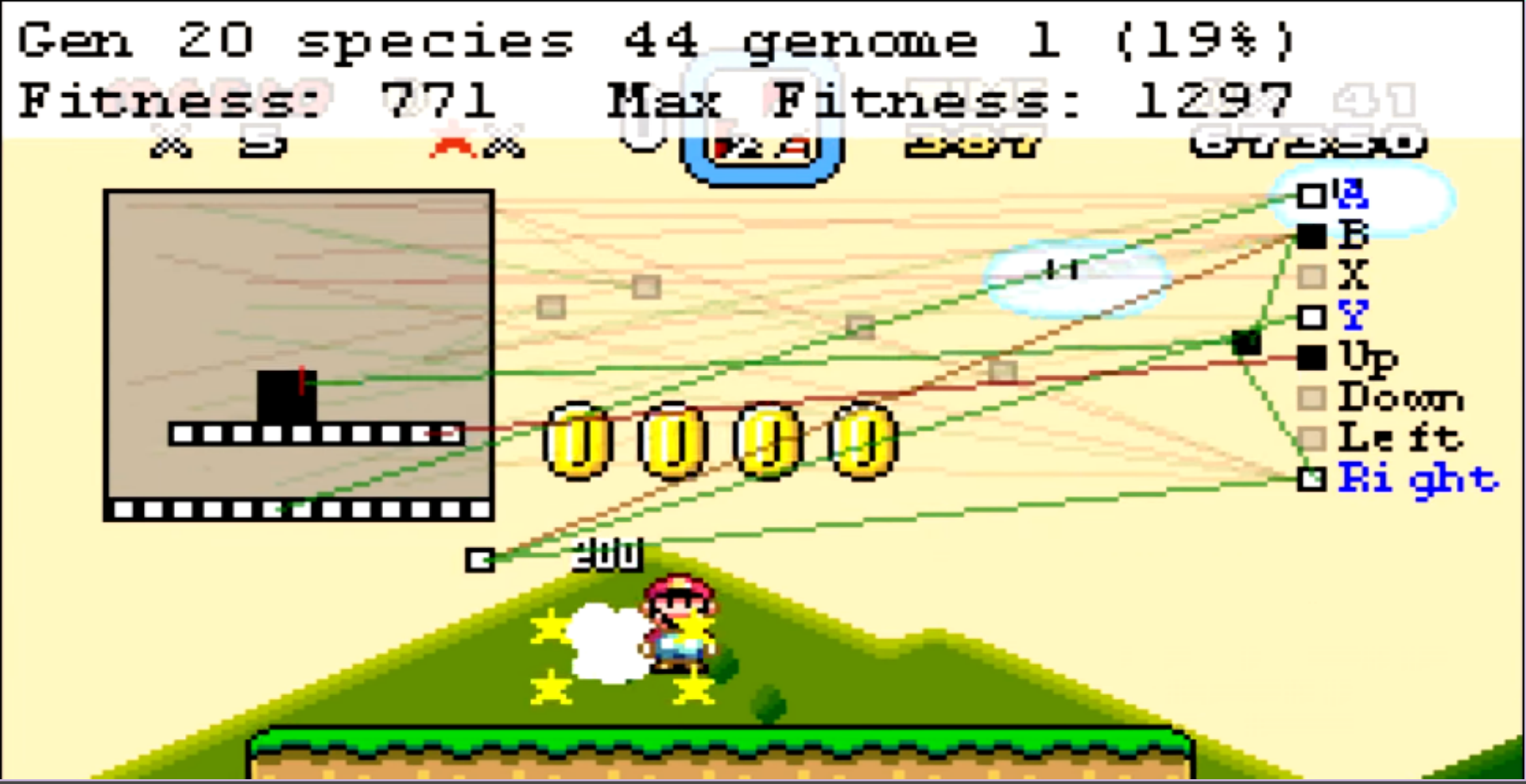


MACHINE LEARNING INTRO

ECS170 Spring 2018

Josh McCoy, @deftjams

We need an algorithm

...but we don't have a sufficient one for:

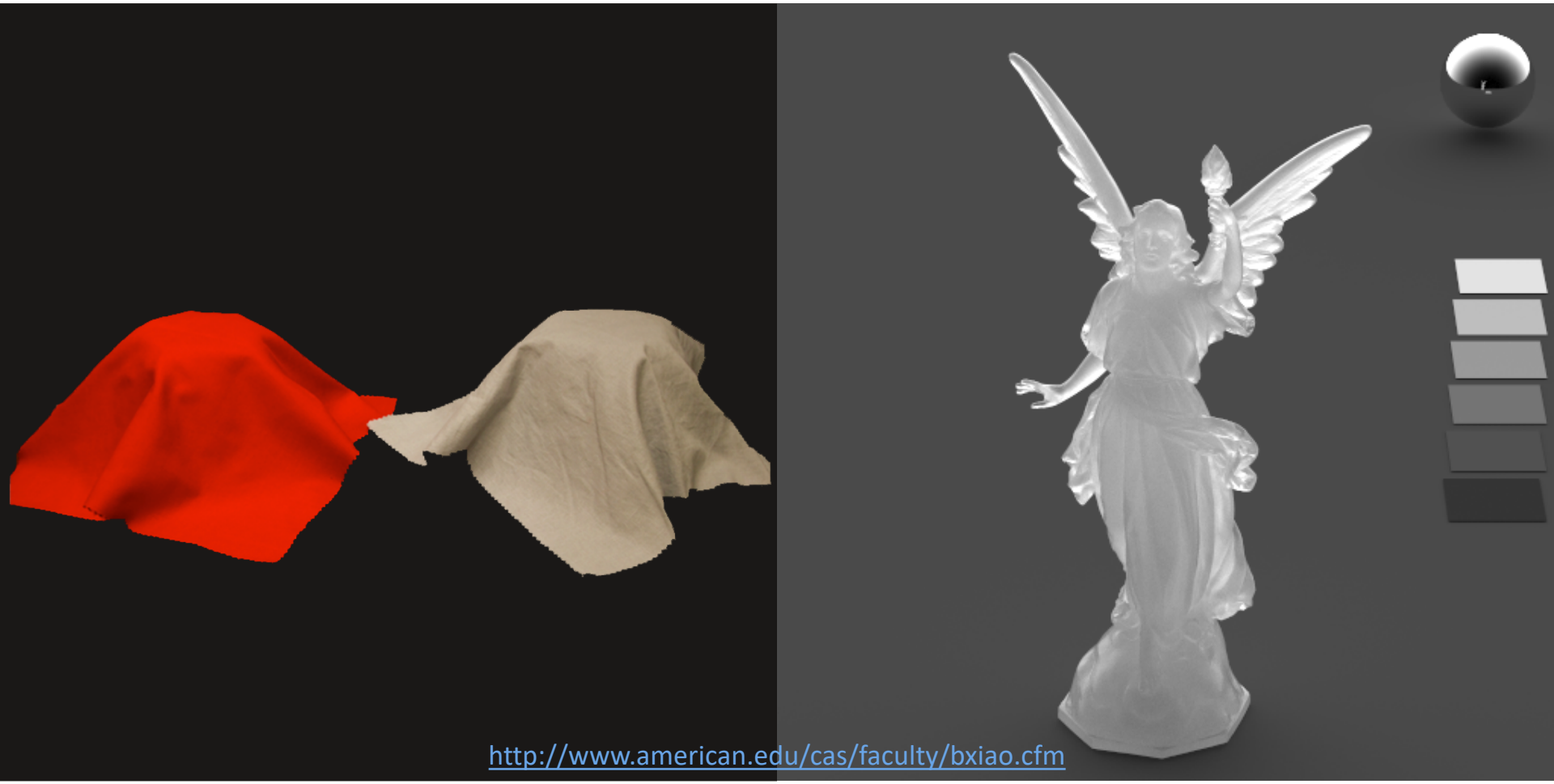
- Spam detection
- Recognizing hand-written digits
- Utility function for Go
- Image segmentation
- Playing Atari games
- Credit applications

We know what we need

mapping from input to:

- Spam detection ⇒Spam or not spam
- Recognizing hand-written digits ⇒0,1,2,3,4,5,6,7,8,9,.
- Utility function for Go ⇒f(state)=value
- Image segmentation ⇒Segments
- Playing Atari games ⇒High score
- Credit applications ⇒How much $ to loan

Estimation of Material Properties



http://www.american.edu/cas/faculty/bxiao.cfm

DARPA Grand Challenge 2005

Stanford

Sebastian Thrun

https://youtu.be/TDqzyd7fDRc?t=1m27s

Autonomous Drifting

MIT Aerospace Controls Lab

https://www.youtube.com/watch?v=opsmd5yuBF0

Obama-RNN

Samim.io

https://medium.com/@samim/obama-rnn-machine-generated-political-speeches-c8abd18a2ea0#.nr6vqxlgi

https://github.com/samim23/obama-rnn/

“Good afternoon. God bless you.

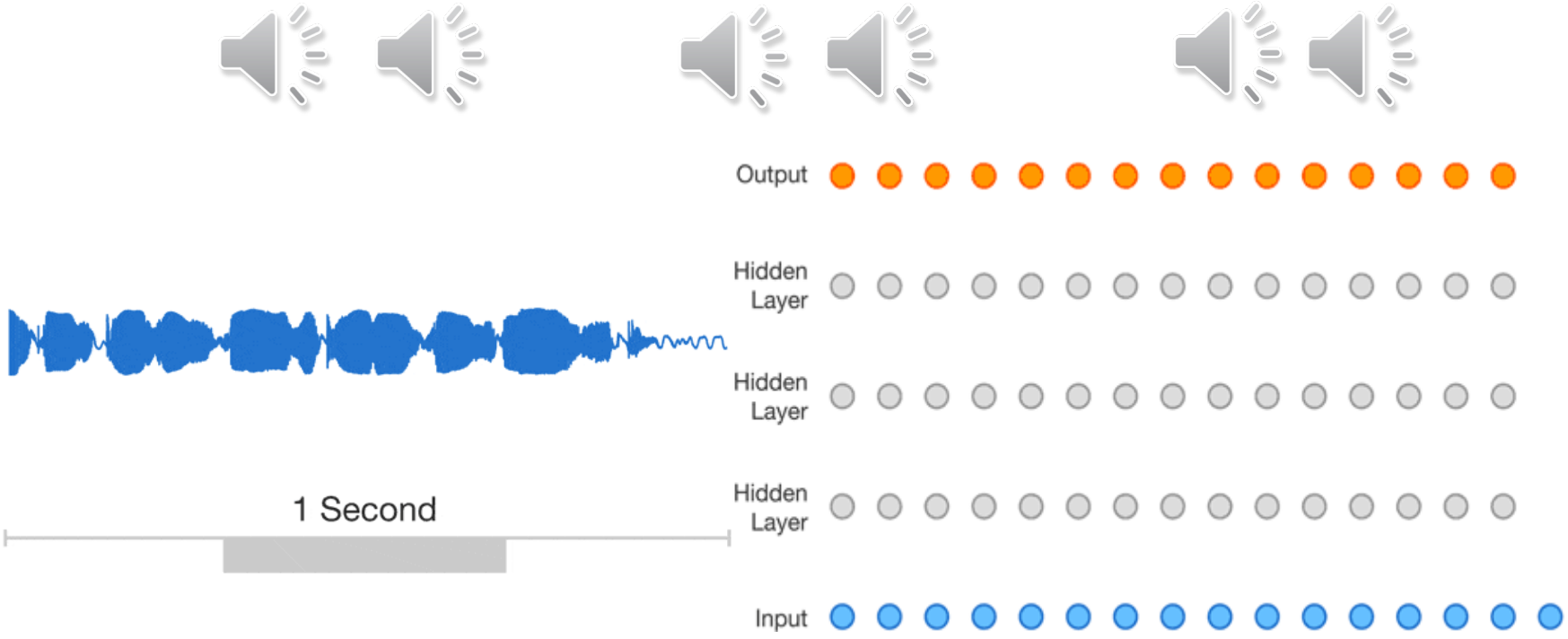
The United States will step up to the cost of a new challenges of the American people that will share the fact that we created the problem. They were attacked and so that they have to say that all the task of the final days of war that I will not be able to get this done. The promise of the men and women who were still going to take out the fact that the American people have fought to make sure that they have to be able to protect our part. It was a chance to stand together to completely look for the commitment to borrow from the American people. And the fact is the men and women in uniform and the millions of our country with the law system that we should be a strong stretcks of the forces that we can afford to increase our spirit of the American people and the leadership of our country who are on the Internet of American lives. Thank you very much.

