INTRO TO MACHINE LEARNING

WHAT IS MACHINE LEARNING?

- "Machine learning is the field of study that gives computers the ability to learn without being explicitly programmed"

  —Arthur L. Samuel, AI pioneer, 1959

WHAT IS MACHINE LEARNING?

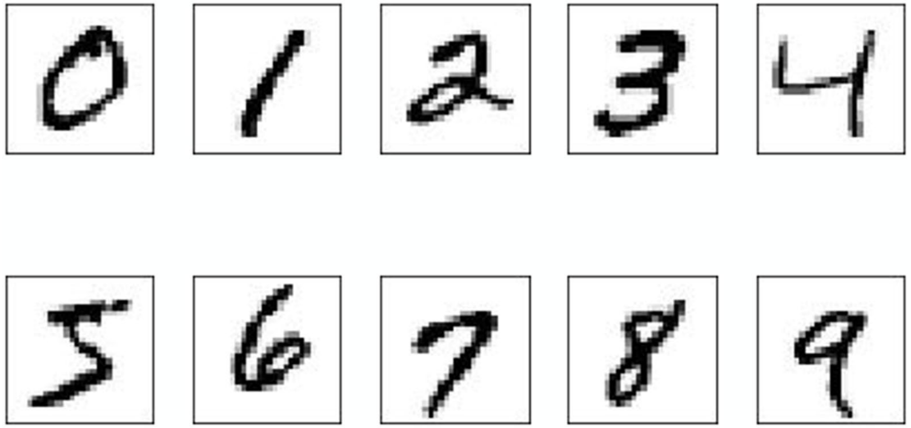
- A program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.

  — Tom Mitchell, Professor at Carnegie Mellon

  - Experience: Data
  - Task: Prediction, etc.
  - Performance: RSME, etc.

MACHINE LEARNING TASKS

Handwriting Recognition Example:



- Experience:
- Task:
- Performance:

WAIT, I'VE ALSO HEARD ABOUT STATISTICAL LEARNING…

- Statistical Learning = Machine Learning (ish)
  - "Statistical" learning emphasizes modern ML's roots and foundation in probability and statistics for handling uncertainty
- So what is statistical/machine learning?
  - A field of study concerned with making quantitative inferences and predictions from data
  - Uses statistical techniques and computational power to "learn" these inferences and predictions without being explicitly programmed
  - A little more formally: an algorithm is said to learn from data if its performance improves as the amount of data increases

