INTRODUCTION

DATA/MSML 603: Principles of Machine Learning

SCIENCE ACADEMY

scienceacademy.umd.edu

Machine Learning

- Question: What is Machine Learning? What comes to your mind?

Source: https://bestarion.com/what-is-artificial-intelligence/

SCIENCE ACADEMY

Machine Learning

"Learning is any process by which a system improves performance from experience."

Herbert Simon, *Nobel Prize winner in economics and leader in AI*

SCIENCE ACADEMY

Machine Learning

Definition by Tom Mitchell (1998):

- Machine Learning is the study of algorithms that:
  - Improve their performance (P)
  - At some task (T)
  - With experience (E)
  - A well-defined learning task is given by <P,T,E>

SCIENCE ACADEMY

Machine Learning (1/4)

Traditional programming:

Machine learning:

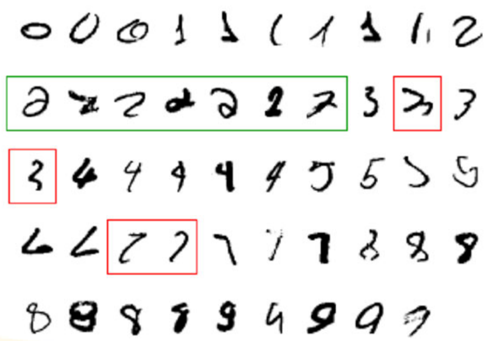
Machine Learning (2/4)

- When do we need machine learning?
  - Human expertise does not exist (navigating on Mars)
  - Human cannot explain their expertise (speech recognition)
  - Models must be customized (personalized medicine)
  - Models are based on huge amounts of data (genomics)

Machine Learning (3/4)

A classic example of a task that requires machine learning

- It is rather hard to distinguish what is a `2'



Machine Learning (4/4)

Some more examples of tasks that are best solved by a learning algorithm:

- **Recognizing patterns**
  - Facial identities or facial expressions
  - Handwritten or spoken words
  - Medical images
- **Generating patterns**
  - Generating images or motion sequences
- **Recognizing anomalies**
  - Unusual credit card transactions
  - Unusual patterns of sensor readings in a nuclear power plant
  - Unusual activities in a network

SCIENCE ACADEMY

© 2020, UNIVERSITY OF MARYLAND

DATA/MSML 603

Sample Applications

- Web search
- Computational biology
- Finance
- E-commerce
- Space exploration and exoplanet discovery
- Robotics
- Information extraction
- Social networks
- Debugging software
- Network management

SCIENCE ACADEMY

DATA/MSML 603

Samuel's Checkers-Player

"Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed." –Arthur Samuel (1959)

SCIENCE ACADEMY

© 2020, UNIVERSITY OF MARYLAND

DATA/MSML 603

Defining the Learning Task (1/2)

- Improve on task **T**, with respect to performance metric **P**, based on experience **E**
  - T: Playing checkers
  - P: Percentage of games won against an arbitrary opponent
  - E: Playing practice games against itself

  - T: Recognizing hand-written words
  - P: Percentage of words correctly classified
  - E: Database of human-labeled images of handwritten words

Defining the Learning Task (2/2)

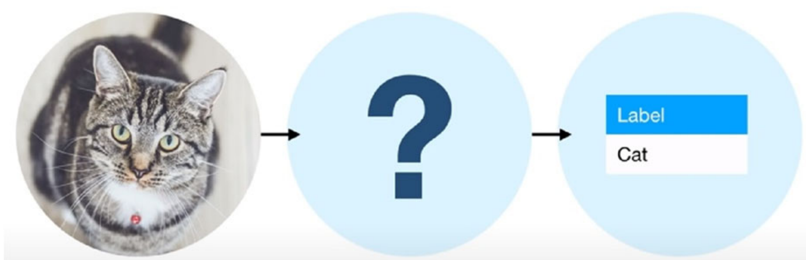
- Improve on task **T**, with respect to performance metric **P**, based on experience **E**
  - T: Driving on four-lane highways using vision sensors
  - P: Average distance traveled before a human-judged error
  - E: A sequence of images and steering commands recorded while observing a human driver

SCIENCE ACADEMY
© 2020, UNIVERSITY OF MARYLAND
DATA/MSML 603

Two Key Problems In Machine Learning

Two Kinds of Problem (1/4)

- According to the types of output ground truth
- Classification: the output ground truth is discretized values (or labels)



Input Model Output

Two Kinds of Problem (2/4)

- According to the types of output ground truth
- Classification: the output ground truth is discretized values (or labels)
- Regression: the output ground truth is continuous values

Training Data

Age (days)	Weight (grams)
2	49
12	122
8	74
21	205
4	80

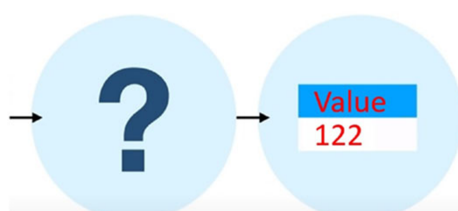
Test Data

Age (days)	Weight (grams)
18	?

Two Kinds of Problem (3/4)

- According to the types of output ground truth
- Classification: the output ground truth is discretized values (or labels)
- Regression: the output ground truth is continuous values

18 days

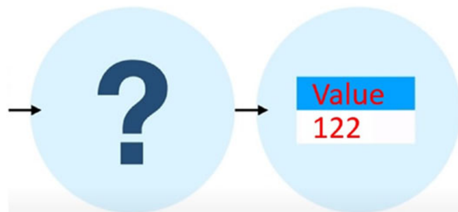


Input Model Output

Two Kinds of Problem (4/4)

- According to the types of output ground truth
- Classification: the output ground truth is discretized values (or labels)
- Regression: the output ground truth is continuous values

18 days;
Food type;
Sunshine;
....