God bless you, and God bless the United States of America. “

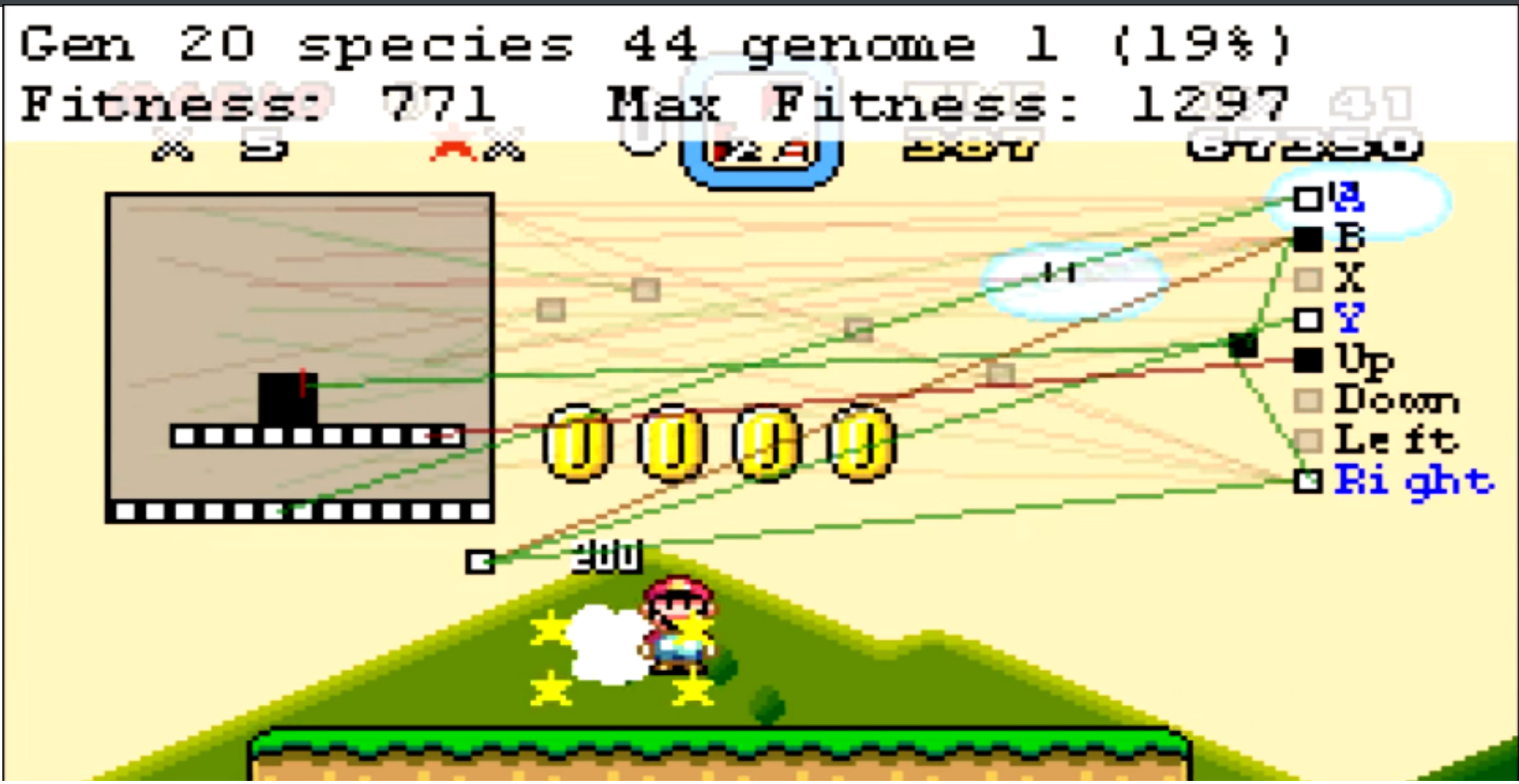
DeepMind Wavenet

https://github.com/ibab/tensorflow-wavenet

https://deepmind.com/blog/wavenet-generative-model-raw-audio/



MarI/O



https://www.youtube.com/watch?v=qv6UVOQ0F44

Data instead of knowledge

What we lack in knowledge we make up for in data to find patterns and regularities.

