UVA CS 4774: Machine Learning

S6: Lecture 26: Review

Dr. Yanjun Qi

University of Virginia
Department of Computer Science

Final Review

- ❑ Review of covered so far
- ❑ Five Tribes of Machine Learning
- ❑ Four books to recommend

Objective

- To help students be able to build machine learning tools
  - (not just a tool user!!!)

- Key Results:
  - Able to build a few simple machine learning methods from scratch
  - Able to understand a few complex machine learning methods at the source code and equation level

Digital Over Physical

Who has:

Cable or satellite TV

Internet

2+ cell phones

Premium TV (HBO)

Internet TV (Netflix)

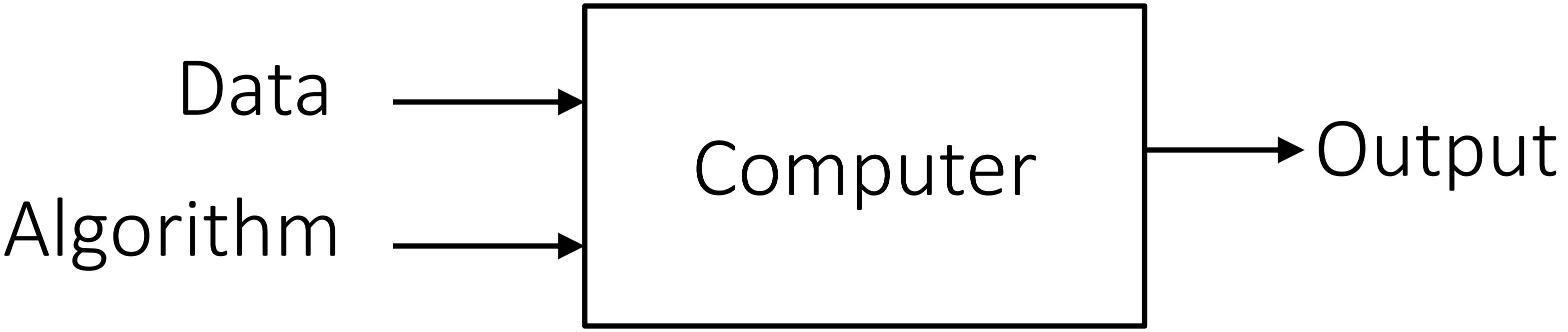
\+ XM Radio

Almost all aspects of planet earth go digital NOW

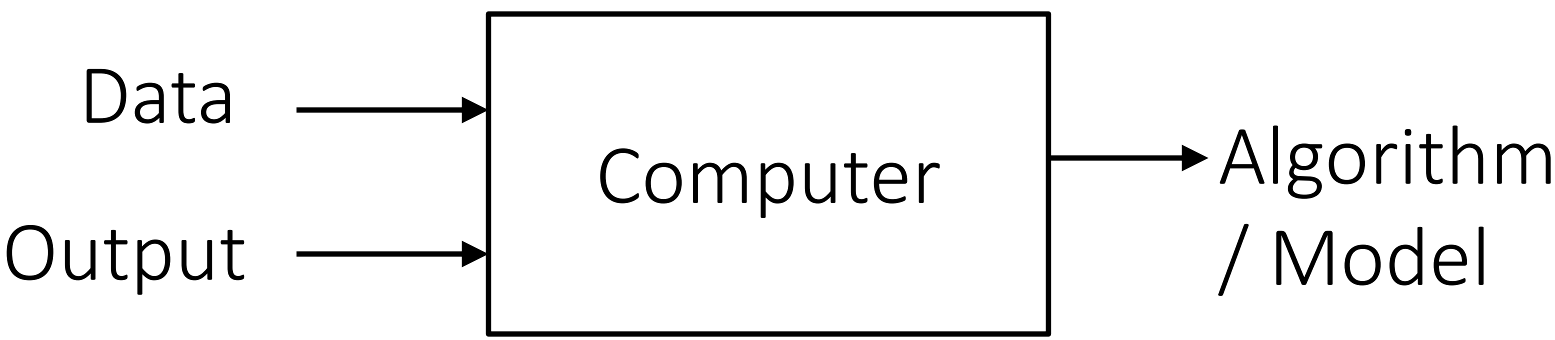
➔ Accessible / Large Amount of Data Samples, Streams, …