Input Model Output

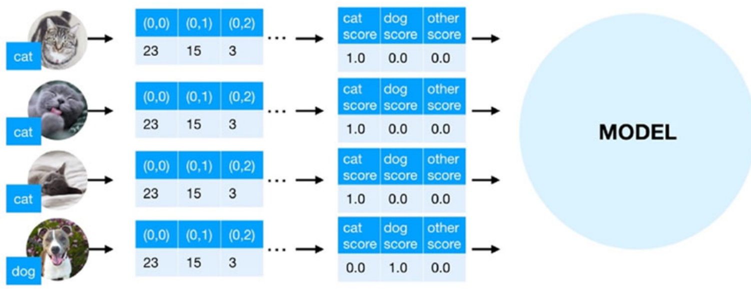
SCIENCE ACADEMY

Two Procedures

SCIENCE ACADEMY

Machine Learning

What should we do to use the models in either classification or regression?



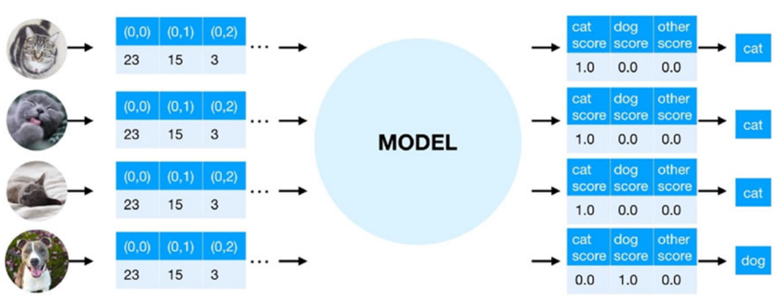
Two procedures

1: Training: learn the parameters in a model

SCIENCE ACADEMY

Machine Learning

What should we do to use the models in either classification or regression?

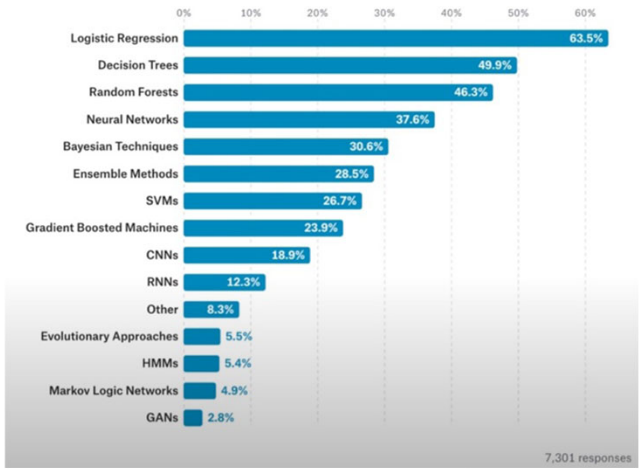


Two procedures

2: Testing: use an unseen input to go through the model with the learned parameters

SCIENCE ACADEMY

Kaggle Survey: Most Common ML Models



Source: https://www.vidora.com/general/kaggle-survey/

SCIENCE ACADEMY

DATA/MSML 603

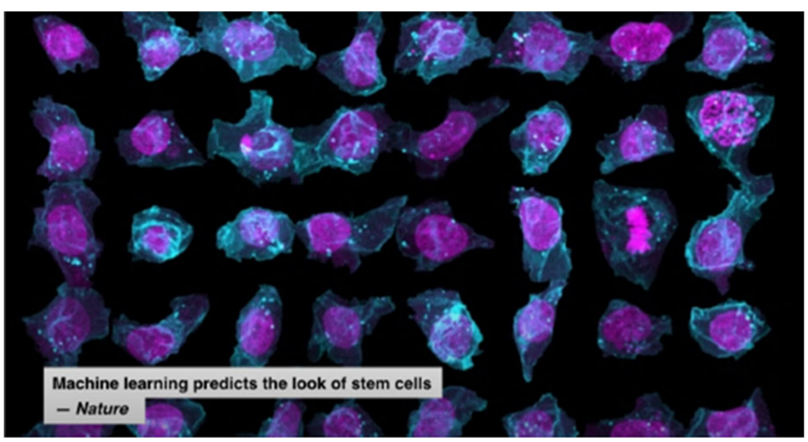
State of the Art Applications of Machine Learning

SCIENCE ACADEMY

scienceacademy.umd.edu

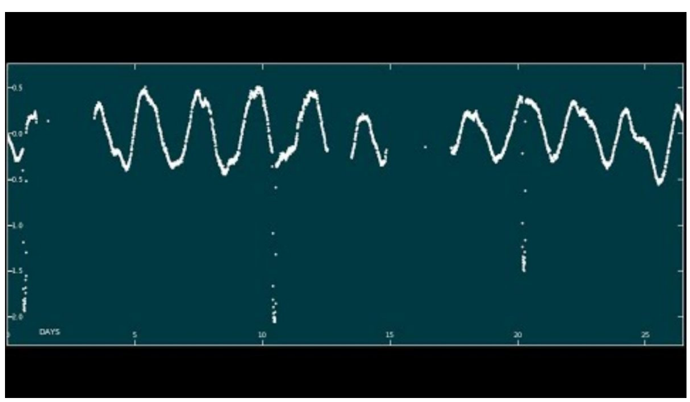
Security

Biology



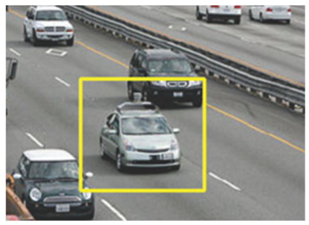
Asteroid Tracking

Planet Hunters



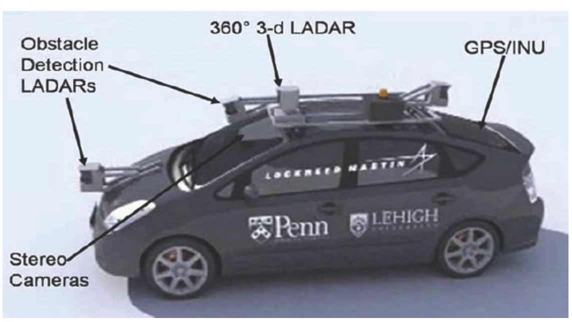
Autonomous Cars

- Nevada made it legal for autonomous cars to drive on roads in June 2011
- As of August 2024, 24 states allow driverless operations (in seven states, no human driver is required with SAE Level 4 or 5 capable AI)



