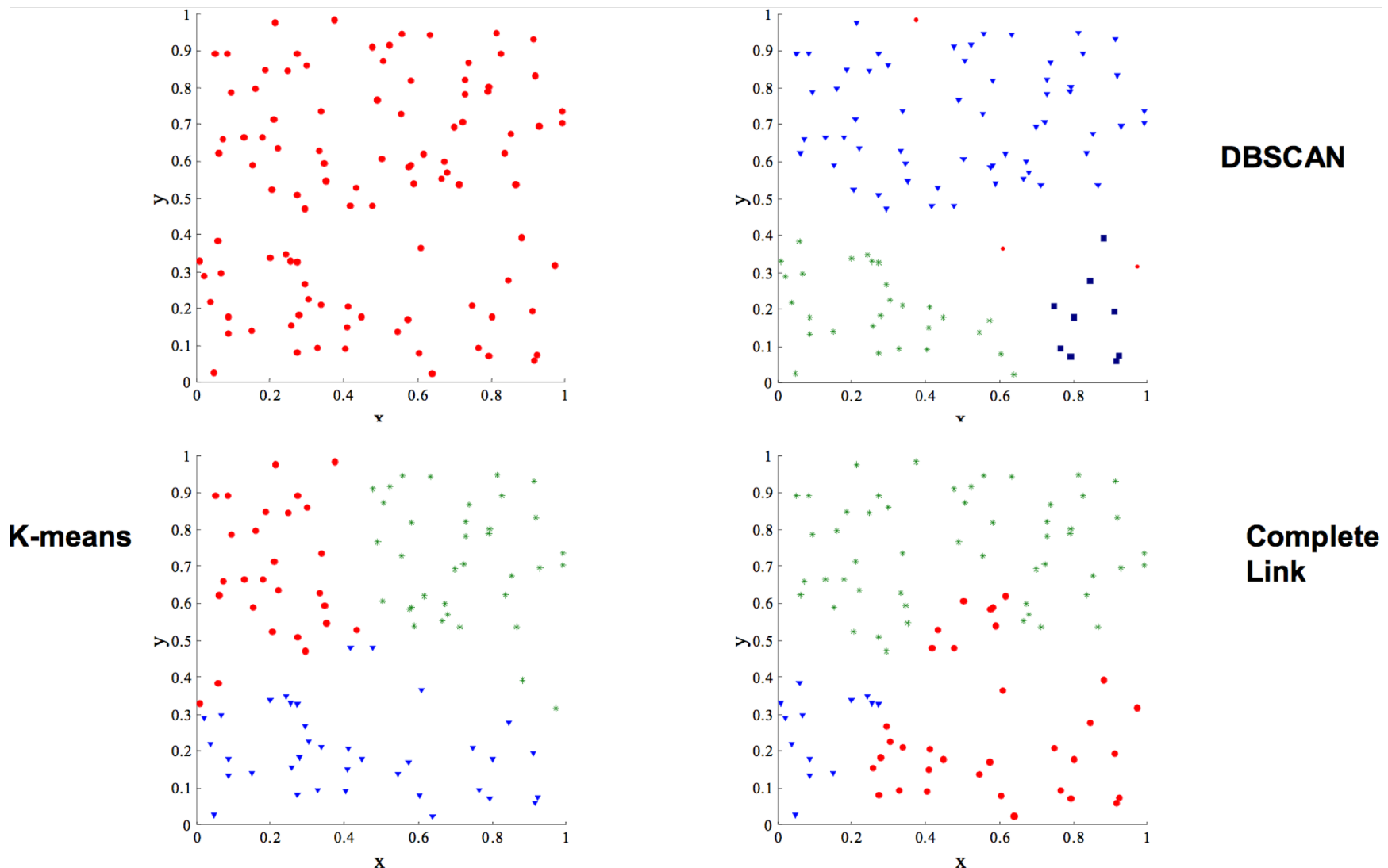
Did I Learn *Safely & Ethically*?
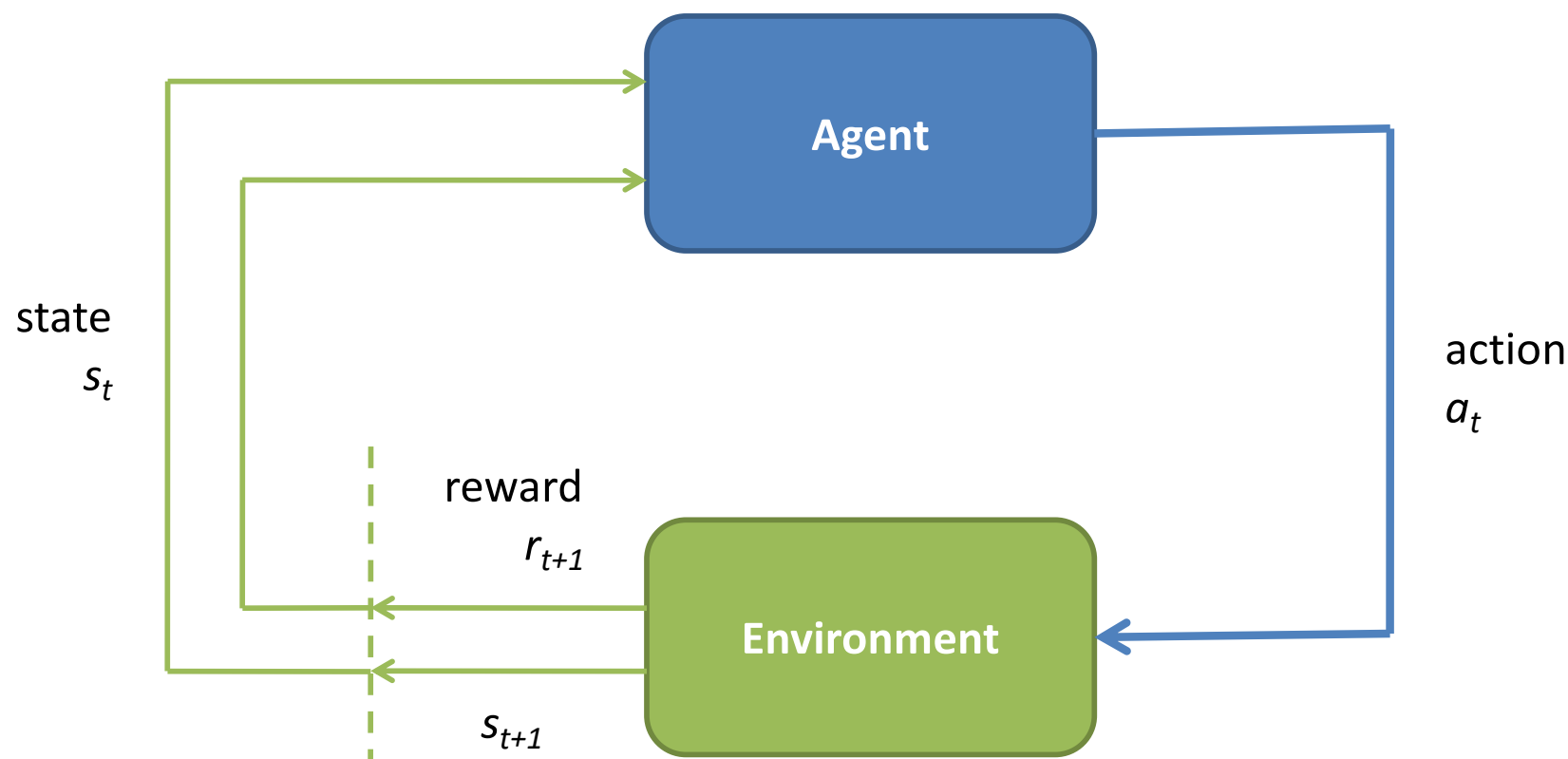