From GTNEXUS

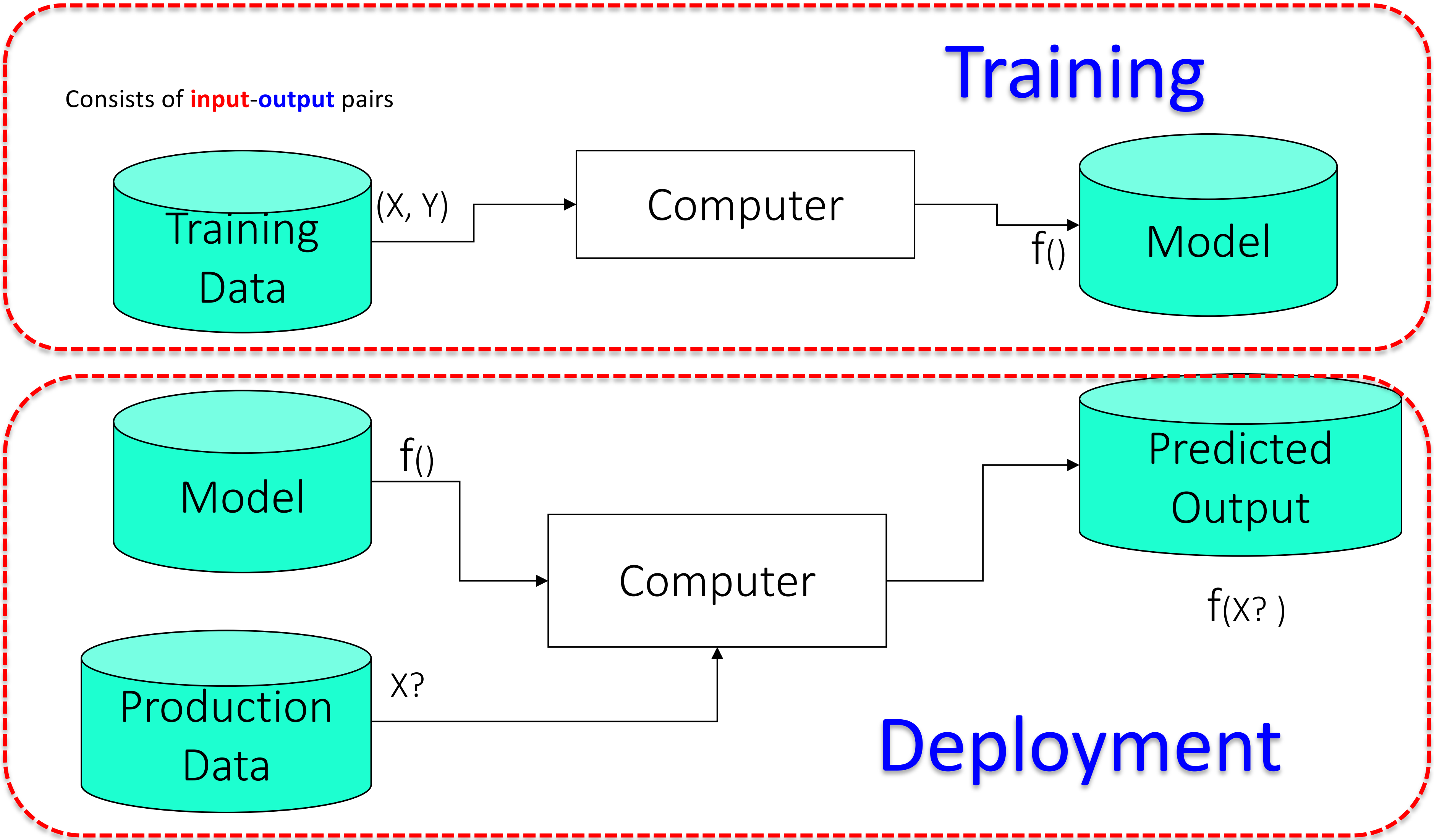
Traditional Programming



Machine Learning



Two Modes of Machine Learning



Machine Learning in a Nutshell

Data
↓
Task
↓
Representation
↓
Score Function
↓
Search/Optimization
↓
Models, Parameters

ML grew out of work in AI

Optimize a performance criterion using example data or past experience,

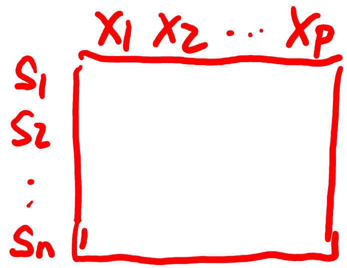
Aiming to generalize to unseen data

Rough Sectioning of this Course

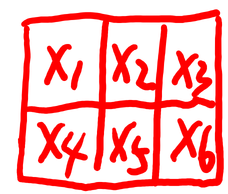
- S1. Basic Supervised Regression + Tabular Data
- S2. Basic Deep Learning + 2D Imaging Data
- S3. Generative and Deep + 1D Sequence Text Data
- S4. Advanced Supervised learning + Tabular Data
- S5. Not Supervised
- S6: Wrap Up + (a few invited tasks, e.g. on AWS)

Course Content Plan ➔ Regarding Data

- ❑ Tabular / Matrix



- ❑ 2D Grid Structured: Imaging



- ❑ 1D Sequential Structured: Text

- ❑ Graph Structured (Relational)

- ❑ Set Structured / 3D /

Course Content Plan ➔ Regarding Tasks

- ❑ Regression (supervised) — Y is a continuous
- ❑ Learning theory — About f()
- ❑ Classification (supervised) — Y is a discrete
- ❑ Unsupervised models — NO Y
- ❑ Graphical models — About interactions among Y,X1,. Xp
- ❑ Reinforcement Learning — Learn to Interact with environment

Three major sections for classification

- We can divide the large variety of classification approaches into roughly three major types

1. Discriminative
   - directly estimate a decision rule/boundary
   - e.g., logistic regression, neural networks
   - e.g., support vector machine, decisionTrees

2. Generative:
   - build a generative statistical model
   - e.g., naïve bayes classifier, Bayesian networks

3. Instance based classifiers
   - Use observation directly (no models)
   - e.g. K nearest neighbors

Selected Deep Trends

https://qdata.github.io/deep2Read/

Module I: Data → Representation f()

1. DNN on graphs / trees / sets
2. NTM 4program induction

Representation f():

1. CNN / Residual / Memory
2. RNN / Attention / Seq2Seq / Transformer …

Module II: Task

1. Deep Generative models/ DeepFake
2. Deep reinforcement learning
3. Few-shots / Meta learning / AGI?

Module III: Training Workflow: Score Function → Search/Optimization

1. Autoencoder / self supervised training
2. Generative Adversarial Networks (GAN)
3. Learning to optimize /Learning to search architecture (AutoML)

Module IV: On Deployment: Models, Parameters, Hyperparameter, Metrics → Hardware

Models, Parameters, Hyperparameter, Metrics:

1. Validate / Evade / Test / Verify
2. Understand DNNs

Hardware:

1. Model Compression / Efficient Net

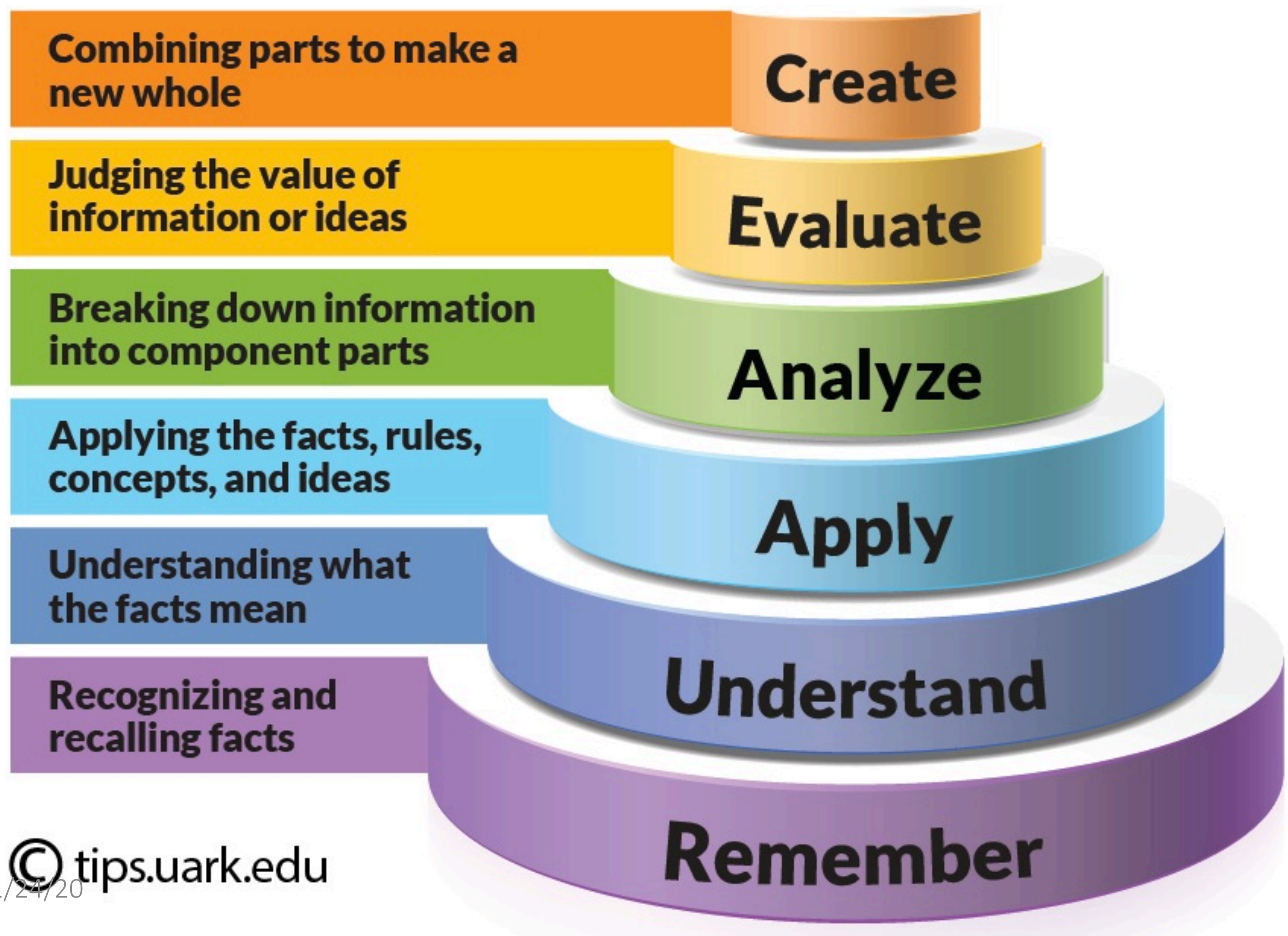
What we have covered (more)

- Learning theory / Model selection
  - K-folds cross validation / Model Selection
  - Expected prediction error
  - Bias and variance tradeoff (overfit / underfit)
  - Generative vs. Discriminative Classifiers
  - Remedy when Overfit / Underfit
    - Control / adjust model complexity, capacity
    - Control / adjust training size
  - Three plots:
    - Train / Vali Loss vs. Epochs
    - Train / Vali Loss vs. hyperparameter Values
    - Train / Vali Loss vs. Varying Size of Trainin