Autonomous Car Sensors

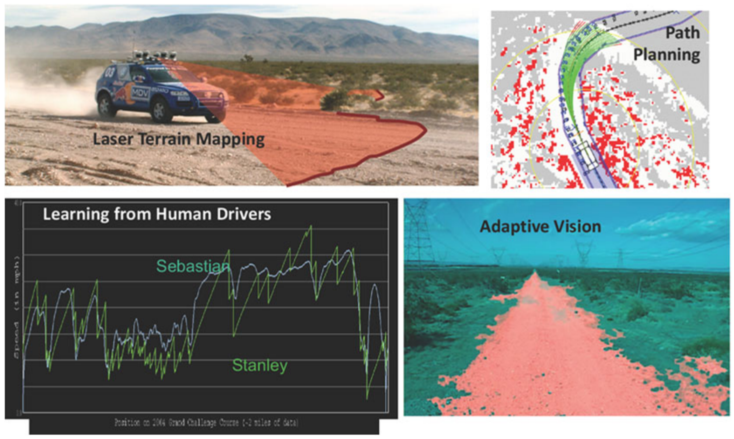
- Penn's autonomous car (Ben Franklin Racing Team)



https://vaderlab.wordpress.com/the-ben-franklin-racing-team/

SCIENCE ACADEMY
© 2020, UNIVERSITY OF MARYLAND
DATA/MSML 603

Autonomous Car Technology



Images and video source: Sebastian Thrun's multimedia website

DATA/MSML 603

Autonomous Car

Train in Germany, Test in The USA: Making 3D Object Detectors Generalize

CVPR2020

Yan Wang*[1] Xiangyu Chen*[1] Yurong You[1] Li Erran Li[2,3]
Bharath Hariharan[1] Mark Campbell[1] Kilian Q. Weinberger[1] Wei-Lun Chao[4]
[1]Cornell University [2]Scale AI [3]Columbia University [4]The Ohio State University
{yw763, xc429, yy785, bh497, mc288, kqw4}@cornell.edu erranlli@gmail.com chao.209@osu.edu

Abstract

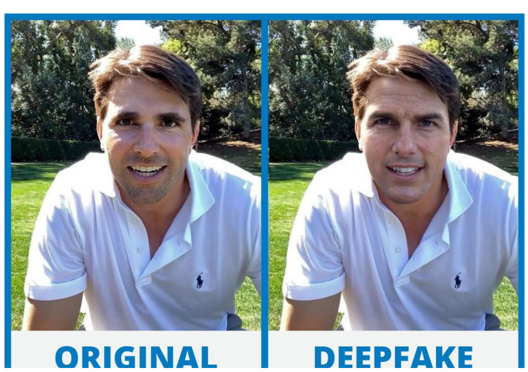
In the domain of autonomous driving, deep learning has substantially improved the 3D object detection accuracy for LiDAR and stereo camera data alike. While deep networks are great at generalization, they are also notorious to over-fit to all kinds of spurious artifacts, such as brightness, car sizes and models, that may appear consistently throughout the data. In fact, most datasets for autonomous driving are collected within a narrow subset of cities within one country, typically under similar weather conditions. In this paper we consider the task of adapting 3D object detectors from one dataset to another. We observe that naïvely, this appears to be a very challenging task, resulting in drastic drops in accuracy levels. We provide extensive experiments to investigate the true adaptation challenges and

KITTI
Argoverse
nuScenes

DATA/MSML 603

Deep Learning in the Headlines

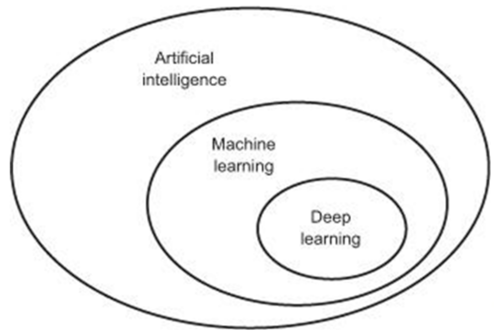
Deep Learning in the Headlines



(source: trymaverick.com)

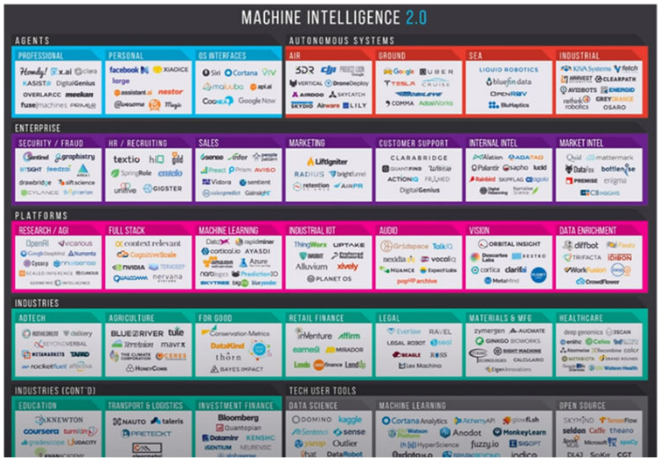
Machine Learning and Deep Learning

- AI is the most general
- Many types of ML algorithms:
  - Nearest Neighbor
  - Naive Bayes
  - Decision Trees
  - Linear Regression
  - Support Vector Machines (SVMs)
  - Neural Networks
- Deep learning is associated with neural networks



