Introduction to Machine Learning in Geosciences
Introduction II

GEO371T/398D.1

Mrinal K. Sen

August 29, 2023

Content

- Motivation
- Human Learning/How do we learn to learn?
- Artificial intelligence
- Data Analytics
- What is Machine Learning?
- History of Machine Learning
- Types of ML: Supervised, Unsupervised, Semi-supervised and Reinforced Learning
- Classification and Regression
- ML background linear algebra, statistics, computing
- Learning under uncertainty!
- shallow and deep learning

Human Learning

"The truth is that a human is just a brief algorithm — 10,247 lines. They are deceptively simple. Once you know them, their behavior is quite predictable."
— Westworld

What is intelligence?

- Intelligence is an umbrella term used to describe a property of the mind that encompasses many related abilities, such as the capacities
  - to reason,
  - to plan,
  - to solve problems,
  - to think abstractly,
  - to comprehend ideas,
  - to use language, and
  - **to learn and to solve problems**

[from Wikipedia]

Human Learning

- We learn from the things that happen to us - our experiences. For example, we learned that lightning is followed by thunder, we learned not to tell lies because it can cause us to lose our credibility and to lose our friends, or that we learned how to dance by watching others demonstrate dance steps to us.
- We can say that we have learned these things because we have acquired appropriate responses for them - we cover our ears when lightning strikes, we try to avoid telling lies, and we dance.
- Learning is acquiring relatively permanent change in behavior through experience. We experience things and learn to modify our behaviors based on what we know.
- Learning applies not just to humans, but also to animals.

https://general-psychology.weebly.com/how-do-we-learn.html

Human Learning

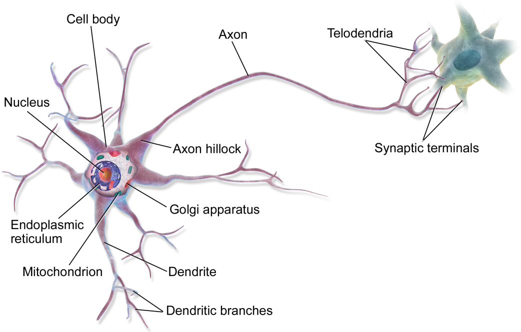
- **Observational Learning** is learning by watching others engage in different behaviors. You probably have learned to dance by watching your teacher demonstrate some dance steps to you.
- **Associative Learning** is learning by establishing connections between events. Conditioning is the method for teaching associations: classical and operant conditioning.
- **Classical conditioning** is the method of teaching associations between two different stimuli. We learn the connection between lightning and thunder because they almost always occur together. Because of this, whenever we see lightning, we cover our ears in anticipation of thunder.
- **Operant conditioning** is the method of teaching associations between behaviors and consequences. Operant conditioning uses rewards and punishments to strengthen or weaken behaviors. For example - the connection between telling lies and losing credibility and friends.

https://general-psychology.weebly.com/how-do-we-learn.html

Human Learning

Human Learning: Physiology

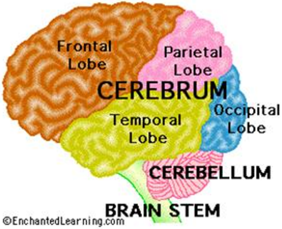
- Neuron, neurone (old British spelling) or nerve cell, is an electrically excitable cell[1] that communicates with other cells via specialized connections called synapses. It is the main component of nervous tissue in all animals except sponges and placozoa. Plants and fungi do not have nerve cells.
- A group of connected neurons is called a neural circuit.



https://trainingindustry.com/articles/content-development/how-the-brain-learns/

Human Learning: Physiology

- The brain acts as a dense network of fiber pathways consisting of approximately 100 billion 10^{10} **neurons**.
- Three principal parts – **stem**, **cerebellum** and **cerebrum**
- the cerebrum is most important in learning, since this is where higher-ordered functions like memory and reasoning occur. Each area of the cerebrum specializes in a function – sight, hearing, speech, touch, short-term memory, long-term memory, language and reasoning abilities are the most important for learning.