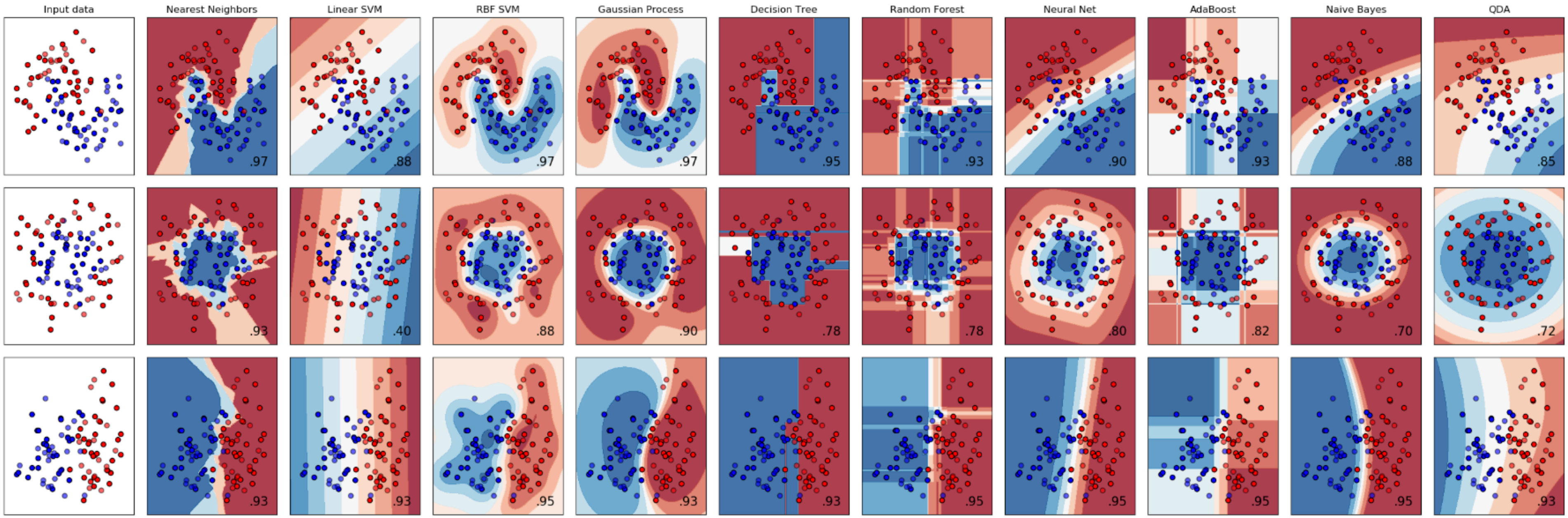
What we have covered for each component

Data	Tabular, 1-D sequential, 2-D Grid like Imaging, 3-D VR, Graph, Set
Task	Regression, classification, clustering, dimen-reduction
Representation	Linear func, nonlinear function (e.g. polynomial expansion), local linear, logistic function (e.g. p(c|x)), tree, multi-layer, prob-density family (e.g. Bernoulli, multinomial, Gaussian, mixture of Gaussians), local func smoothness, kernel matrix, local smoothness, partition of feature space,
Score Function	MSE, Margin, log-likelihood, EPE (e.g. L2 loss for KNN, 0-1 loss for Bayes classifier), cross-entropy, cluster points distance to centers, variance, conditional log-likelihood, complete data-likelihood, regularized loss func (e.g. L1, L2) , goodness of inter-cluster similar
Search/ **Optimization**	Normal equation, gradient descent, stochastic GD, Newton, Linear programming, Quadratic programming (quadratic objective with linear constraints), greedy, EM, asyn-SGD, eigenDecomp, backprop
Models, Parameters	Linear weight vector, basis weight vector, local weight vector, dual weights, training samples, tree-dendrogram, multi-layer weights, principle components, member (soft/hard) assignment, cluster centroid, cluster covariance (shape), …

My Teaching Guide: Bloom's Taxonomy on Cognitive Learning



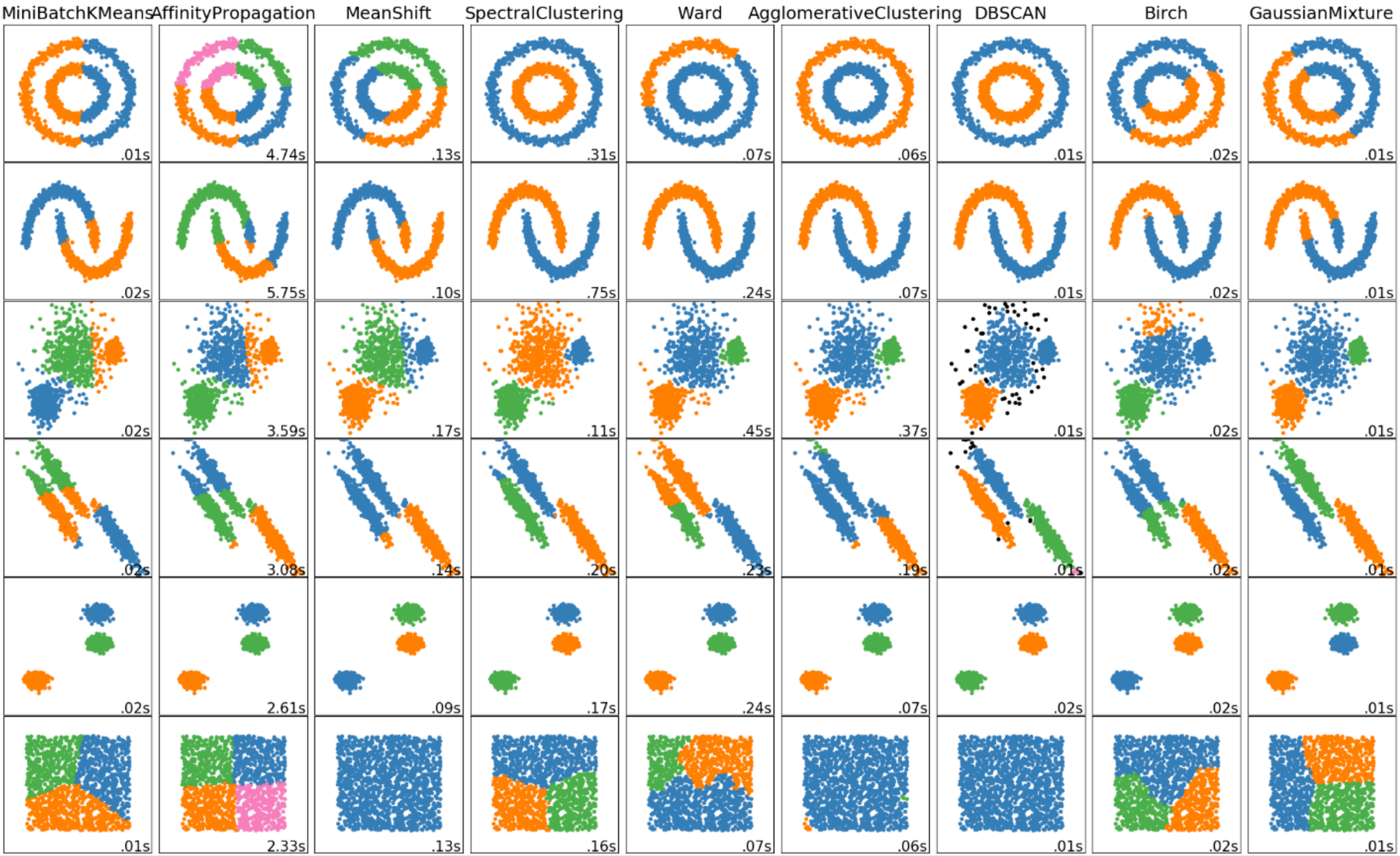
https://scikit-learn.org/stable/auto_examples/classification/plot_classifier_comparison.html