Random Points
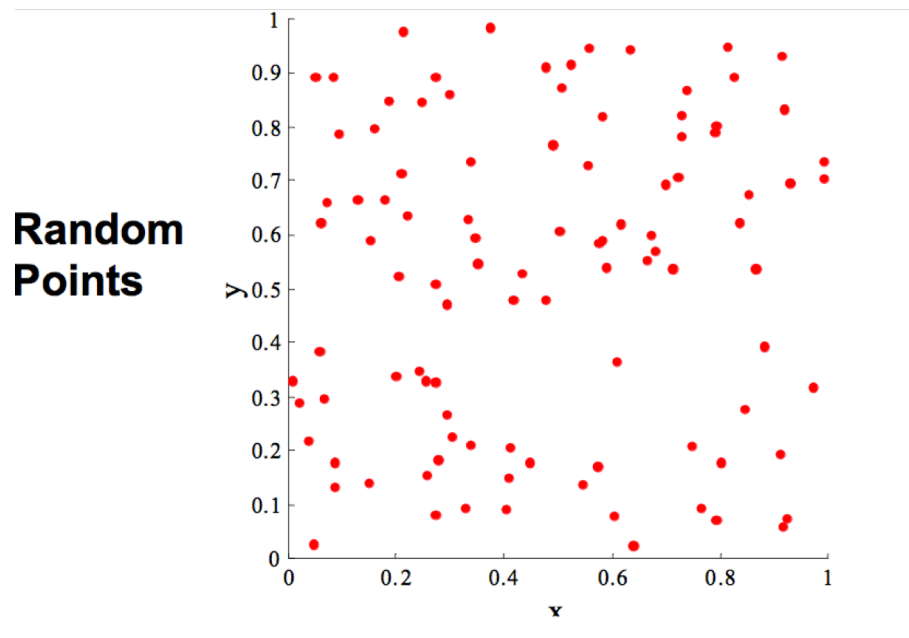
- 👵👴 VS 👶

- 🚐 VS 🚲 VS 🚶

Reinforcement Learning (RL)

Choose actions to maximize future reward

The RL Cycle

Issues. credit assignment, exploration vs. exploitation, reward function, …

state
s_t

action
a_t

reward
r_{t+1}

s_{t+1}

Agent

Environment

Temporal Difference (TD) Learning

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_{t+1} + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)]$$
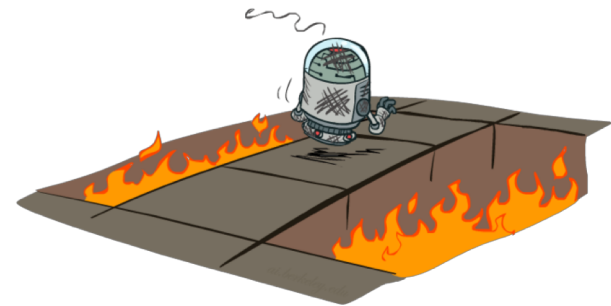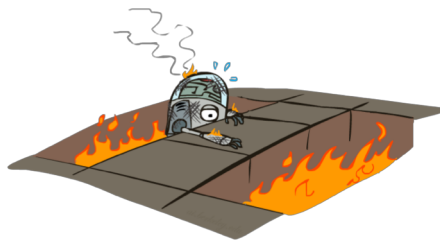
- Evidence that some neurons (dopamine) operate similarly
- Lead to world-class play via TD-Gammon (neural network trained via TD-learning)

Checkup

1. Build an Atari system that learns game-winning techniques via actually playing and adjusting actions based upon score changes

2. Given a dataset of past credit-card transactions (known to be fraudulent or not), build a system to identify future fraud