We may not be able to model the process completely but we can construct a good and useful approximation.

Original Definition of Machine Learning

Field of study that gives computers the ability to learn without being explicitly programmed.

Arthur Samuel, 1959

First Application

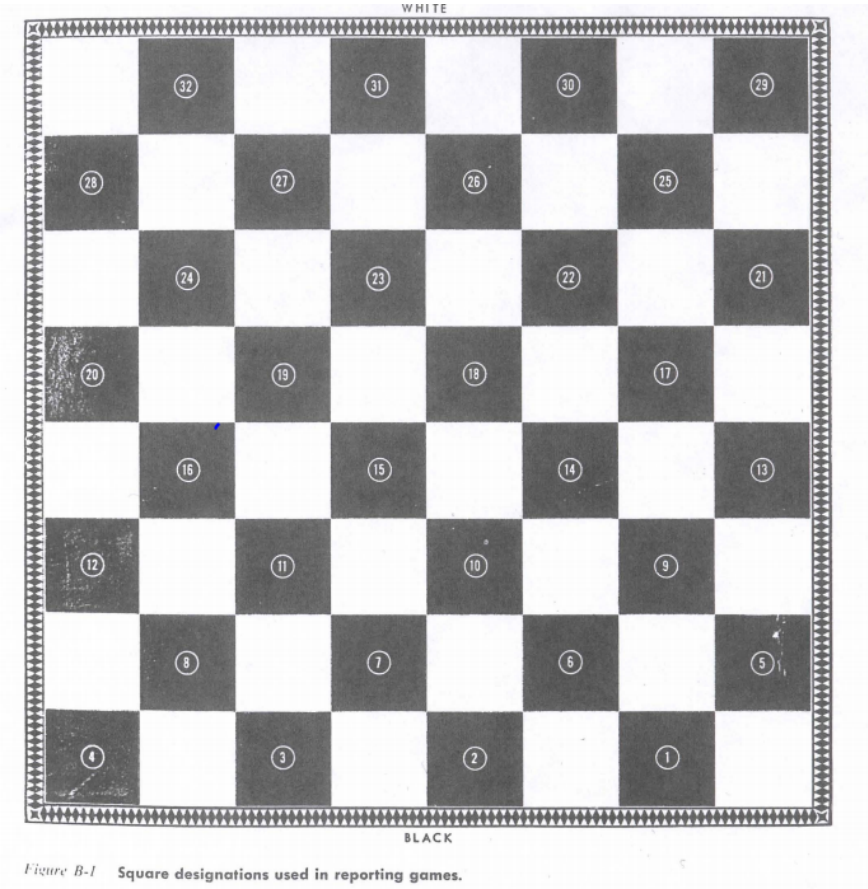
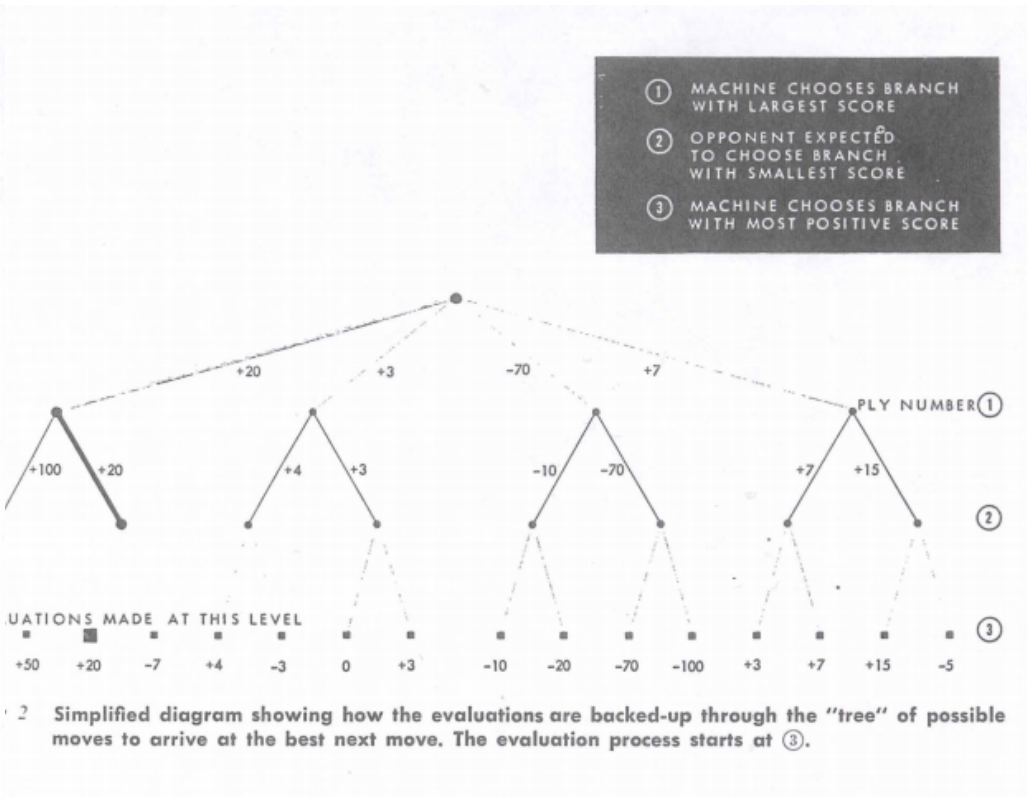