- ✓ different assumptions on data
- ✓ different scalability profiles at training time
- ✓ different latencies at prediction (test) time
- ✓ different model sizes (embedability in mobile devices)
- ✓ different level of model interpretability / robustness

Adapt from Olivier Grisel's talk

https://scikit-learn.org/stable/auto_examples/cluster/plot_cluster_comparison.html



- ✓different assumptions on data
- ✓different scalability profiles
- ✓different model sizes (embedability in mobile devices)

Final Review

- ❑ Review of covered so far
- ❑ Five Tribes of Machine Learning
- ❑ Four books to recommend

Highly Recommend One Book:
0. By Dr. Domingos: Master Algorithm

So How Do Computers Discover New Knowledge?

1. **Symbolists**--Fill in gaps in existing knowledge
2. **Connectionists**--Emulate the brain
3. **Evolutionists**--Simulate evolution
4. **Bayesians**--Systematically reduce uncertainty
5. **Analogizers**--Notice similarities between old and new

SRC: Pedro Domingos ACM Webinar Nov 2015
http://learning.acm.org/multimedia.cfm

The Five Tribes of Machine Learning:

Tribe	Origins	Key Algorithm
Symbolists	Logic, philosophy	Inverse **deduction**
Connectionists	Neuroscience	Backpropagation
Evolutionists	Evolutionary **biology**	Genetic programming
Bayesians	Statistics	Probabilistic **inference**
Analogizers	Psychology	Kernel machines

SRC: Pedro Domingos ACM Webinar Nov 2015
http://learning.acm.org/multimedia.cfm

From: Dr. Pedro Domingos

Symbolists



Tom Mitchell Steve Muggleton Ross Quinlan

Tribe	Origins	Key Algorithm
Symbolists	Logic, philosophy	Inverse **deduction**

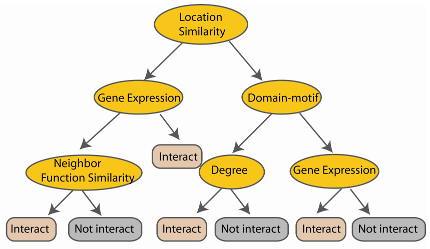
From: Dr. Pedro Domingos

e.g., Decision Tree-building algorithms (1990s)

ID3: Iterative Dichotomiser 3. Developed in the 80s by Ross Quinlan.

C4.5: Successor of ID3, also developed by Quinlan ('93). Main improvements over I3D:

Adaboost: by Robert Schapire (1999)



Connectionists

Yann LeCun

Geoff Hinton

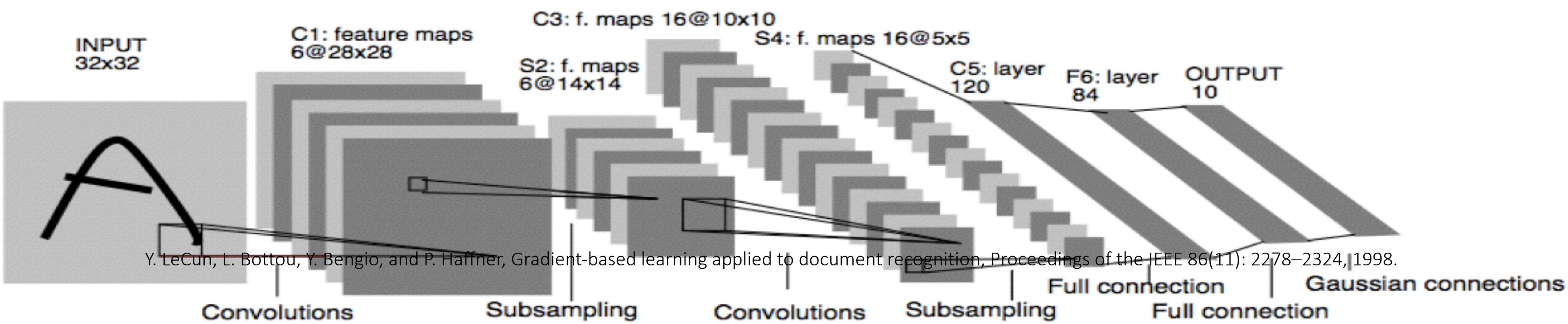
Yoshua Bengio

Tribe	Origins	Key Algorithm
Connectionists	Neuroscience	Backpropagation

From: Dr. Pedro Domingos

Deep Learning (CNN) in the 90's

- Prof. Yann LeCun invented Convolutional Neural Networks (CNN) in 1998
- First NN successfully trained with many layers



Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE 86(11): 2278–2324, 1998.