https://trainingindustry.com/articles/content-development/how-the-brain-learns/

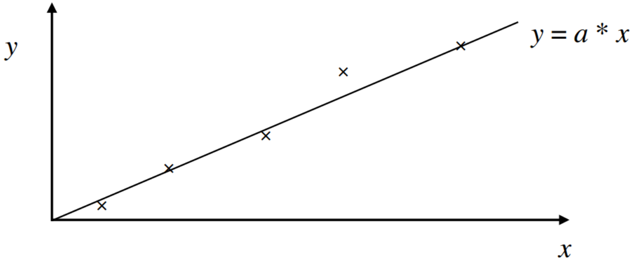
Artificial Intelligence (AI)

- Artificial intelligence (AI) is the simulation of human intelligence processes by machines, especially computer systems. Specific applications of AI include expert systems, natural language processing (NLP), speech recognition and machine vision.
- Artificial Intelligence (AI) is the branch of computer sciences that emphasizes the development of intelligence machines, thinking and working like humans.
- Artificial Intelligence (AI) involves using computers to do things that traditionally require human intelligence. This means creating algorithms to classify, analyze, and draw predictions from data. It also involves acting on data, learning from new data, and improving over time.

https://medium.com/mytake/artificial-intelligence-explained-in-simple-english-part-1-2-1b28c1f762cf

Empirical Inference

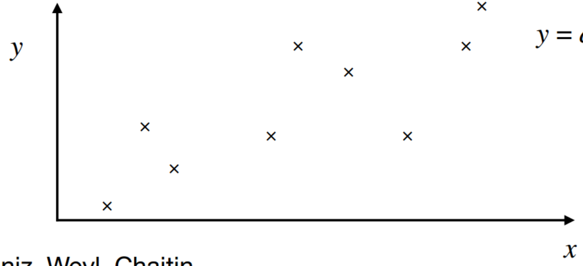
- Drawing conclusions from empirical data (observations, measurements)
- Example 1: scientific inference



slide by Bernhard Schölkopf

Empirical Inference

- Drawing conclusions from empirical data (observations, measurements)
- Example 1: scientific inference

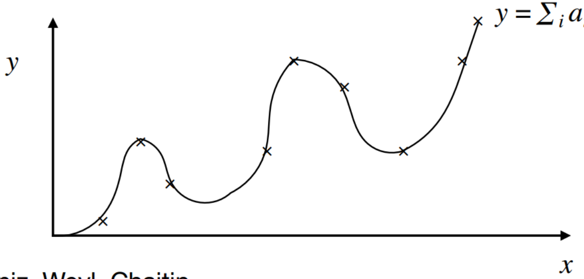


Leibniz, Weyl, Chaitin

slide by Bernhard Schölkopf

Empirical Inference

- Drawing conclusions from empirical data (observations, measurements)
- Example 1: scientific inference