Figure B-1 **Square designations used in reporting games.**



2 **Simplified diagram showing how the evaluations are backed-up through the "tree" of possible moves to arrive at the best next move. The evaluation process starts at ③.**

https://www.cs.virginia.edu/~evans/greatworks/samuel1959.pdf

Updated Definition

Field of study that gives computers the ability to learn without being explicitly programmed.

Arthur Samuel, 1959

Updated Definition

Field of study that gives computers the ability to learn without being explicitly programmed.

Arthur Samuel, 1959

A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P* if its performance at tasks in *T*, as measured by *P*, improves with experience *E*.

Tom M. Mitchell, 1997

Learning

A agent is learning if it improves its performance on future tasks after making observations about the world

Russell and Norvig p. 693

- What is learning?

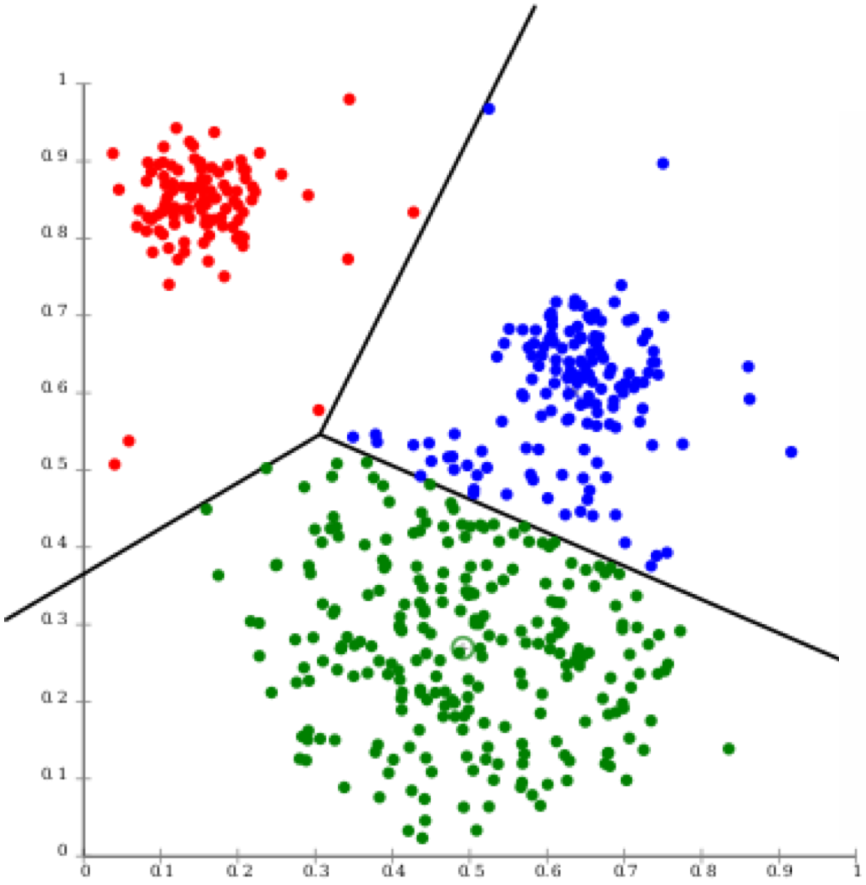
observations → Learning → *Skill*

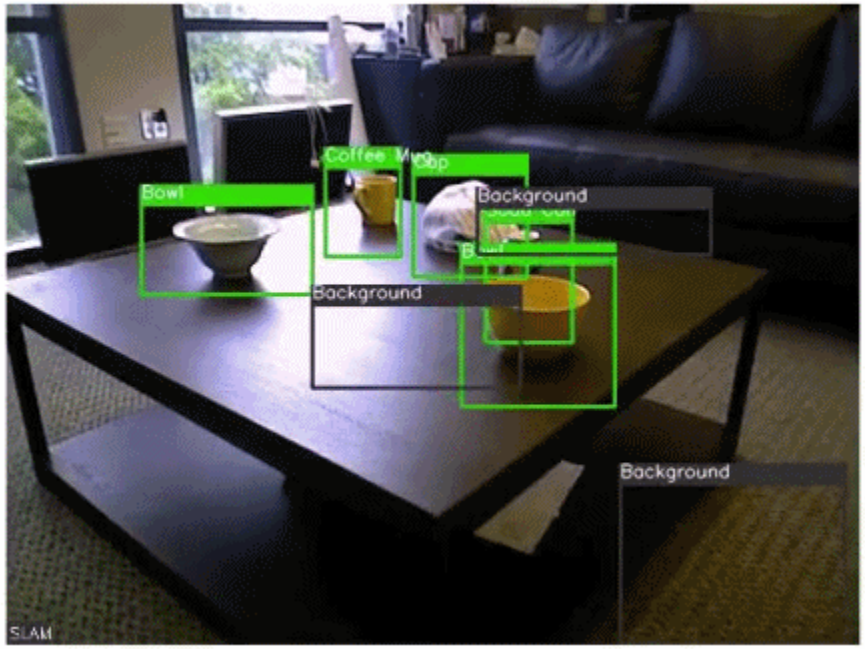
- Machine learning:

data → Machine Learning → *Skill*

http://www.stat.ucdavis.edu/~chohsieh/teaching/ECS171_Winter2018/lecture1.pdf

Data(base|science) vs Unforeseen





https://en.wikipedia.org/wiki/Cluster_analysis#/media/File:KMeans-Gaussian-data.svg

Monocular SLAM-Supported Object Recognition

Approximating Functions

Goal: find a function $h(x)$ that approximates the true but unknown function $f(x)$.