3. If we assume incoming CS1 students are bi-modal, but normally distributed, find the average grades of the two groups

1. RL

2. Supervised
   – Classification
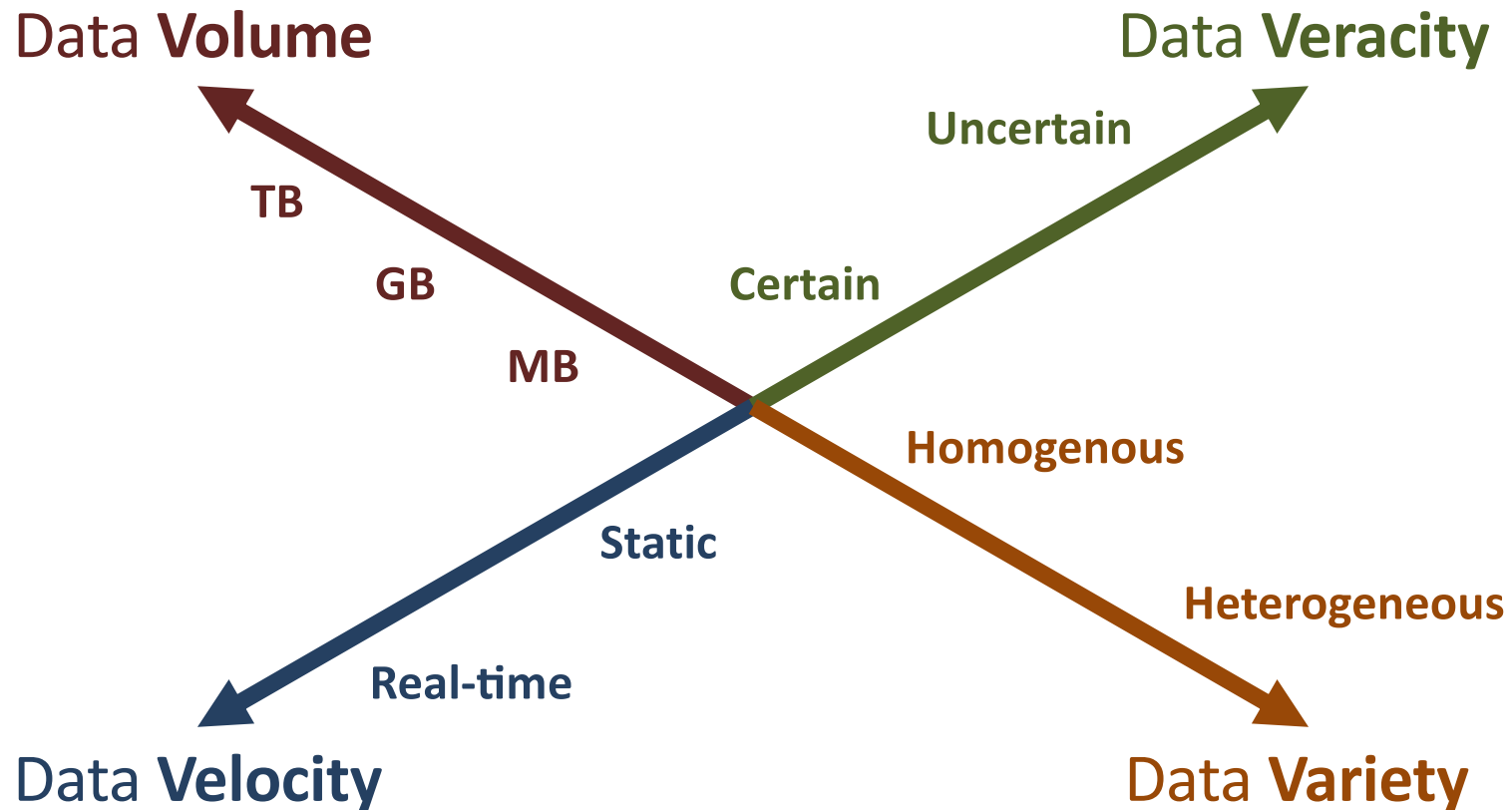
3. Unsupervised
   – Parameter estimation

What is Machine Learning?

Issues/Challenges

- Big Data

- Curse of Dimensionality

- No Free Lunch

Big Data – The Four V's

Parametric algorithm: model does not grow with data size

What is Machine Learning?

The Curse of Dimensionality

"Various phenomena that arise when analyzing and organizing data in high-dimensional spaces (often with hundreds or thousands of dimensions) that do not occur in low-dimensional settings such as the three-dimensional physical space of everyday experience." – Wikipedia

- Memory requirement increases
- Required sampling increases
- Distance functions become less useful

…

What is Machine Learning?

No Free Lunch

- There is no universally best model – a set of assumptions that works well in one domain may work poorly in another

- We need many different models, and algorithms that have different speed-accuracy-complexity tradeoffs

Thank You :)

Questions?

Nate Derbinsky

Associate Teaching Professor

Director of Teaching Faculty

Khoury College of Computer Sciences

n.derbinsky@northeastern.edu

https://derbinsky.info

What is Machine Learning?