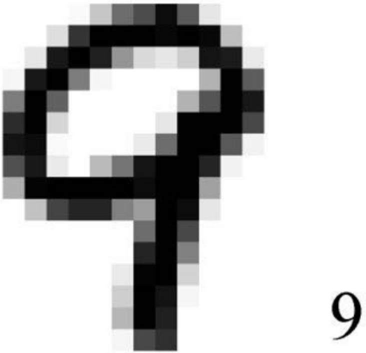
Leibniz, Weyl, Chaitin

slide by Bernhard Schölkopf

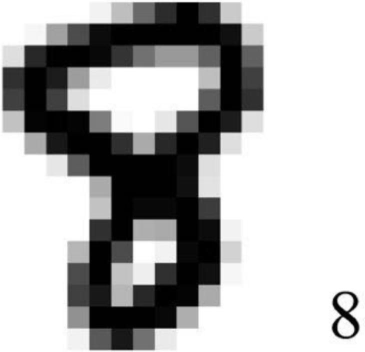
Empirical Inference

- Example 2: perception

slide by Bernhard Schölkopf



slide by Bernhard Schölkopf



slide by Bernhard Schölkopf

8

slide by Bernhard Schölkopf

9

slide by Bernhard Schölkopf

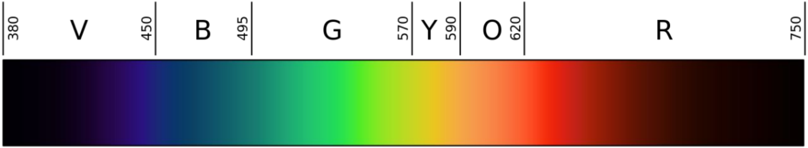
Empirical Inference

- Example2: perception

"The brain is nothing but a statistical decision organ"
H. Barlow

slide by Bernhard Schölkopf

Color Perception



Hard Inference Problems

- High dimensionality – consider many factors simultaneously to find regularity
- Complex regularities – nonlinear; nonstationary, etc.
- Little prior knowledge – e.g. no mechanistic models for the data
- Need large data sets – processing requires computers and automatic inference methods

slide by Bernhard Schölk

What is machine learning?

Example: Netflix Challenge

- Goal: Predict how a viewer will rate a movie
- 10% improvement = 1 million dollars

slide by Yaser Abu-Mostapha

Example: Netflix Challenge

- Goal: Predict how a viewer will rate a movie
- 10% improvement = 1 million dollars
- Essence of Machine Learning:
  - A pattern exists
  - We cannot pin it down mathematically
  - We have data on it

slide by Yaser Abu-Mostapha

What is Machine Learning?

- [Arthur Samuel, 1959]
  - Field of study that gives computers
  - the ability to learn without being explicitly programmed

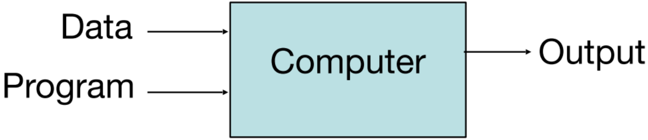
- [Kevin Murphy] algorithms that
  - automatically detect patterns in data
  - use the uncovered patterns to predict future data or other outcomes of interest

- [Tom Mitchell] algorithms that
  - improve their performance (P)
  - at some task (T)
  - with experience (E)

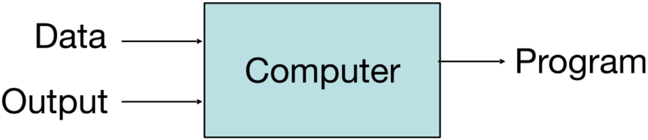
slide by Dhruv Batra

Comparison

- **Traditional Programming**



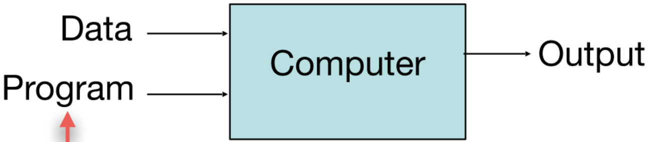
- **Machine Learning**



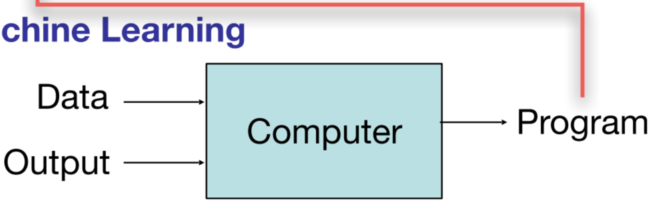
slide by Pedro Domingos, Tom Mitchel, Tom Dietterich