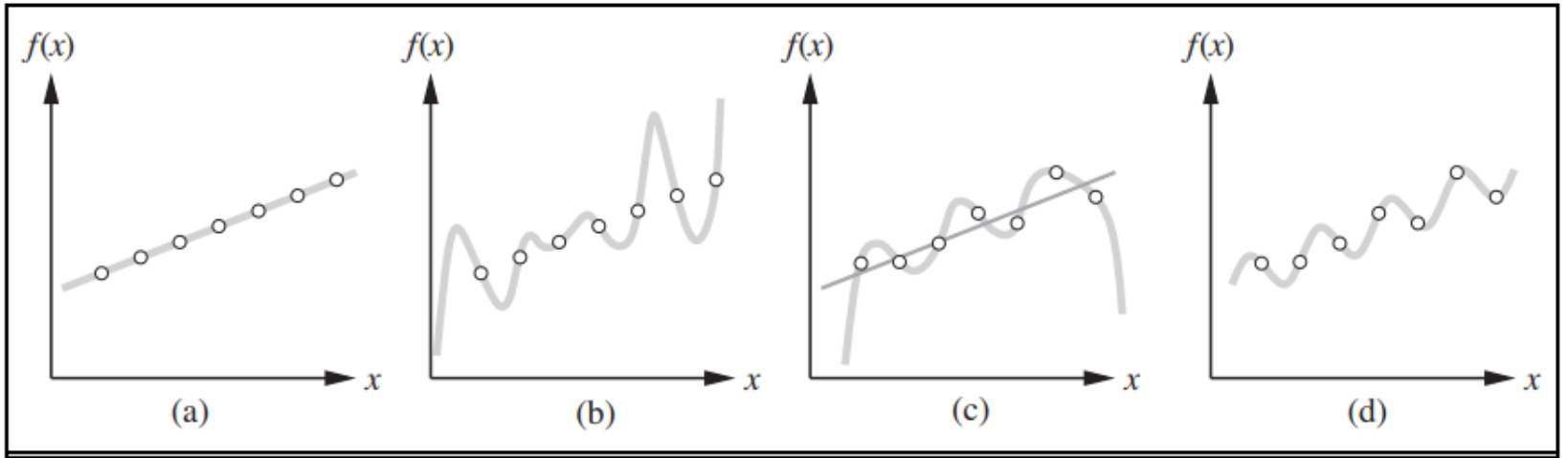


Figure 18.1 (a) Example $(x, f(x))$ pairs and a consistent, linear hypothesis. (b) A consistent, degree-7 polynomial hypothesis for the same data set. (c) A different data set, which admits an exact degree-6 polynomial fit or an approximate linear fit. (d) A simple, exact sinusoidal fit to the same data set.