Commercial Area

Commercial Area



SCIENCE ACADEMY

DATA/MSML 603

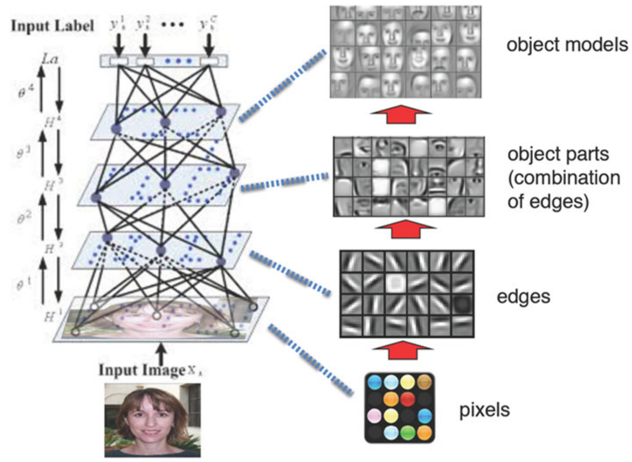
Generative models

- More cool generative models
  - 1D-Text-Visual answering
  - 2D-Image-Style face
  - 3D-Video-Fake face
  - 3D-3D shapes or scene
  - 4D-3D video

SCIENCE ACADEMY

DATA/MSML 603

Deep Belief Net on Face Images

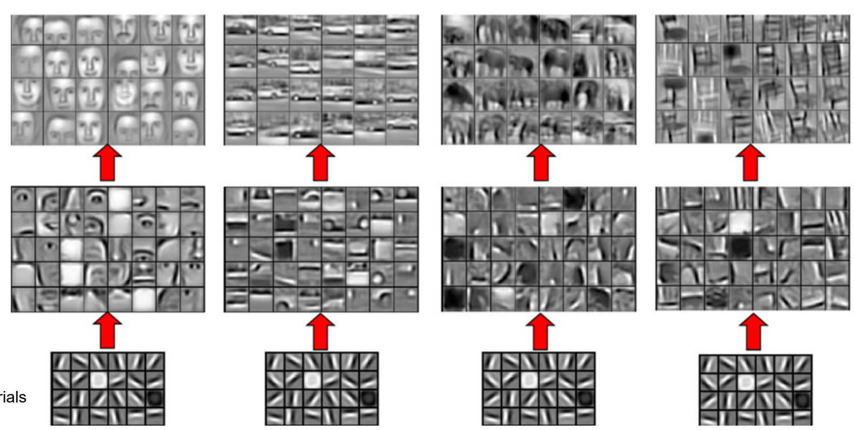


Based on materials
by Andrew Ng

SCIENCE ACADEMY

DATA/MSML 603

Learning of Object Parts



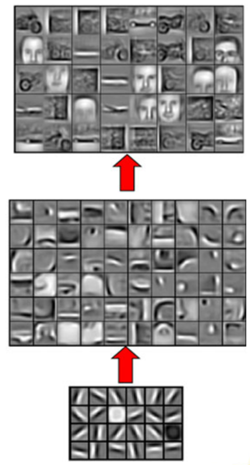
Based on materials
by Andrew Ng

SCIENCE ACADEMY

DATA/MSML 603

Training on Multiple Objects

- Trained on 4 classes (cars, faces, motorbikes, airplanes).
  - Second layer: Shared-features and object-specific features.
  - Third layer: More specific features.



SCIENCE ACADEMY

Scene Labeling via Deep Learning

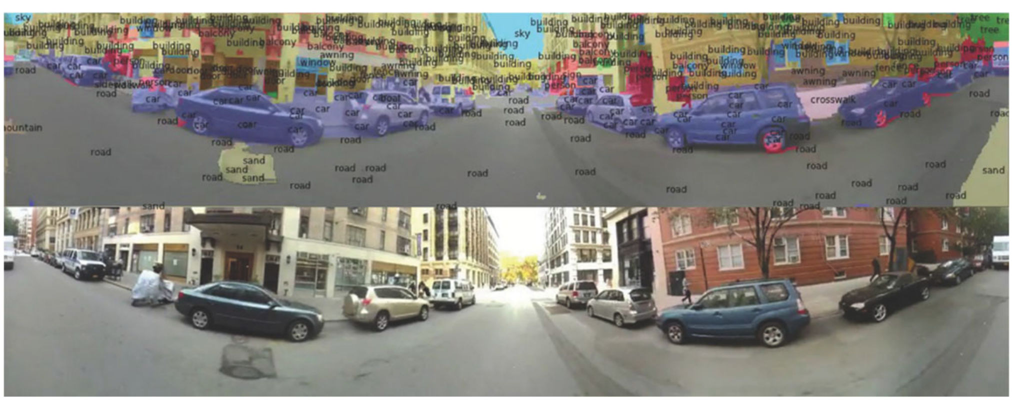


Image source: Farabet *et al.* ICML 2012, PAMI 2013

SCIENCE ACADEMY

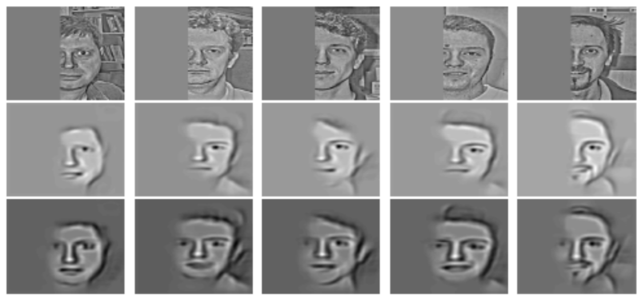
Inference From Deep Learned Models

- Generating posterior samples from faces by “filling in” experiments (cf. Lee and Mumford, 2003). Combine bottom-up and top-down inference.