SUCCESS OF MACHINE LEARNING

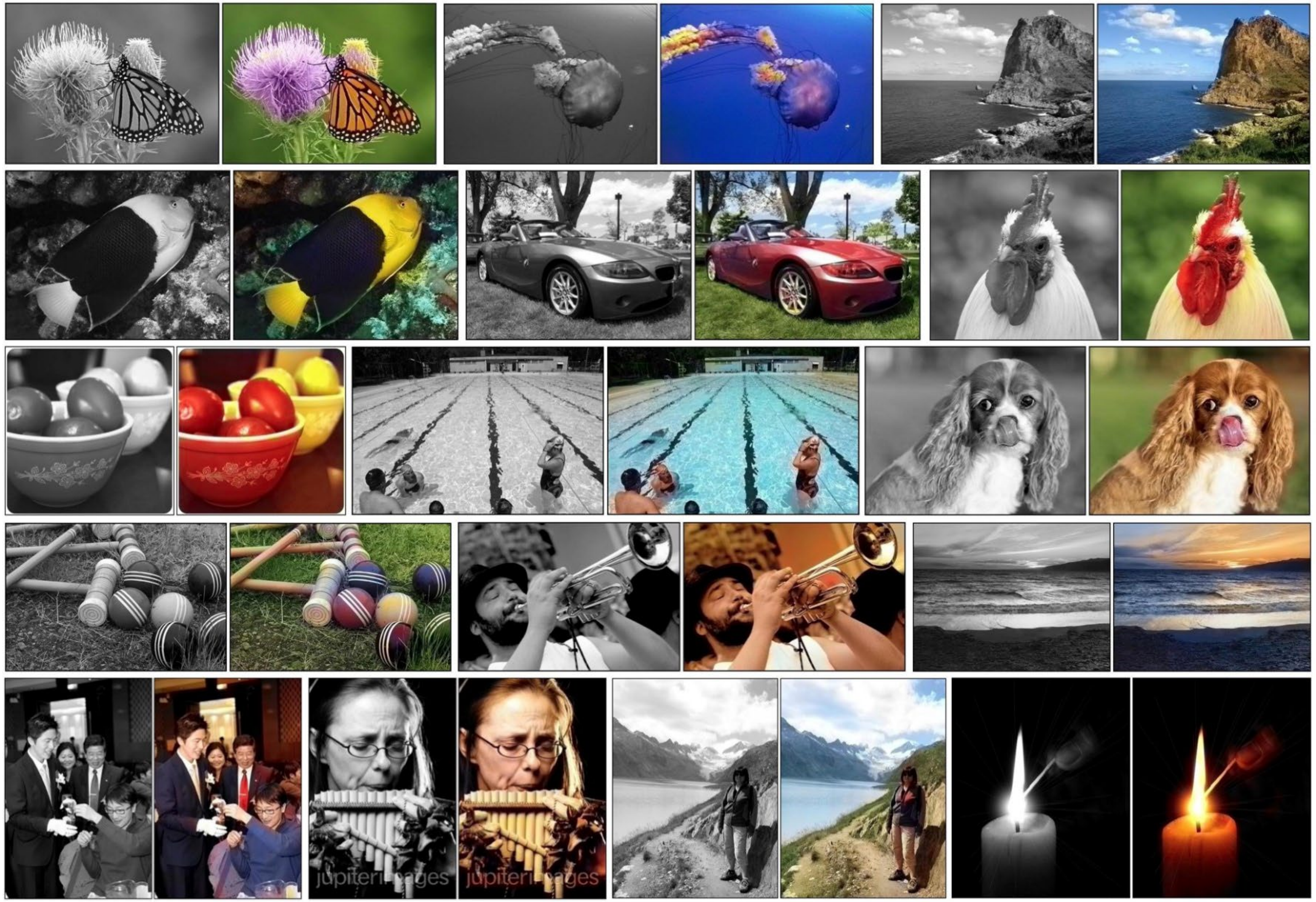
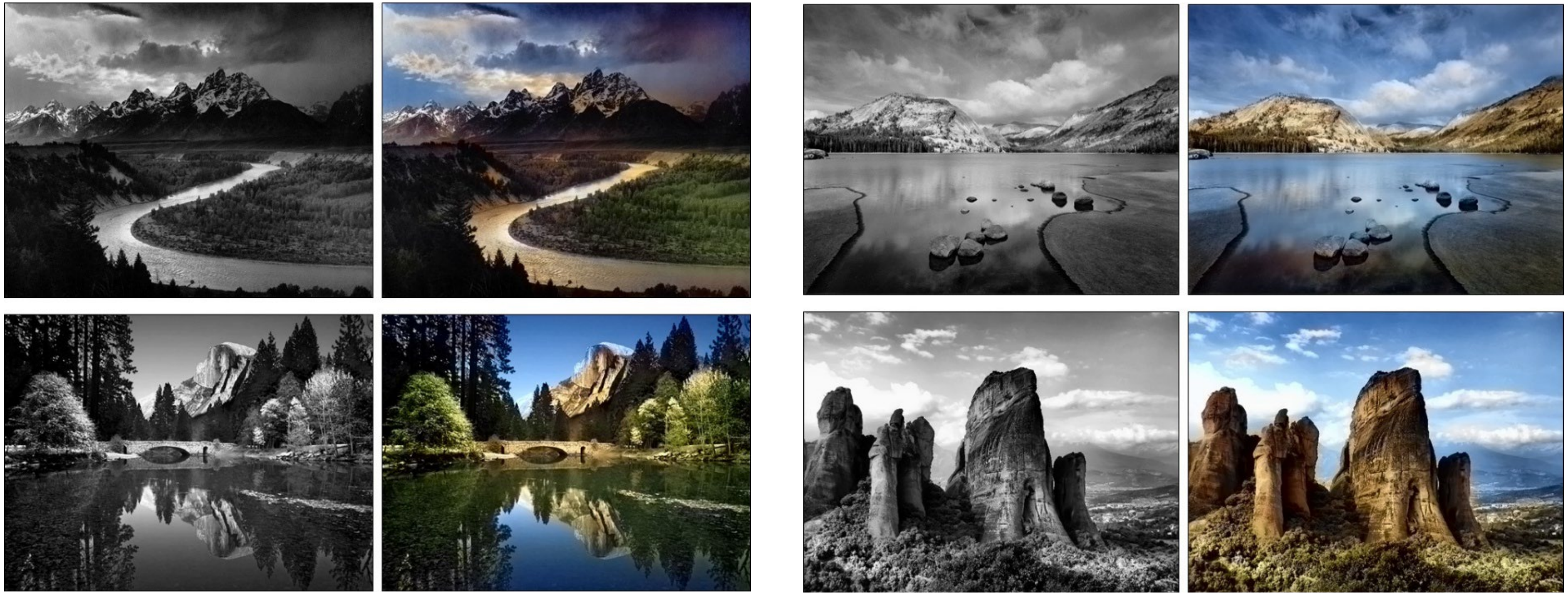


Image colorization
(Zhang et al., 2016)

SUCCESS OF MACHINE LEARNING

Image colorization (Zhang et al., 2016)