Hypothesis

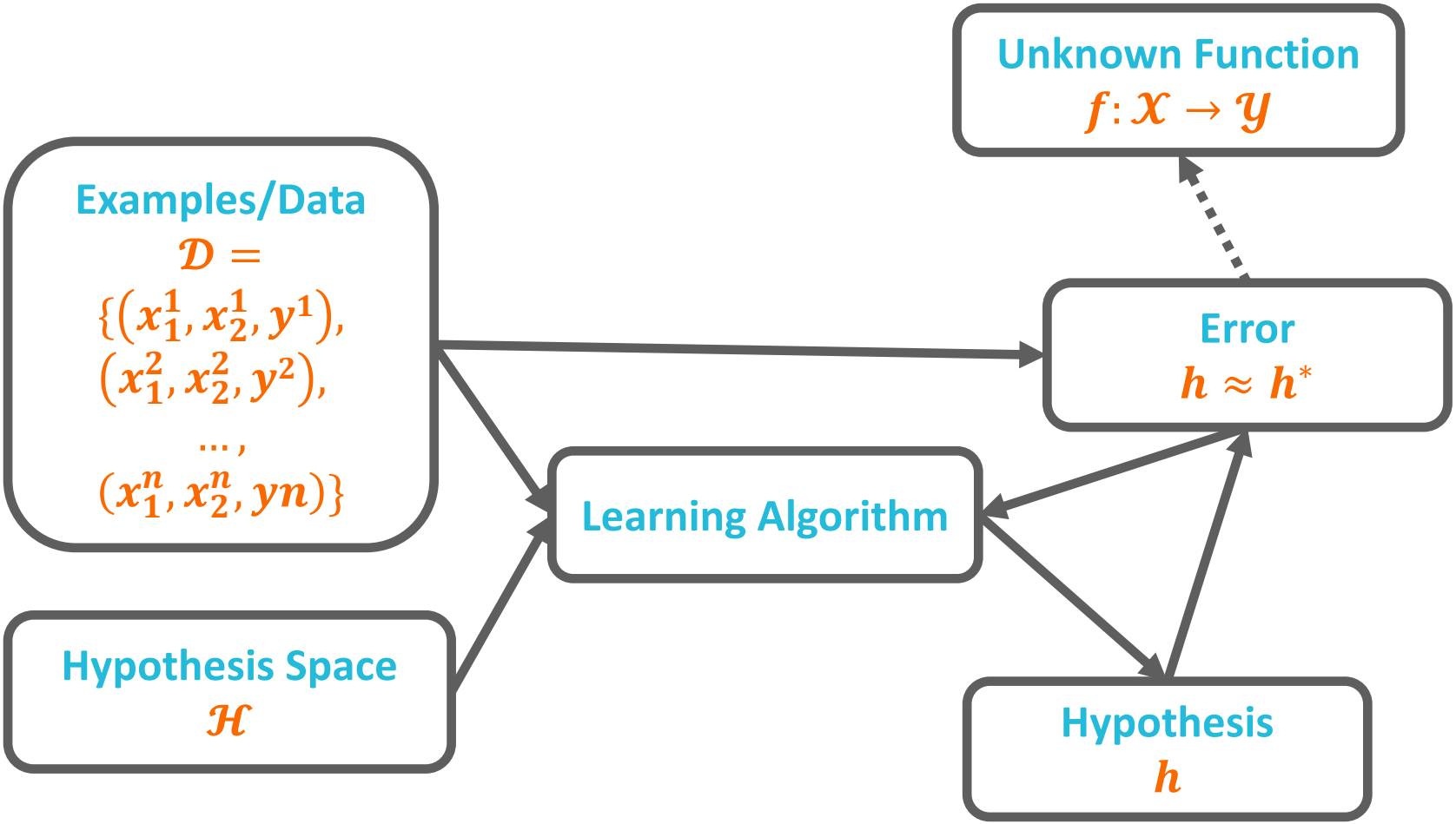
- The **hypothesis** h is one possible approximation of $f(x)$.
- All possible hypotheses are in the **hypothesis space**:
  - $h \in \mathcal{H}$
- **Most probable hypothesis** h^* given the data:
  - $h^* = argmax\ P(h|data)$ where $h^* \in \mathcal{H}$

There is a tradeoff between the expressiveness of a hypothesis space and the complexity of finding a good hypothesis within that space.

Machine Learning Formalization

- **Input**: $x \in \mathcal{X}$
  - Ex. input with two features: $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$
- **Output**: $y \in \mathcal{Y}$
- **Approximation** of the unknown function
  - $f: \mathcal{X} \rightarrow \mathcal{Y}$
- Labeled **Data**
  - $\mathcal{D} = \{(x_1^1, x_2^1, y^1), (x_1^2, x_2^2, y^2), \ldots, (x_1^n, x_2^n, yn)\}$

Process


Unknown Function
$f: X \rightarrow Y$
Examples/Data
$\mathcal{D} =$
$\{(x_1^1, x_2^1, y^1),$
$(x_1^2, x_2^2, y^2),$
$\ldots,$
$(x_1^n, x_2^n, yn)\}$
Error
$h \approx h^*$
Learning Algorithm
Hypothesis Space
$\mathcal{H}$
Hypothesis
h

Common Mode

Factored Representation - a vector of attribute values that are often seen as tensors.

- A.K.A. feature vectors

Output - continuous (probabilities, scalars) or discrete values (group membership).

Example: StarCraft II

Example: StarCraft II

Input features:

- Unit positions
- Resources
- Health/shields
- Vision
- Terrain/map
- Energy
- Cooldowns
- …

Output:

- A single action or a set of game actions.

Machine Learning Styles

- Supervised
  - All $y^i \in \mathcal{Y}$ are in $\mathcal{D}$.
- Unsupervised
  - No $y^i \in \mathcal{Y}$ are present in $\mathcal{D}$.
- Semi-supervised
  - Some $y^i \in \mathcal{Y}$ are missing from $\mathcal{D}$.
- Reinforcement
  - Reward and punishment based on state transitions.

- Offline
  - Trains h then uses h.
- Online
  - Trains h, uses h, trains h, uses h, ...
- Deep
  - More layers and hierarchy in NNs.