Input images

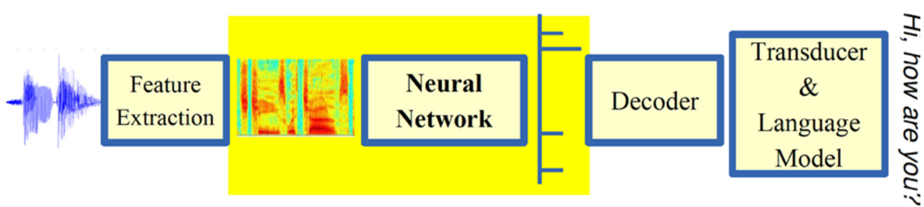
Samples from feedforward Inference (control)

Samples from Full posterior inference

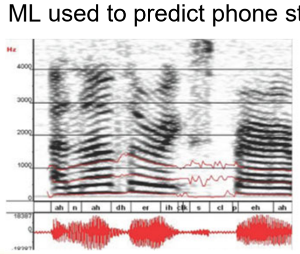


SCIENCE ACADEMY

Typical Speech Recognition System



ML used to predict phone states from the sound spectrogram



Deep learning has state-of-the-art results

# Hidden Layers	1	2	4	8	10	12
Word Error Rate %	16.0	12.8	11.4	10.9	11.0	11.1

Baseline GMM performance = 15.4%

[Zeiler et al. “On rectified linear units for speech recognition” ICASSP 2013]

SCIENCE ACADEMY

Impact of Deep Learning in Speech Technology

(https://www.cartoonmovement.com/cartoon/chatgpt-and-others)

SCIENCE ACADEMY

Types of Learning

SCIENCE ACADEMY

Types of Learning

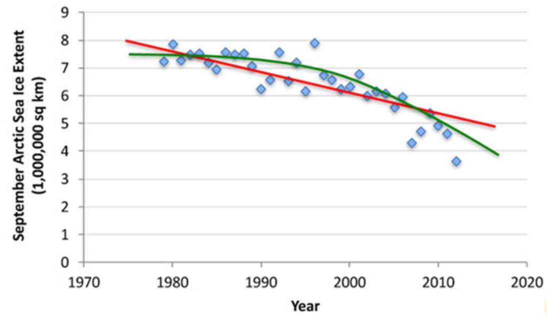
- **Supervised (inductive) learning**
  - Given: training data + desired outputs (labels)
- **Unsupervised learning**
  - Given: training data (without desired outputs)
- **Semi-supervised learning**
  - Given: training data + a few desired outputs
- **Reinforcement learning**
  - Rewards from sequence of actions

Supervised Learning

Machine Learning (1/4)

Supervised vs. Unsupervised

- **Supervised learning:** Given training data + desired output (labels)
- Given (x1, y1), (x2, y2), (x3, y3), …, (xn, yn)
- Learn a function f(x) to predict y given x
- **y is real-valued == regression**



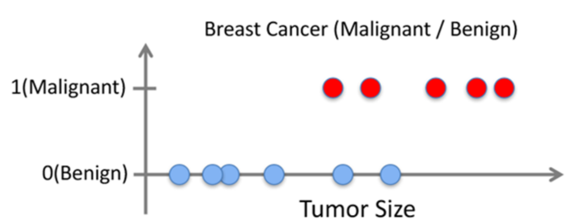
SCIENCE ACADEMY

DATA/MSML 603

Machine Learning (2/4)

Supervised vs. Unsupervised

- **Supervised learning:** Given training data + desired output (labels)
- Given (x1, y1), (x2, y2), (x3, y3), …, (xn, yn)
- Learn a function f(x) to predict y given x
- **y is categorical == classification**



Based on materials by Andrew Ng

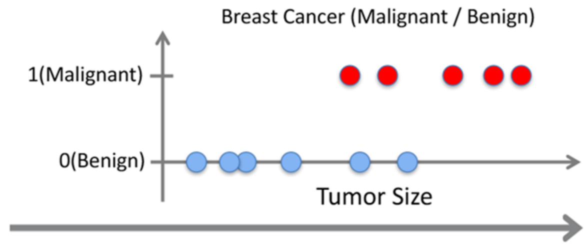
SCIENCE ACADEMY

DATA/MSML 603

Machine Learning (3/4)

Supervised vs. Unsupervised

- **Supervised learning:** Given training data + desired output (labels)
- Given (x1, y1), (x2, y2), (x3, y3), …, (xn, yn)
- Learn a function f(x) to predict y given x
- **y is categorical == classification**



Based on materials by Andrew Ng

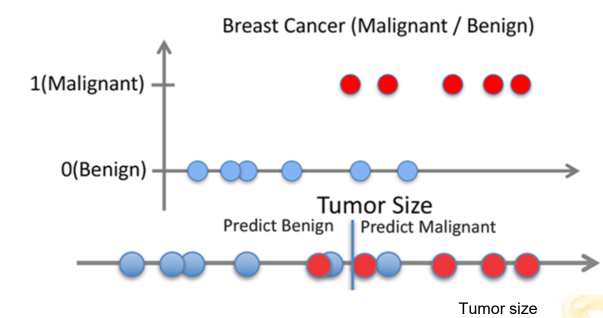
SCIENCE ACADEMY

DATA/MSML 603

Machine Learning (4/4)