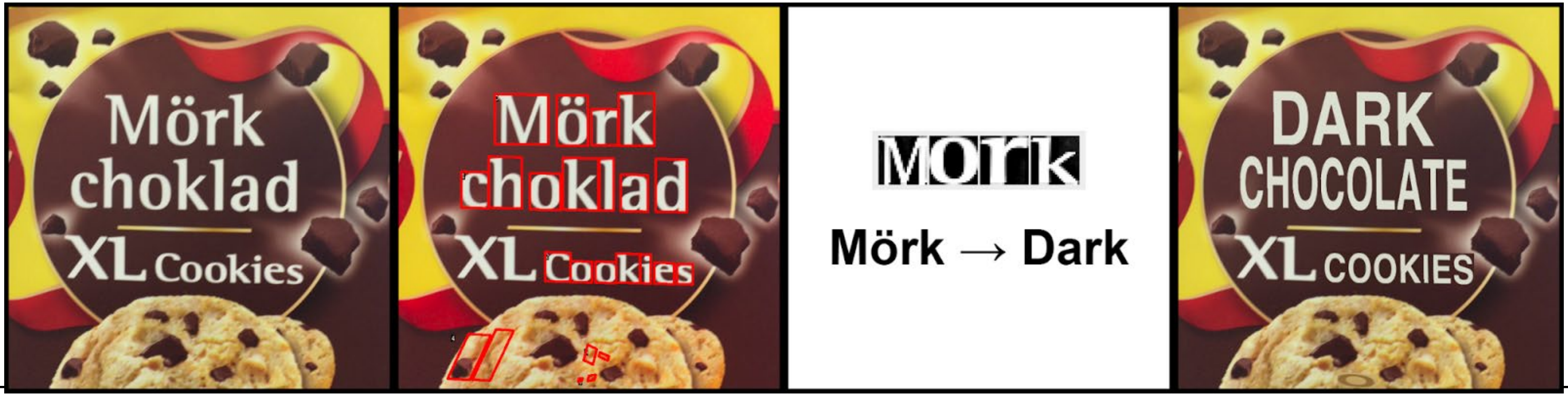


Colorized classical photographs by Ansel Adams

SUCCESS OF MACHINE LEARNING

Real-time visual translation on smartphones

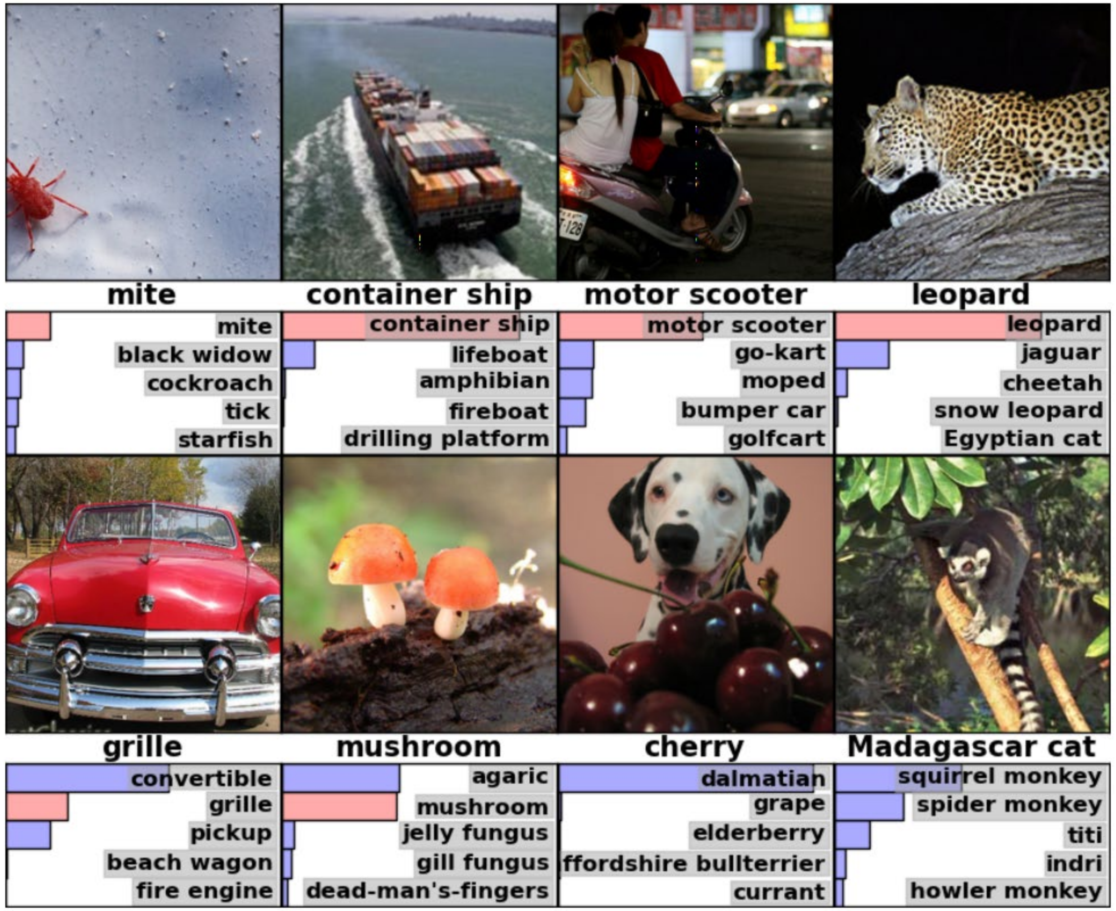
1. Find the letters
2. Recognize the letters
3. Translate
4. Render the translation in the same style