Groups of Methods

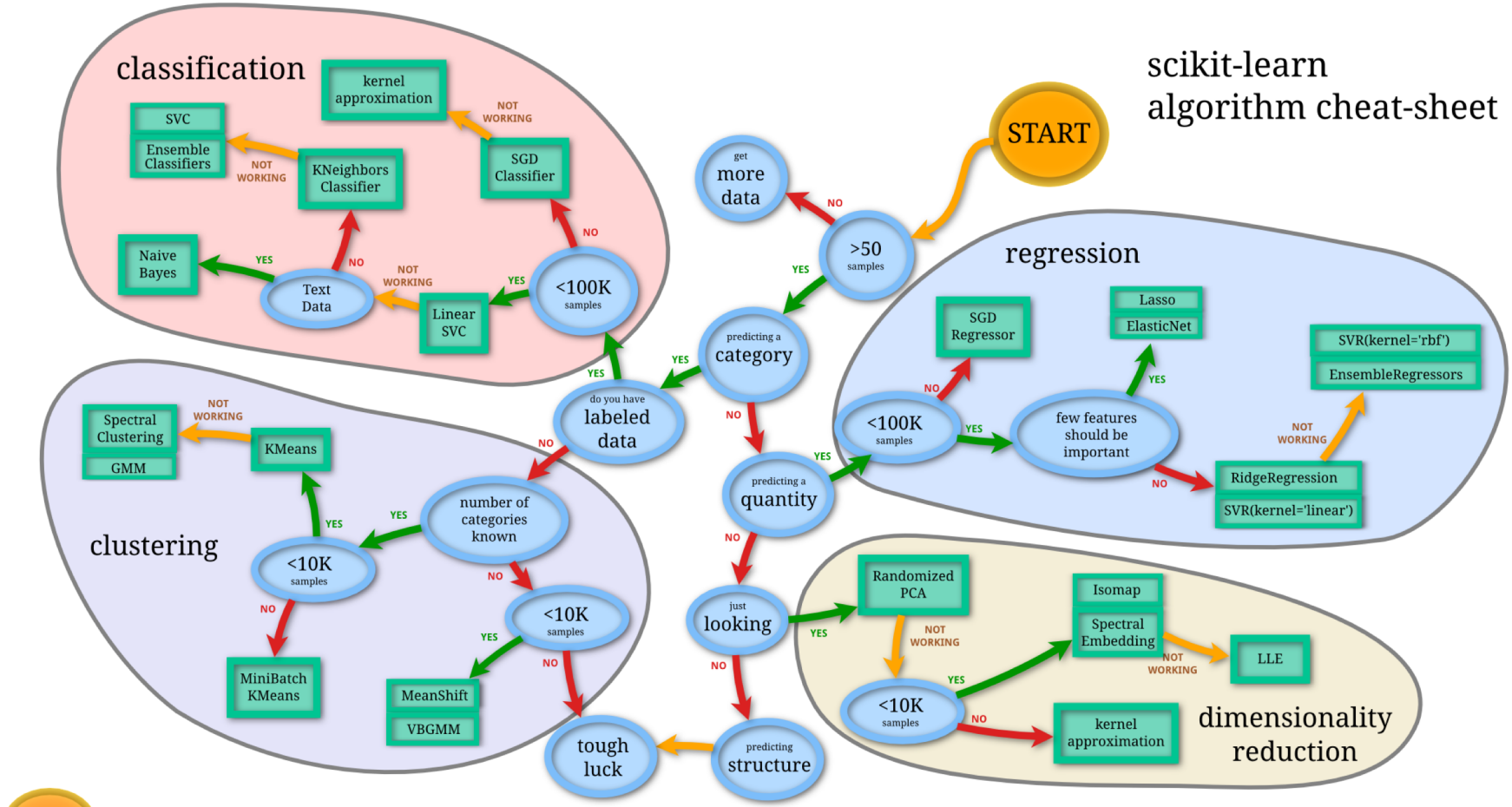
- Classification*
  - Multi-class
- Regression*
- Instance-based*
- Decision Trees
- Bayesian
- Clustering
- Association Rule
- Neural Networks
  - Deep
- Case-based
- Dimensionality Reduction
- Symbolic
- Kernel
- Anomaly Detection
- Reinforcement
- Genetic
- Markov Processes

**Broader categories based more on output type and less on algorithmic similarity.*

Enhancements, Add-ons, and Meta

- Regularization
- Ensemble
  - Boosting
  - Random Forest
- Meta
  - Inductive bias
  - Metadata

The right tool



http://scikit-learn.org/stable/tutorial/machine_learning_map/

The right tool

Microsoft Azure Machine Learning: Algorithm Cheat Sheet

This cheat sheet helps you choose the best Azure Machine Learning Studio algorithm for your predictive analytics solution. Your decision is driven by both the nature of your data and the question you're trying to answer.

START

Finding unusual data points

Discovering structure

Predicting categories

Predicting values

Three or more

Two

ANOMALY DETECTION

- **One-class SVM** — >100 features, aggressive boundary
- **PCA-based anomaly detection** — Fast training

CLUSTERING

- **K-means**

MULTI-CLASS CLASSIFICATION

- Fast training, linear model — **Multiclass logistic regression**
- Accuracy, long training times — **Multiclass neural network**
- Accuracy, fast training — **Multiclass decision forest**
- Accuracy, small memory footprint — **Multiclass decision jungle**
- Depends on the two-class classifier, see notes below — **One-v-all multiclass**

REGRESSION

- **Ordinal regression** — Data in rank ordered categories
- **Poisson regression** — Predicting event counts
- **Fast forest quantile regression** — Predicting a distribution
- **Linear regression** — Fast training, linear model
- **Bayesian linear regression** — Linear model, small data sets
- **Neural network regression** — Accuracy, long training time
- **Decision forest regression** — Accuracy, fast training
- **Boosted decision tree regression** — Accuracy, fast training, large memory footprint

TWO-CLASS CLASSIFICATION

- **Two-class SVM** — >100 features, linear model
- **Two-class averaged perceptron** — Fast training, linear model
- **Two-class logistic regression** — Fast training, linear model
- **Two-class Bayes point machine** — Fast training, linear model

- Accuracy, fast training — **Two-class decision forest**
- Accuracy, fast training, large memory footprint — **Two-class boosted decision tree**
- Accuracy, small memory footprint — **Two-class decision jungle**
- >100 features — **Two-class locally deep SVM**
- Accuracy, long training times — **Two-class neural network**

Created by the Azure Machine Learning Team

Email: AzurePoster@microsoft.com

Download this poster: http://aka.ms/MLCheatSheet

Microsoft

The right tool

One of the most comprehensive overviews:

https://en.wikipedia.org/wiki/Outline_of_machine_learning#Machine_learning_methods

Lacks ontology and consistency but has most algorithms listed.