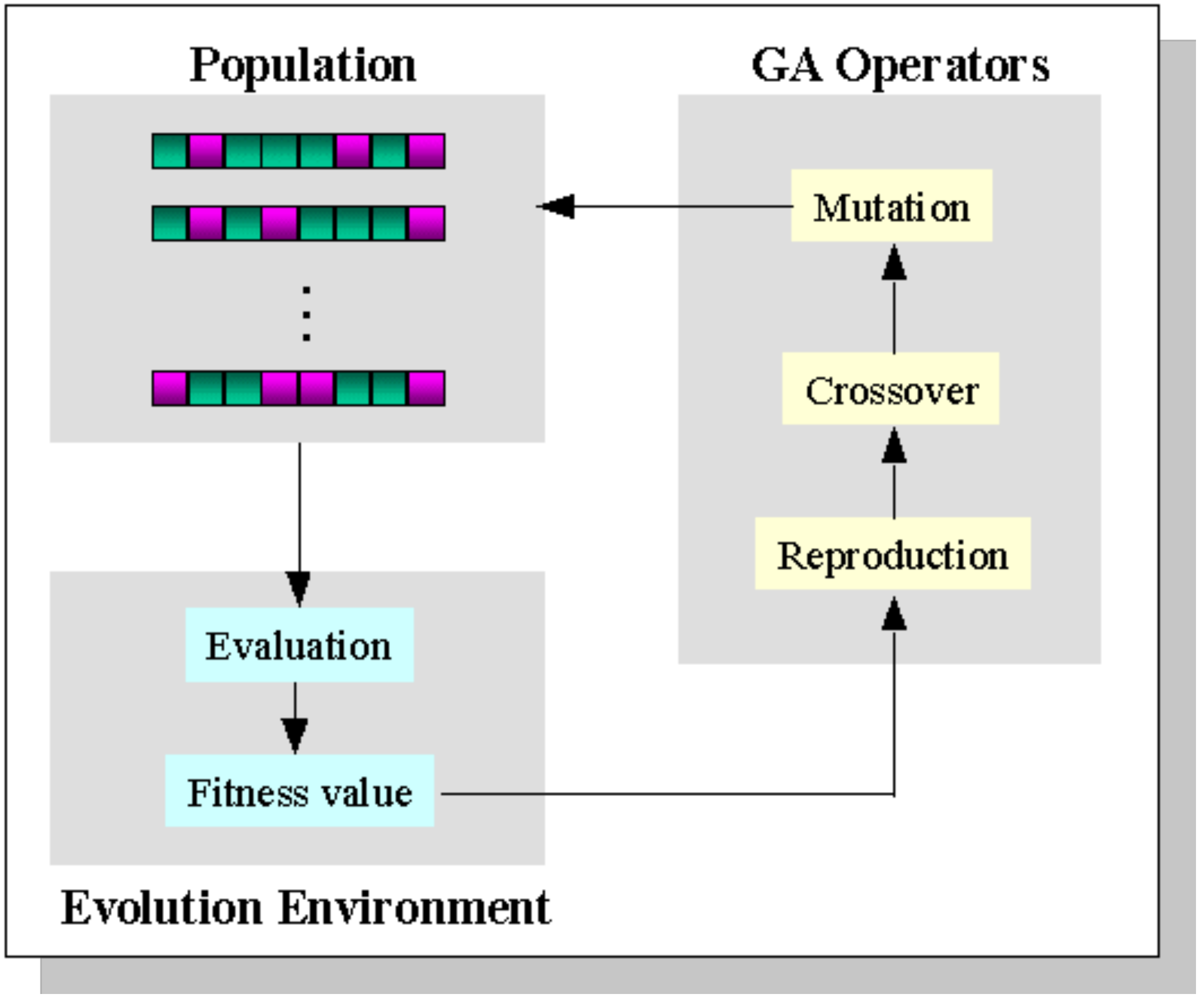
Evolutionaries

John Koza John Holland Hod Lipson

Tribe	Origins	Key Algorithm
Evolutionists	Evolutionary **biology**	Genetic programming