Google blog, 2015

SUCCESS OF MACHINE LEARNING

Object classification/detection in images (Krizhevsky et al., 2012)



SUCCESS OF MACHINE LEARNING

Find the bug with this code:

```
for (var i = 0; i < 5; i++) {
  setTimeout(() => console.log(i), 1000)
}
```

The code contains a bug that will cause it to always print the number 5 to the console, rather than the expected sequence of numbers from 0 to 4. This is because the `setTimeout` function creates a closure that captures the value of the `i` variable, but the `i` variable is changed before the callback function is executed.

To fix this bug, the code can be modified to create a new variable in the loop that is passed to the `setTimeout` callback function, rather than using the loop variable directly. This will ensure that each callback function captures the correct value of the loop variable, and the expected output will be produced. Here is an example of how the code could be fixed:

Copy code

```
for (var i = 0; i < 5; i++) {
  // Create a new variable and pass its value to setTimeout
  const value = i;
  setTimeout(() => console.log(value), 1000)
}
```

This code will print the numbers 0 to 4 to the console, as expected. Alternatively, the `let` keyword can be used in place of `var` to declare the loop variable, which will automatically create a new variable for each iteration of the loop and avoid the need to create a new variable manually. This is a common pattern for avoiding closure-related bugs in JavaScript.

Automated programming with ChatGPT

SUCCESS OF MACHINE LEARNING

AI generated art

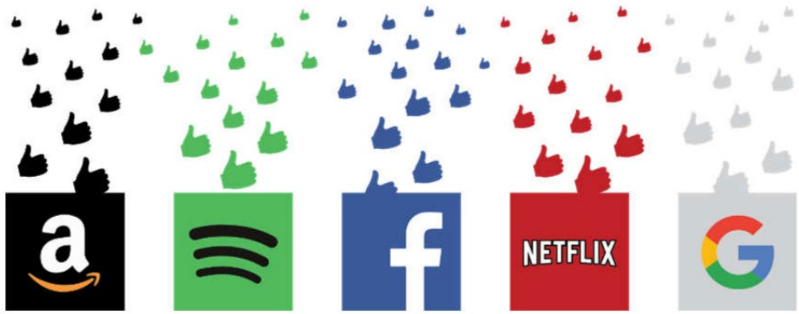
Best Dalle2 Pics
@Dalle2Pics

A raccoon playing tennis at Wimbledon in the 1990s
#dalle2 #dalle

3:33 AM · May 14, 2022

OTHER MACHINE LEARNING APPLICATIONS

- Chatbots (NLP)
- Fraud detection
- Route planning
- Search engine refinement
- Face recognition/detection
- Credit risk assessment
- Financial market prediction
- Medical diagnosis
- Personalized medicine
- Electricity demand forecasting
- Spam filtering
- Collision avoidance systems
- Speech synthesis/analysis (e.g. Siri, Alexa)
- Recommender systems (e.g. Netflix, Spotify)



SUCCESS OF MACHINE LEARNING

Alpha Go Zero

nature
International journal of science

Access provided by Yale University

Altmetric: 2188 Citations: 1 More detail »

Article

Mastering the game of Go without human knowledge

David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel & Demis Hassabis

Nature **550**, 354–359 (19 October 2017)
doi:10.1038/nature24270
Download Citation

Computational science
Computer science Reward

Received: 07 April 2017
Accepted: 13 September 2017
Published online: 18 October 2017

Alpha Zero

arXiv.org > cs > arXiv:1712.01815

Search or
(Help | Advanc

Computer Science > Artificial Intelligence

Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm

David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, Timothy Lillicrap, Karen Simonyan, Demis Hassabis

(Submitted on 5 Dec 2017)

The game of chess is the most widely-studied domain in the history of artificial intelligence. The strongest programs are based on a combination of sophisticated search techniques, domain-specific adaptations, and handcrafted evaluation functions that have been refined by human experts over several decades. In contrast, the AlphaGo Zero program recently achieved superhuman performance in the game of Go, by tabula rasa reinforcement learning from games of self-play. In this paper, we generalise this approach into a single AlphaZero algorithm that can achieve, tabula rasa, superhuman performance in many challenging domains. Starting from random play, and given no domain knowledge except the game rules, AlphaZero achieved within 24 hours a superhuman level of play in the games of chess and shogi (Japanese chess) as well as Go, and convincingly defeated a world-champion program in each case.