Supervised vs. Unsupervised

- **Supervised learning:** Given training data + desired output (labels)
- Given (x1, y1), (x2, y2), (x3, y3), …, (xn, yn)
- Learn a function f(x) to predict y given x
- **y is categorical == classification**



Based on materials by Andrew Ng

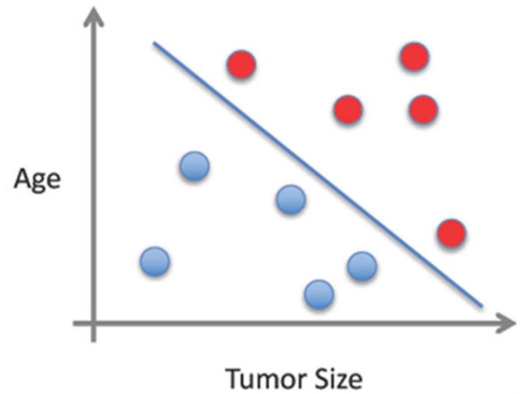
SCIENCE ACADEMY

DATA/MSML 603

Supervised Learning

x Can Be Multi-dimensional

- Each dimension corresponds to an attribute
  - Clump Thickness
  - Uniformity of Cell Size
  - Uniformity of Cell Shape

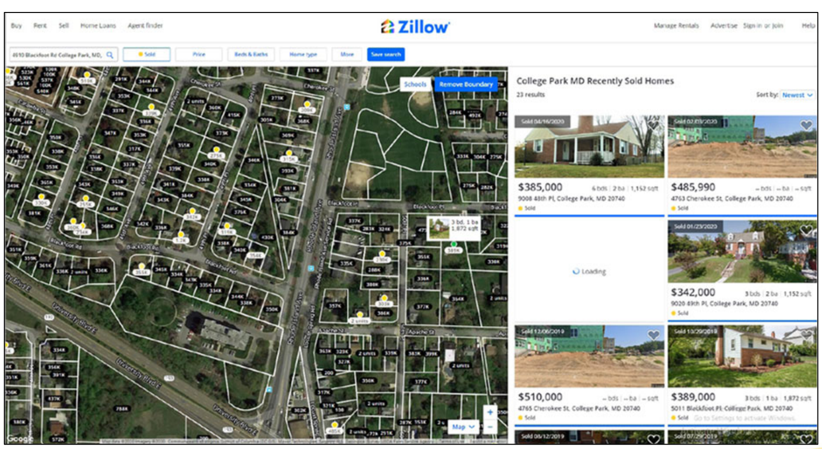


Based on materials by Andrew Ng

SCIENCE ACADEMY

DATA/MSML 603

House Price Prediction



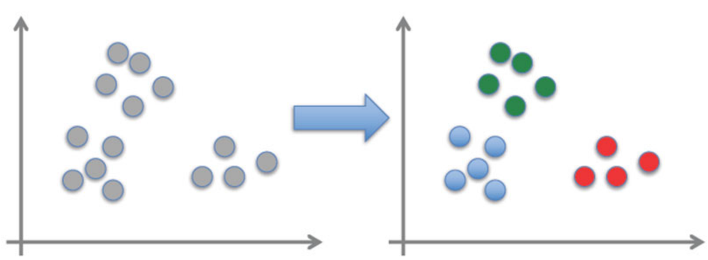
SCIENCE ACADEMY

DATA/MSML 603

Unsupervised Learning

Unsupervised Learning (1/2)

- Given $x_1, x_2, \ldots, x_n$ (without labels)
- Output hidden structure behind the x's
  - E.g., clustering



Unsupervised Learning (2/2)

- Genomics application: group individuals by genetic similarity

Genes

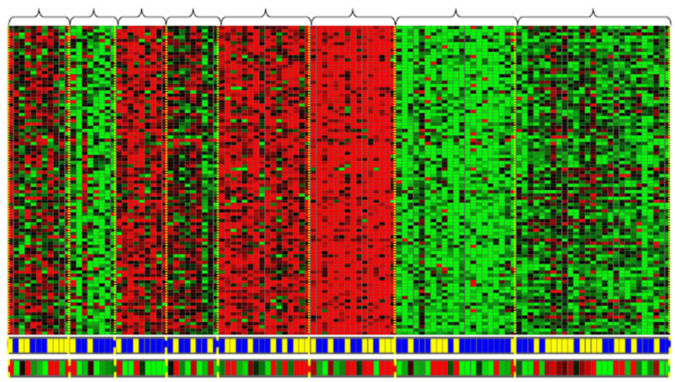


Image source:
Daphne Koller

Individuals

SCIENCE ACADEMY

DATA/MSML 603

MNIST

MNIST: Modified National Institute of Standards and Technology

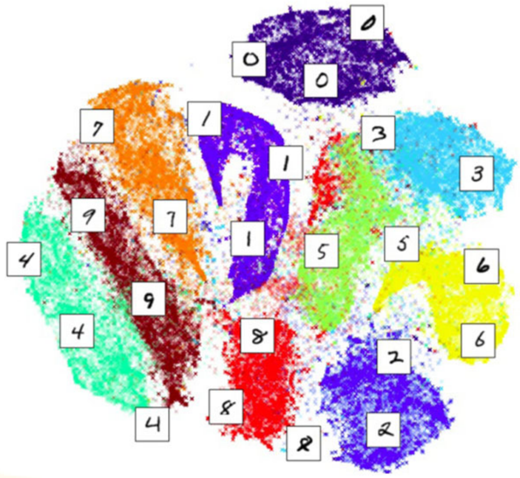
SCIENCE ACADEMY

DATA/MSML 603

Visualization of MNIST (1/2)