Comparison

- **Traditional Programming**

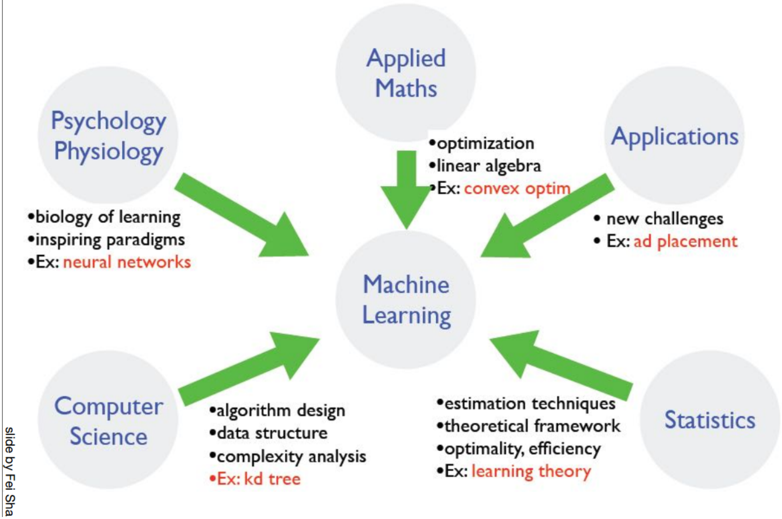


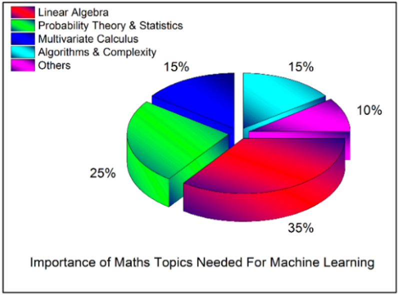
- **Machine Learning**



slide by Pedro Domingos, Tom Mitchel, Tom Dietterich

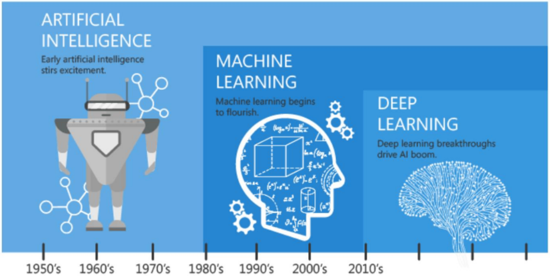
Where does ML fit in?




Linear Algebra
Probability Theory & Statistics
Multivariate Calculus
Algorithms & Complexity
Others
15%
15%
10%
25%
35%
Importance of Maths Topics Needed For Machine Learning

History of Machine Learning

- Neural Networks (1960)
- Multi-layer Perceptions (1985)
- Restricted Boltzman Machines (1986)
- Support Vector Machine (1995)
- Deep Belief Networks – New interest in deep learning (2005) CNN
- Deep Recurrent Neural Network (2009)
- Convolutional DBN (2010)
- Max Pooling CDBN (2011)



Since an early flush of optimism in the 1950's, smaller subsets of artificial intelligence - first machine learning, then deep learning, a subset of machine learning - have created ever larger disruptions.

Summary of differences: AI vs. machine learning

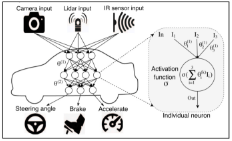
	Artificial Intelligence	**Machine Learning**
What is it?	AI is broad term for machine-based applications that mimic human intelligence. Not all AI solutions are ML.	ML is an artificial intelligence methodology. All ML solutions are AI solutions.
Best suited for	AI is best for completing a complex human task with efficiency.	ML is best for identifying patterns in large sets of data to solve specific problems.
Methods	AI may use a wide range of methods, like rule-based, neural networks, computer vision, and so on.	For ML, people manually select and extract features from raw data and assign weights to train the model.
Implementation	AI implementation depends on the task. AI is often prebuilt and accessed via APIs.	You train new or existing ML models for your specific use case. Prebuilt ML APIs are available.