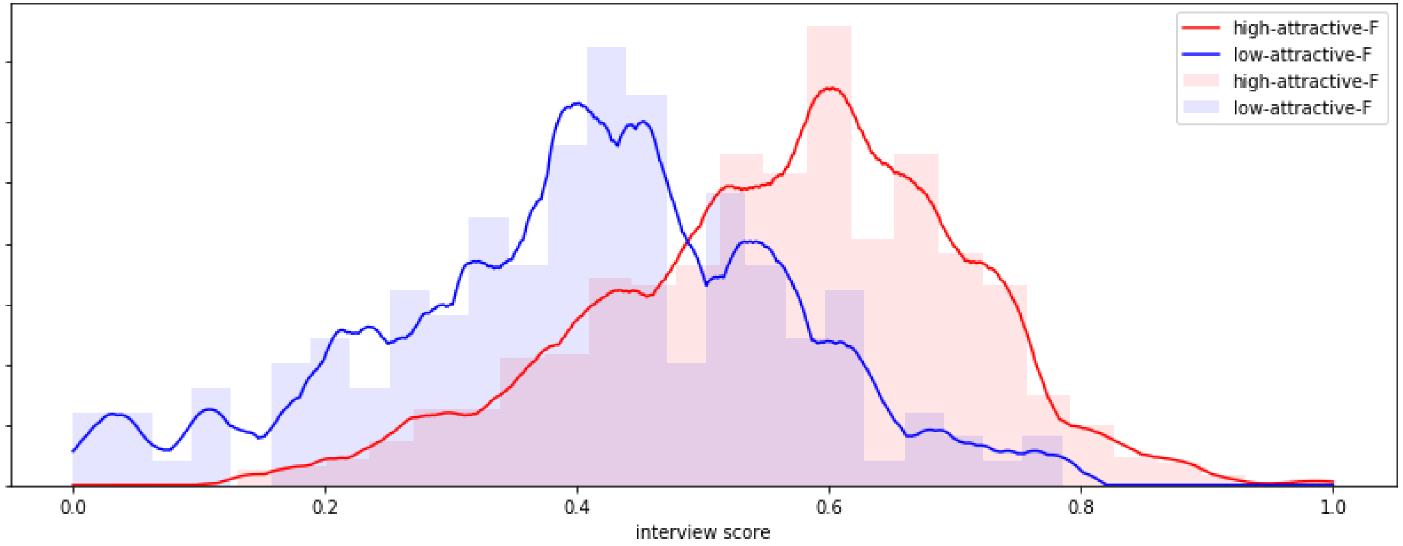
Subjects: **Artificial Intelligence (cs.AI)**; Learning (cs.LG)
Cite as: arXiv:1712.01815 **[cs.AI]**
(or **arXiv:1712.01815v1** **[cs.AI]** for this version)

Submission history
From: David Silver [view email]
[v1] Tue, 5 Dec 2017 18:45:38 GMT (272kb,D)

Automatic game playing

SUCCESS OF MACHINE LEARNING

- Finding bias in human-based job interviews
- More attractive women are rated higher by humans



"Finding racism under every rock" by Ben Taylor

- Similar method finds evidence of some racial bias

CATEGORIES OF ML

- Supervised Learning
  - Labeled Data
  - Predict Outcome
  - Direct Feedback (E.g., accuracy, RMSE. Etc.)
- Unsupervised Learning
  - No Labels/Targets
  - “Fuzzy” Feedback: No Target = No Objective Less Clear
- Reinforcement Learning
  - Set of states/actions
  - Maximize “reword”

MACHINE LEARNING VOCABULARY

Term used in Statistics	Term used in Machine Learning
Explanatory variables / Predictors	
Response variable	
Parameter estimates	
Intercept	

SOME NOTATION

- Typically will denote a measurement as $\mathbf{x} \in \mathbb{R}^d$
- $\mathbf{x}$ is usually a Euclidean vector, denoted

$$\mathbf{x} = \begin{bmatrix} x^{(1)} \\ \vdots \\ x^{(d)} \end{bmatrix} \qquad \text{or} \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix}$$

- $\mathbf{x}$ is a pattern, signal, input, instance, or feature vector
- x_i is a feature, attribute, predictor, or covariate
- $\mathbf{x}$ is typically viewed as a realization of a random variable/vector $\mathbf{X}$

GOALS OF SUPERVISED ML

- Prediction
  - Predict the target for new observations
  - Forecast the target for future observations

- Interpretation
  - Which feature are associated with the target?
  - What is the relationship between target and each feature?
  - Are some features more important than others?
  - Is there an explainable grouping in the data?