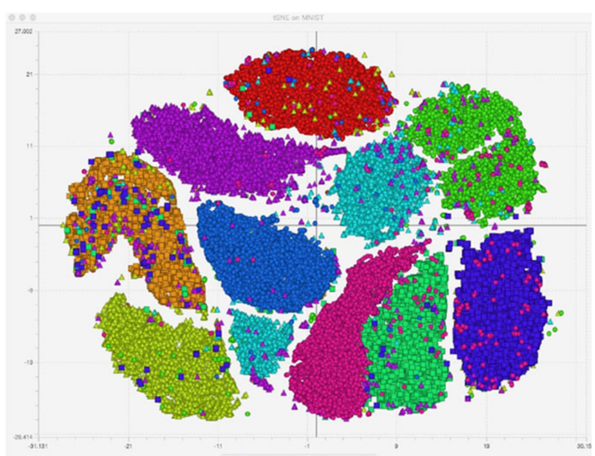
- t-SNE

(t-distributed Stochastic Neighbor Embedding)

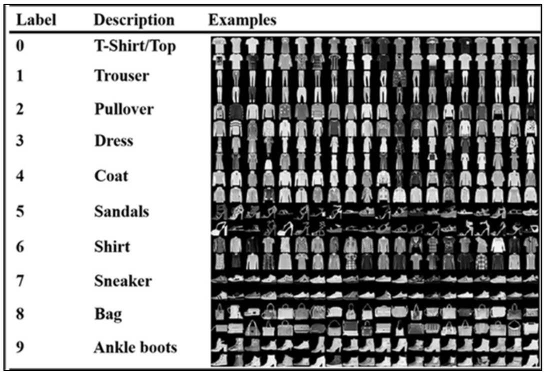


Visualization of MNIST (2/3)

- t-SNE



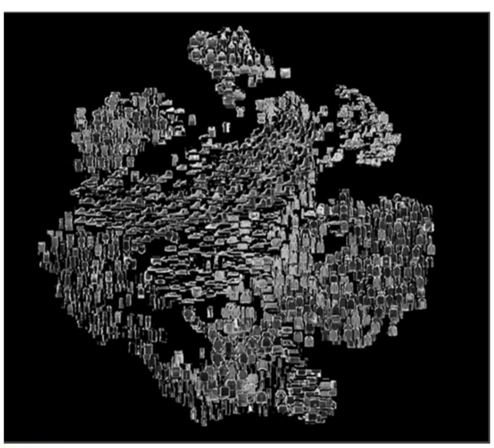
Data (1/2)



Label	Description	Examples
0	T-Shirt/Top	
1	Trouser	
2	Pullover	
3	Dress	
4	Coat	
5	Sandals	
6	Shirt	
7	Sneaker	
8	Bag	
9	Ankle boots	

- Widely used dataset
  - Fashion MNIST
  - Training: 60,000 examples, testing: 10,000 examples (28x28 grayscale image).

Data (2/2)



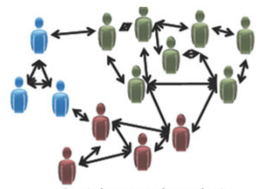
T-SNE visualization

- Widely used dataset
  - Fashion MNIST

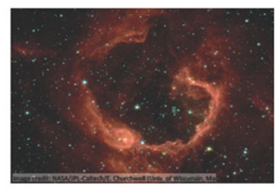
SCIENCE ACADEMY
© 2020, UNIVERSITY OF MARYLAND
DATA/MSML 603

Unsupervised Learning (1/3)


Organize computing clusters


Social network analysis


Market segmentation


Astronomical data analysis

Based on materials by Andrew Ng

Unsupervised Learning (2/3)

- Independent component analysis (ICA) – separate a combined signal into its original sources

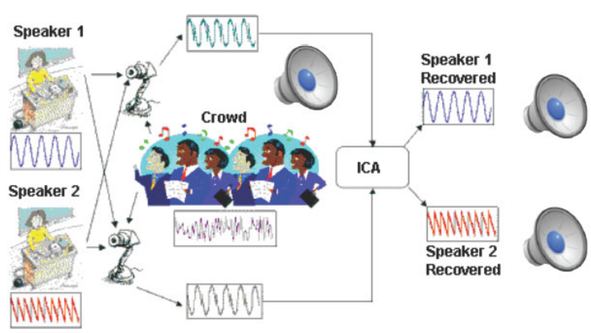
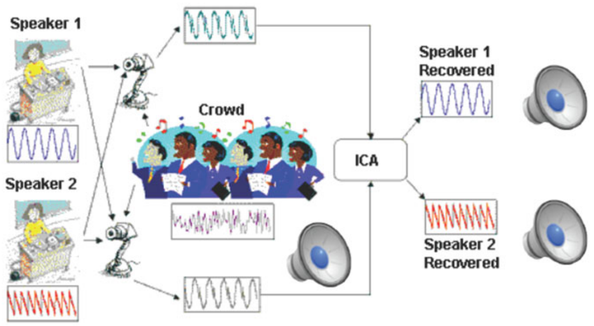


Image source: statsoft.com.

Unsupervised Learning (3/3)

- Independent component analysis – separate a combined signal into its original sources



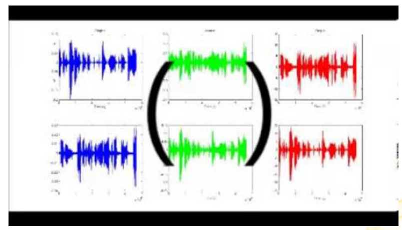


Image source: statsoft.com.