Today AI is ubiquitous

- Automate routine labor
  - Search

- Understand speech
  - SIRI, Alexa

- Autonomous Vehicles



Why are things working today?

- More compute power
- More data
- Better algorithms/ models

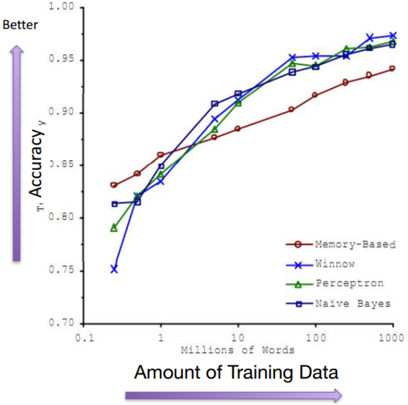


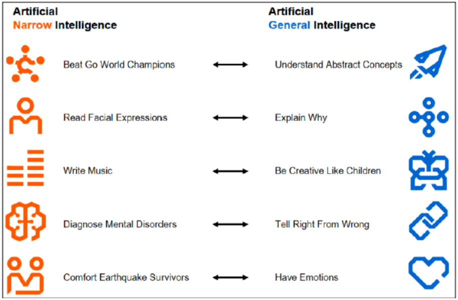
Figure Credit: Banko & Brill, 2011

slide by Dhruv Batra

AI Paradox

- Hard problems for people are easy for AI
- Easy problems are hard for AI
  - Narrow Intelligence General Intelligence

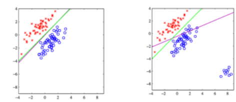
People easy tasks:



The Machine Learning approach

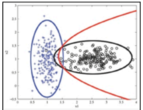
- Difficulties of hard-coded approach suggests:
  – Allow computers to learn from experience
- First determine what features to use
- Learn to map the features to outputs

Linear classifier



Magenta: Least Squares
Green: Logistic Regression (more robust)

Quadratic classifier



2D binary classification with Naive Bayes. A density contour is drawn for the Gaussian model of each class and the decision boundary is shown in red

The ML Approach

Data Collection
- Samples

Model Selection
- Probability distribution to model process

Parameter Estimation
- Values/distributions

Inference
- Find responses to queries

Generalization

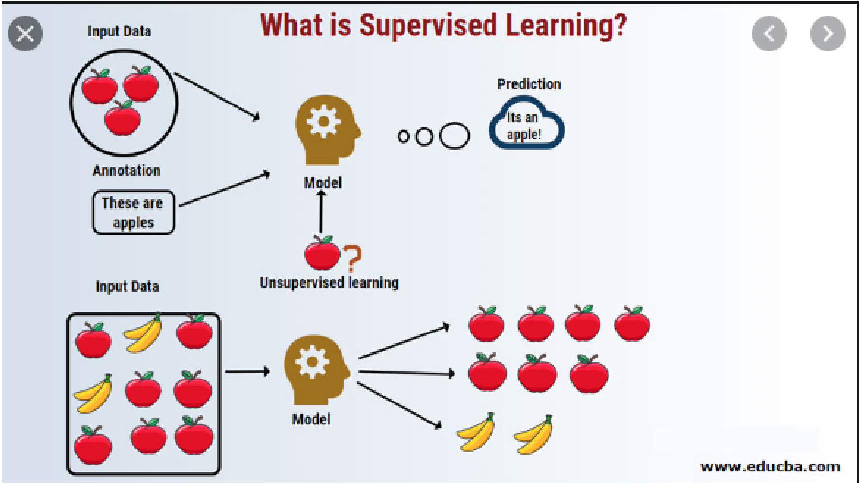
(Training)

Decision
(Inference
OR
Testing)

11

ML Problem Types

1. Based on Type of Data
   1. Supervised, Unsupervised, Semi-supervised
   2. Reinforcement Learning
2. Based on Type of Output
   - Regression, Classification
3. Based on Type of Model
   - Generative, Discriminative