SUPERVISED LEARNING

- The learner/user are given training data

$$(x_1, y_1), \ldots, (x_n, y_n)$$

- Each y_i is the output of an unknown and possibly noisy function with input $\mathbf{x}_i$
- Goal: Given a test input $\mathbf{x}$, predict the correct output
- Two primary types of supervised learning: classification and regression

CLASSIFICATION

- Outputs are called labels and are typically finite:

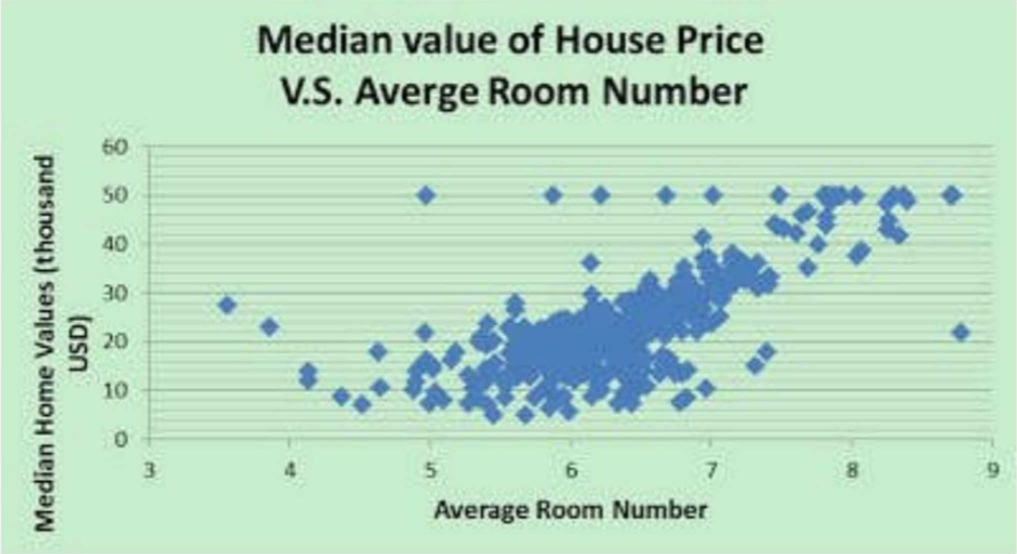
$$y \in \{1, 2, \ldots, C\}$$

  where C is the number of classes

- Example: handwritten digit recognition

REGRESSION

Outputs are called responses and are real valued: $y \in \mathbb{R}$



UNSUPERVISED LEARNING

- Data is unlabeled
- **Goal:** infer a property from the data, not to predict a property from future data
- Examples:
  - **Clustering:** What are the distinct clusters or groups in a dataset?
  - **Density Estimation:** What is the probability density function that generated the data?
  - **Dimensionality Reduction:** How can we reduce the dimensionality of the data without losing (much) information?
  - **Anomaly Detection:** Automatically identifying outliers.

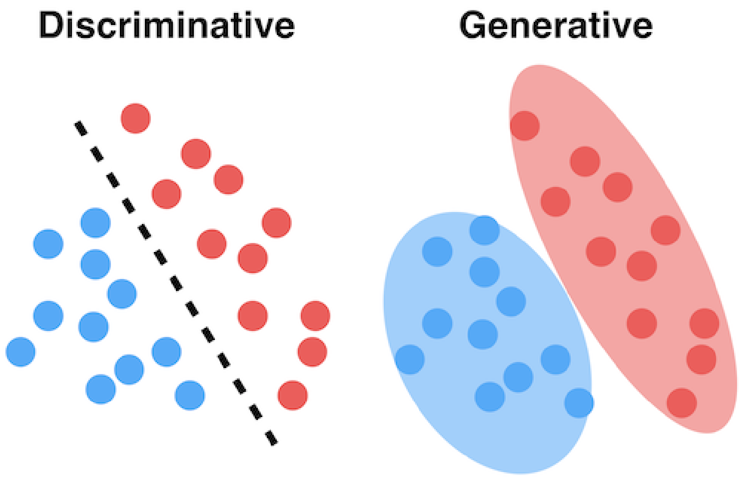
CHALLENGES IN MACHINE LEARNING

- Big data (both sample size and dimension)
- Imbalanced labels
- Noisy labels
- Missing labels
- Missing features
- Uncertainty
- Data representations
- Safety (e.g., self-driving cars)
- Fairness
- Ethics
- Explainability/Interpretability
- Changing environment