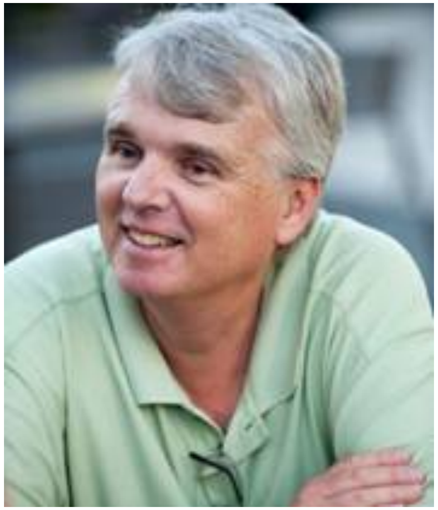
From: Dr. Pedro Domingos

Genetic Algorithms



From: Dr. Pedro Domingos

Bayesians

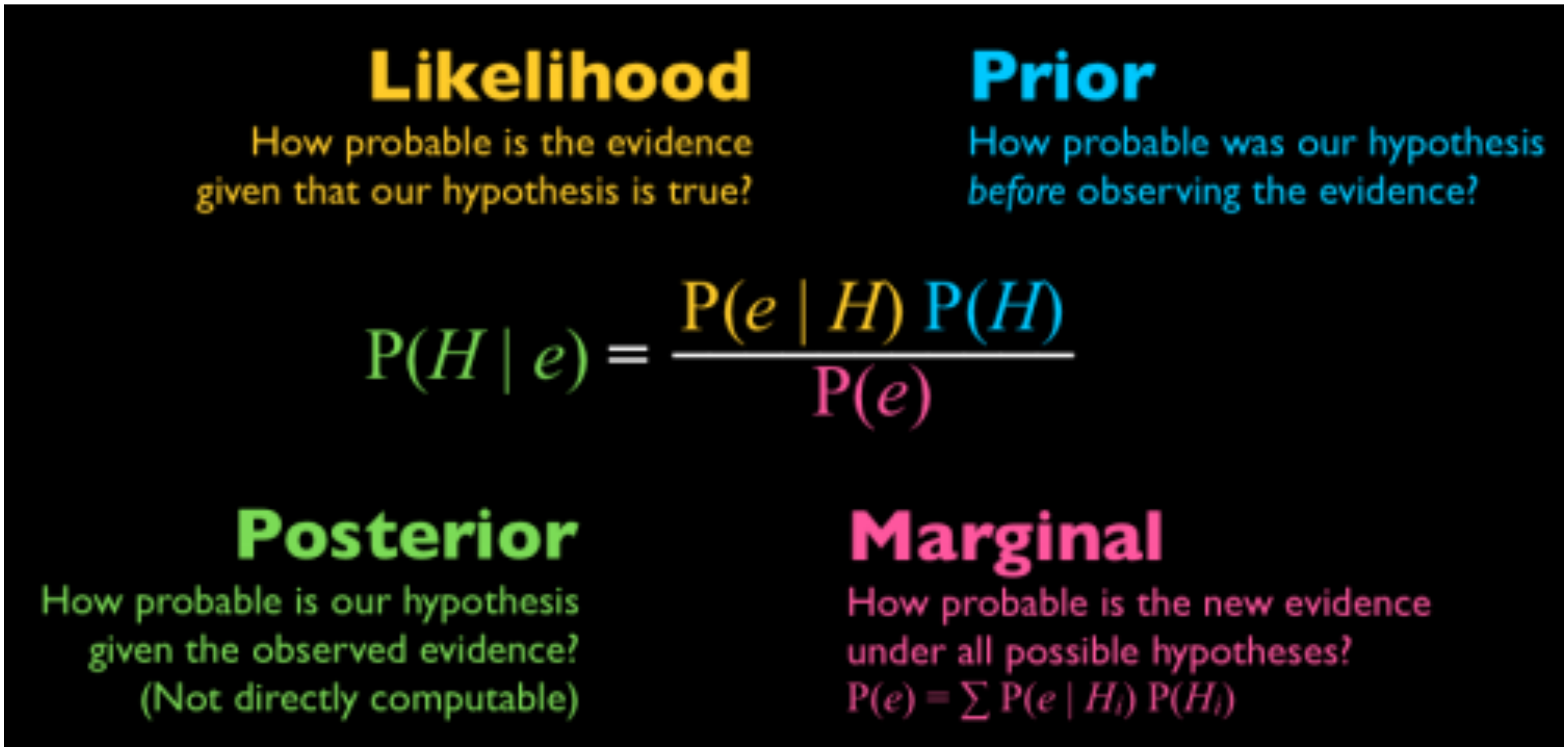


David Heckerman | Judea Pearl | Michael Jordan

Tribe	Origins	Key Algorithm
Bayesians	Statistics	Probabilistic **inference**

From: Dr. Pedro Domingos

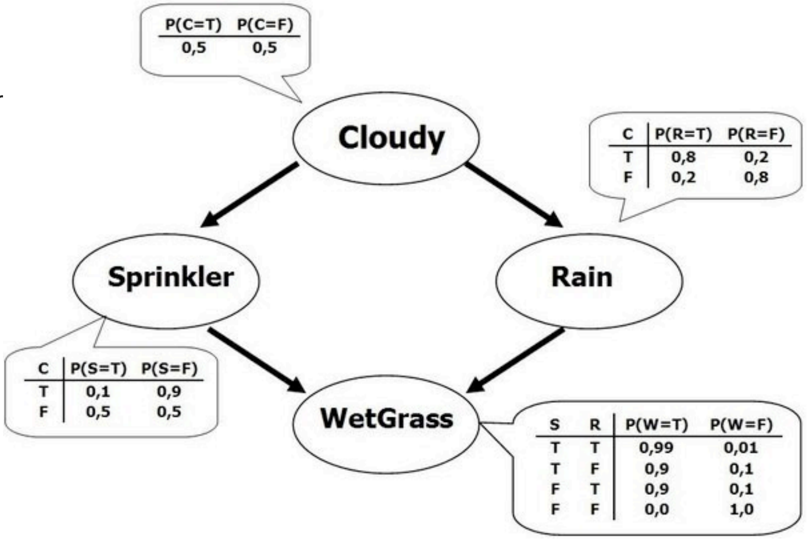
Probabilistic Inference



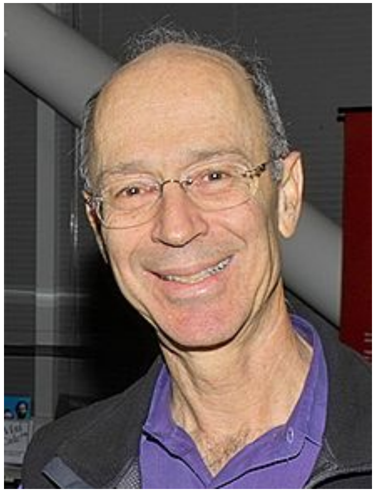
From: Dr. Pedro Domingos

Reasoning with uncertainty

- “Bayesian network” was termed by Judea Pearl in 1985
- Bayes' conditioning is the basis for updating information in the graph
- The distinction between causal and evidential modes of reasoning
- In the late 1980s, established as a field of study.
    - Pearl's Probabilistic Reasoning in Intelligent Systems
    - Neapolitan's Probabilistic Reasoning in Expert Syster