SCIENCE ACADEMY

Reinforcement Learning

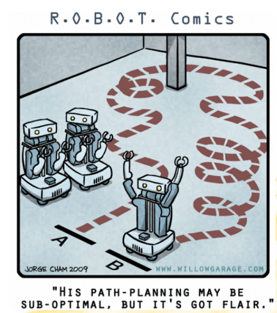
SCIENCE ACADEMY

© 2020, UNIVERSITY OF MARYLAND

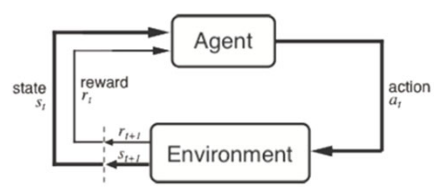
DATA/MSML 603

Reinforcement Learning

- Given a sequence of states and actions with (delayed) rewards, output a policy
  - Policy is a mapping from states' actions that tells you what to do in a given state
- Examples:
  - Credit assignment problem
  - Game playing (chess, Go, etc.)
  - Robot in a maze or path planning with obstacles
  - Balance a pole on your hand



The Agent-Environment Interface



Agent and environment interact at discrete time steps : $t = 0, 1, 2, K$

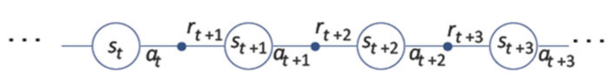
Agent observes state at step t : $s_t \in S$

produces action at step t : $a_t \in A(s_t)$

gets resulting reward : $r_{t+1} \in \Re$

and resulting next state : s_{t+1}

Slide credit:
Sutton & Barto



Reinforcement Learning

SCIENCE ACADEMY

DATA/MSML 603

Reinforcement Learning

- Google DeepMind

SCIENCE ACADEMY

DATA/MSML 603

Inverse Reinforcement Learning

- Learn policy from user demonstrations
- Stanford Autonomous Helicopter

SCIENCE ACADEMY

DATA/MSML 603

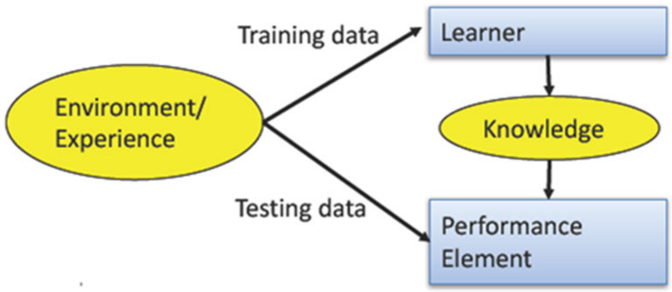
Framing a Learning Problem

SCIENCE ACADEMY

DATA/MSML 603

Designing a Learning System

- Learn policy from user demonstrations
- Choose exactly what is to be learned
  - i.e. the ***target function***
- Choose how to represent the target function
- Choose a learning algorithm to infer the target function from the experience



Based on slide by Ray Mooney

Training vs. Test Distribution

- We generally assume that the training and test examples are independently drawn from the same overall distribution of data
  - We call this "**i.i.d**" which stands for "**independent and identically distributed**"
- If examples are not independent, requires **collective classification**
- If test distribution is different, requires **transfer learning**
  - Transfer learning is a machine learning technique where a model trained on one task is re-purposed for a second related task

ML In a Nutshell

- Tens of thousands of machine learning algorithms
  - Hundreds new every year
- Every ML algorithm has three components:
  - **Representation**
  - **Optimization** (paid advertisement: focus of MSML 604)
  - **Evaluation**

Slide credit: Pedro Domingos

Various Function Representations

- **Numerical functions**
  - Linear regression
  - Neural networks
  - Support vector machines
- **Symbolic functions**
  - Decision trees
  - Rules in propositional logic
  - Rules in first-order predicate logic
- **Instance-based functions**
  - Nearest-neighbor
  - Case-based
- **Probabilistic Graphical Models**
  - Naive Bayes
  - Bayesian networks
  - Hidden-Markov Models (HMMs)
  - Probabilistic Context Free Grammars (PCFGs)
  - Markov networks

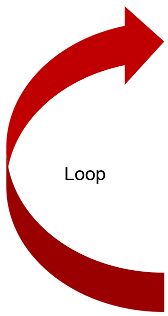
Various Search/Optimization Algorithms

- **Gradient Descent**
  - Perceptron
  - Backpropagation
- **Dynamic Programming**
  - HMM (hidden Markov model) Learning
  - PCFG (probabilistic context-free grammars) Learning
- **Divide and Conquer**
  - Decision tree induction
  - Rule learning
- **Evolutionary Computation**
  - Genetic Algorithms (GAs)
  - Genetic Programming (GP)
  - Neuro-evolution

Evaluation

- Accuracy
- Precision and recall
- Squared error
- Likelihood
- Posterior probability
- Cost / Utility
- Margin
- Entropy
- K-L divergence…

ML in Practice



- Understand domain, prior knowledge, and goals
- Data integration, selection, cleaning, pre-processing, etc.
- Learn models
- Interpret results
- Consolidate and deploy discovered knowledge

Slide credit: Pedro Domingos

Lessons Learned About Learning

- Learning can be viewed as using direct or indirect experience to approximate a chosen target function
- Function approximation can be viewed as a search through a space of hypotheses (representations of functions) for one that best fits a set of training data
- Different learning methods assume different hypothesis spaces (representation languages) and/or employ different search techniques

What We'll Cover in This Course

Supervised learning

- Decision tree induction
- Linear regression
- Logistic regression
- Support vector machines & kernel methods
- Model ensembles
- Bayesian learning
- Neural networks & deep learning
- Learning theory

Unsupervised learning

- Clustering
- Dimensionality reduction

Evaluation

- Applications

Our focus will be on applying machine learning to real applications

SCIENCE ACADEMY DATA/MSML 603

Resources

- Google
- Github
- Optional textbook
  - Pattern Recognition and Machine Learning by Christopher M. Bishop

SCIENCE ACADEMY DATA/MSML 603