ARTIFICIAL INTELLIGENCE VS. MACHINE LEARNING

- **Artificial Intelligence (AI)** is a broad field concerned with building systems that exhibit intelligent behavior
- As a field, AI generally includes machine learning, but also includes more
- Most recent, popular "AI" advances are better classified as machine learning advances
- AI has become the latest buzzword, so many machine learning methods and problems are being rebranded as AI
- Note: Many statistical methods have also been rebranded as ML methods…
  - Linear Regression (and variants, such as LASSO)
  - Logistic Regression
  - Principal Components Analysis

SOME TERMS



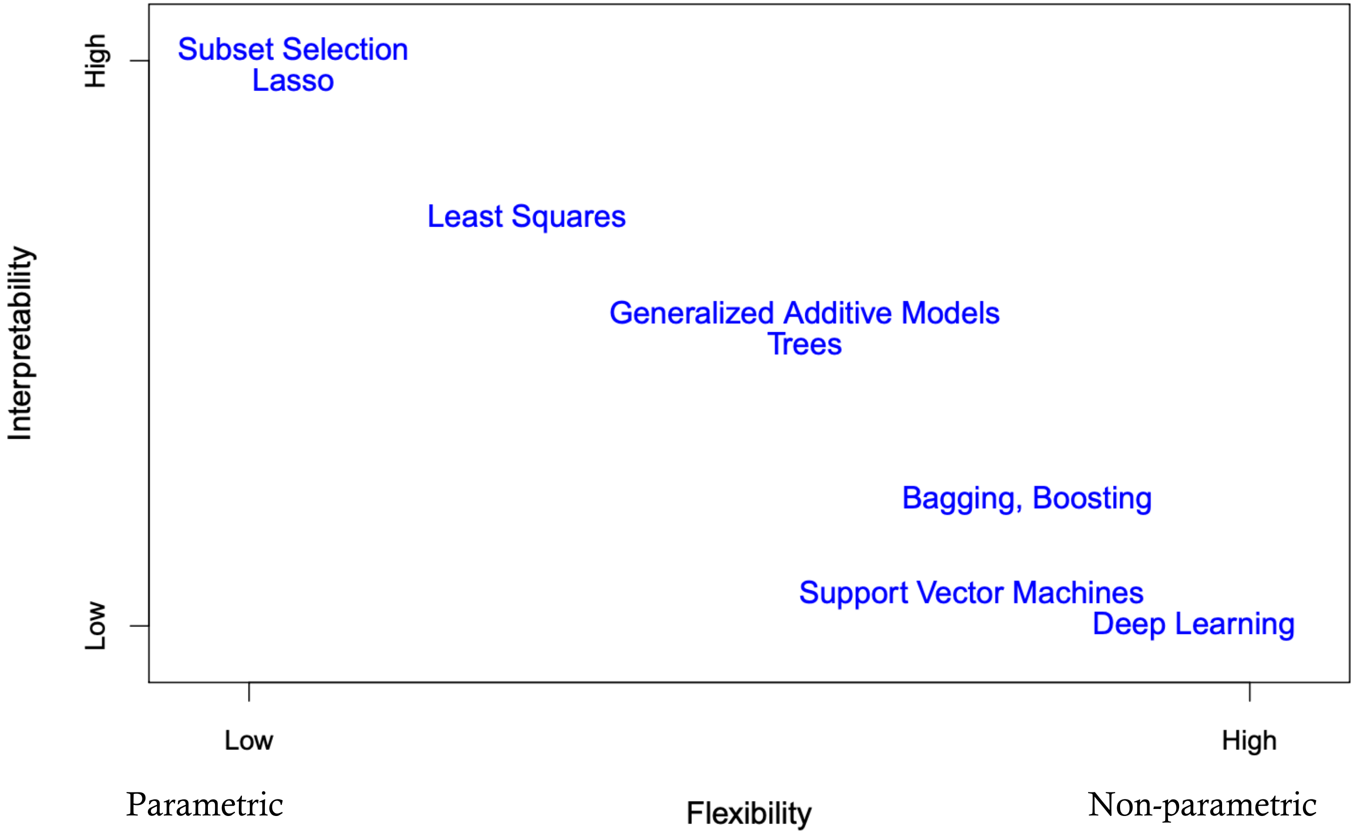
Some adjectives used to describe ML algorithms

- **Distributional assumptions**
  - Generative: assumes a full probabilistic model of the observed data
  - Discriminative: assumes a partial or no probabilistic model
- **Computational form**
  - Linear: produces a linear/affine function
  - Nonlinear: not linear
- **Complexity**
  - Parametric: # of model parameters* is independent of sample size
  - Nonparametric: # of model parameters* grows with sample size
- **Learning Type**
  - Lazy learner: No explicit model is built at training time (usually **instance-based**).
  - Eager learner: A general model is built during training.