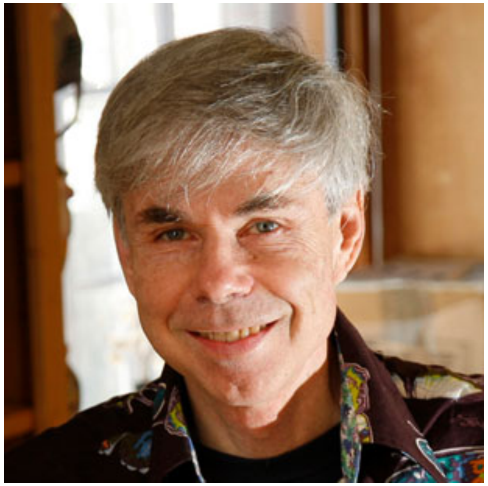


Analogizers



Peter Hart

Vladimir Vapnik



Douglas Hofstadter

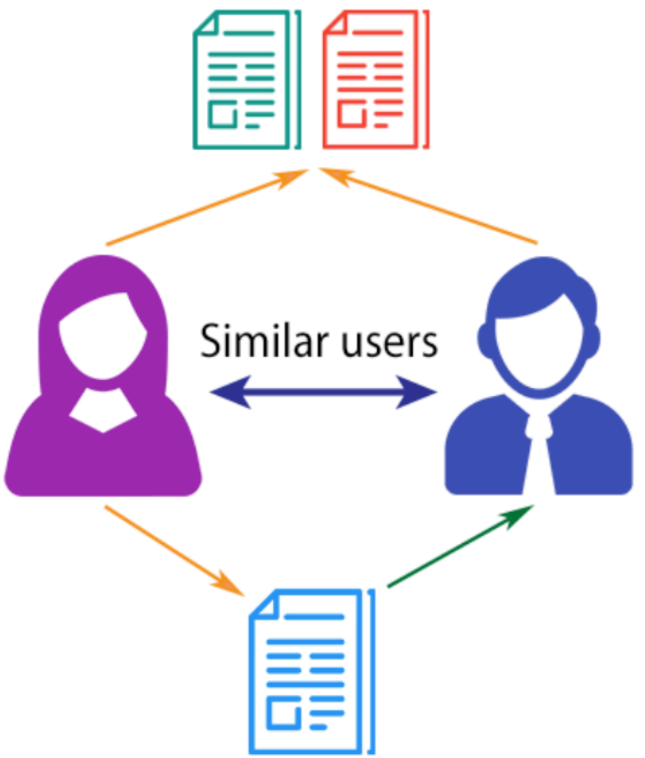
Tribe	Origins	Key Algorithm
Analogizers	Psychology	Kernel machines

From: Dr. Pedro Domingos

A little bit History

- **SVM** : first introduced in 1992, popular because of its success in handwritten digit recognition (1994); Regarded as an important example of "kernel methods"

- **Recommender Systems:**
  – E.g., Matrix Factorization



Recommender Systems

From: Dr. Pedro Domingos

The Big Picture

Tribe	Problem	Origins	Solution	Module in Nutshell
Symbolists	Knowledge composition	Logic, philosophy	Inverse deduction	Representations;
Connectionists	Credit assignment	Neuroscience	Backpropagation	Representations; **Numerical Optimization**
Evolutionaries	Search Structure discovery	Evolutionary **biology**	Genetic programming	Discrete Optimization;
Bayesians	Uncertainty	Statistics	Probabilistic inference	Likelihood type Score function;
Analogizers	Similarity	Psychology	Kernel machines	Representations; Reconstruction loss

Final Review

- ❑ Review of covered so far
- ❑ Five Tribes of Machine Learning
- ❑ Four books to recommend

Highly Recommend One Book:
0. By Dr. Domingos: Master Algorithm

So How Do Computers Discover New Knowledge?

1. **Symbolists**--Fill in gaps in existing knowledge
2. **Connectionists**--Emulate the brain
3. **Evolutionists**--Simulate evolution
4. **Bayesians**--Systematically reduce uncertainty
5. **Analogizers**--Notice similarities between old and new

SRC: Pedro Domingos ACM Webinar Nov 2015
http://learning.acm.org/multimedia.cfm

Highly Recommend
Four Extra-curriculum books

- 1. Book - Algorithms to Live By: The Computer Science of Human Decisions
  - https://books.google.com/books/about/Algorithms_to_Live_By_The_Computer_Scien.html?id=xmeJCgAAQBAJ&source=kp_book_description
  - This book provides a fascinating exploration of how computer algorithms can be applied to our everyday lives.

http://qiyanjun.github.io/books2read//

Highly Recommend Four Extra-curriculum books

- 2. Book: **So Good They Cannot Ignore You-**
  - https://www.amazon.com/Good-They-Cant-Ignore-You/dp/1455509124
  - The idea of Career capital - rare and valuable skills need deliberate practice
  - 10,000 hours of deliberate practice ➔ Expert!

Highly Recommend
Four Extra-curriculum books

- 3. Book: **Ego Is the Enemy** by RYAN HOLIDAY 2016
  - https://www.amazon.com/Ego-Enemy-Ryan-Holiday/dp/1591847818
  - Don’t get fancy. Ego turns minor accomplishments into major events. ...Stay humble through your work.
  - Work! While aspiring, the most important thing you can do to fight your ego is to focus on creating value. Sit down and put in the hours. Invest in yourself by thinking long term.

Highly Recommend Four Extra-curriculum books

- 4. Book: Homo Deus- A Brief History of Tomorrow
  - https://www.goodreads.com/book/show/31138556-homo-deus
  - "Homo Deus explores the projects, dreams and nightmares that will shape the twenty-first century—from overcoming death to creating artificial life. It asks the fundamental questions: Where do we go from here? And how will we protect this fragile world from our own destructive powers? This is the next stage of evolution. This is Homo Deus.""
  - Keep reinventing ourselves in an era of uncertainty !

References

- Hastie, Trevor, et al. The elements of statistical learning. Vol. 2. No. 1. New York: Springer, 2009.
- Prof. Domingos' slides
- Prof. Andrew Ng's slides
- Many wonderful books from Audible