*Model parameters = parameters needed to represent the function produced by the learning algorithm

INTERPRETABILITY VS FLEXIBILITY

(Graphic from ISLP)



Parametric

Non-parametric

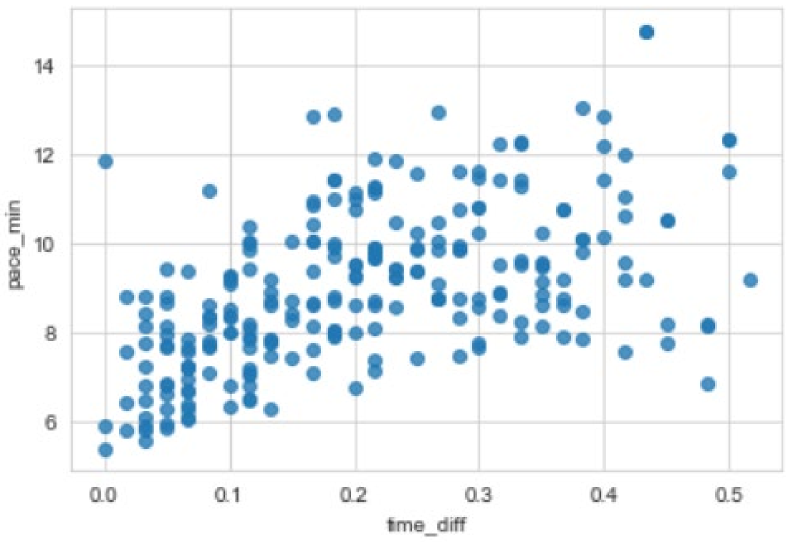
INSTANCE-BASED VS MODEL-BASED

- Instance-Based
    - Learn training data by heart
    - Generalize to new observations based on *similarity* to existing observations
        - Requires a *metric* that quantifies similarity
    - *Example*: k Nearest Neighbors (k-NN)

INSTANCE-BASED VS MODEL-BASED

- Model-Based
  - Estimate the model $f(x)$ that defines the relationship between X and y
  - Predictions are made by introducing new observations into the model: $\hat{y}_{new} = \hat{f}(x_{new})$
  - Most ML algorithms are model-based (e.g., linear regression, neural networks)
  - Model-based algorithms are trained by defining an *objective function* and finding the parameters or values that *optimize* the objective function

EXAMPLE: SIMPLE LINEAR REGRESSION



How does simple linear regression work?

EXAMPLE: SIMPLE LINEAR REGRESSION

- Objective function: (this is a *cost* function since we want to find the *minimum*)

$$L(\beta_0, \beta_1) = \sum_{i=1}^{n} \big(y_i - (\beta_0 + \beta_1 x_i)\big)^2$$

  - We find the $\widehat{\beta_1}$ and $\widehat{\beta_0}$ that minimize the objective function
- Optimization method:
  - *How* do we perform the optimization?

GENERAL MODEL-BASED TRAINING FRAMEWORK

- Define an appropriate **objective function**
  - Cost/Loss functions are minimized
  - Utility/Fitness functions are maximized
    - For example the likelihood in MLE is a utility function
- "Train" the model by solving the optimization problem using an appropriate optimization **method**
  - a "closed-form" solution (rare)
  - a numerical method (usually)

CHALLENGES IN ML

MAIN CHALLENGE OF MACHINE LEARNING

(According to Hands On Machine Learning)

- Insufficient quantity of training data
- Non-representative training data
- Poor-quality data
- Irrelevant features
- Under- and overfitting

MAIN CHALLENGE OF MACHINE LEARNING

- Insufficient quantity of training data
- Non-representative training data (e.g., distribution shift)
- Poor-quality data
- Irrelevant features
- Under- and overfitting

Most of these issues boil down to one idea:
Your results are only as good as your data

THE IMPORTANCE OF DATA

- Data is the fuel that powers our learning machine
- No amount of model improvement can compensate for “bad” data
  - “Bad” data is not always poor quality
  - Sometimes “bad” data is just the wrong data for the task

THINGS TO ASK OF YOUR DATA

- What data is available?
- Is the data still relevant?
- What other data sources might be available?
- Is there *enough* data?
- Where did the data come from?
- Do you know and trust the data source?
- Does the available data actually answer the question?

POTENTIAL BIASES IN DATA

- **Selection bias:** Data is selected for convenience or self-selected
- **Confirmation bias:** We are less critical of data (and results) that confirm our prior beliefs
- **Societal bias:** Data often reflects cultural stereotypes
- **Measurement bias:** Biases due to limitations of measurement devices
- **Proxy measurement bias:** A proxy measurement is used for something that is hard to measure
- **Exclusion bias**: Important variables might not be measured
- **Sampling bias**: Data isn't representative or doesn't reflect realities of where the ML model will run
- **Human bias:** All data is somehow collected by humans