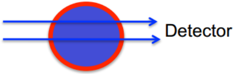

What is Supervised Learning?
Input Data
Annotation
These are apples
Model
Prediction
Its an apple!
Unsupervised learning
Input Data
Model
www.educba.com

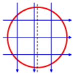
Supervised Learning

- Most widely used methods of ML, e.g.,
  - Spam classification of email
  - Face recognizers over images
  - Medical diagnosis systems
- Inputs $\boldsymbol{x}$ are vectors or more complex objects
  - documents, DNA sequences or graphs
- Outputs are binary, multiclass(K),
  - Multi-label (more than one class), ranking,
  - Structured:
    - $\boldsymbol{y}$ is a graph satisfying constraints, e.g., POS tagging
  - Real-valued or mixture of discrete and real-valued

Supervised Classification Example

- Off-shore oil transfer pipelines
  - Non-invasive measurement of *proportion* of oil,water, gas
    - Called Three-phase Oil/Water/Gas Flow
- Input data: Dual-energy gamma densitometry
  - Beam of gamma rays passed through pipe
  - Attenuation in intensity indicates density of material
  - Single beam insufficient
    - Two degrees of freedom: fraction of oil, fraction of water
    - One beam of Gamma rays of two energies (frequencies)

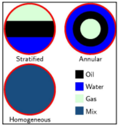




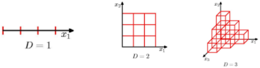
•Six Beams
•12 measurements
•attenuation

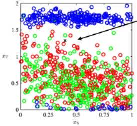
Prediction Problems

1. Predict Volume Fractions of oil/water/gas
2. Predict configuration (one of three)
   - Twelve Features
     - Three classes
     - Two variables, 100 points shown

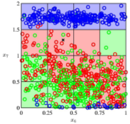


- Naïve cell based voting fails
  - exponential growth of cells with dimensionality
  - 12 dimensions discretized into 6 gives 3 million cells
- Hardly any points in each cell



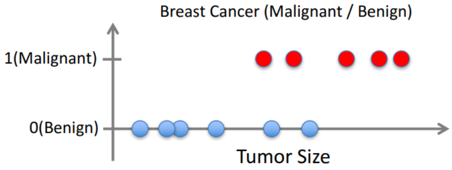


Which class should $\boldsymbol{x}$ belong to?



Supervised Learning: Classification

- Given $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- Learn a function $f(x)$ to predict y given x
  - y is categorical == classification

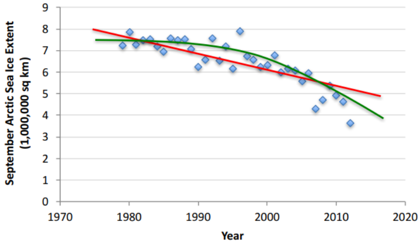


Based on example by Andrew Ng

27

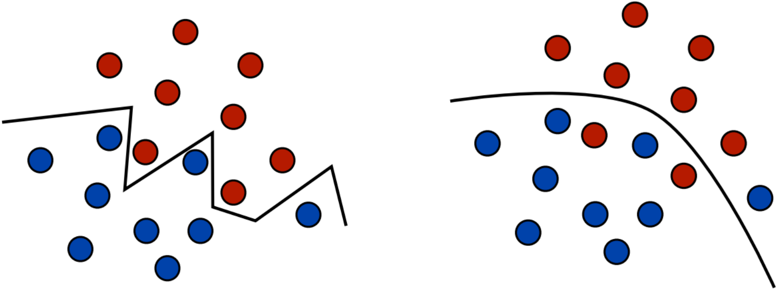
Supervised Learning: Regression

- Given (x_1, y_1), (x_2, y_2), ..., (x_n, y_n)
- Learn a function $f(x)$ to predict y given x
  - y is real-valued == regression



Data from G. Witt. Journal of Statistics Education, Volume 21, Number 1 (2013)

Learning ≠ Fitting



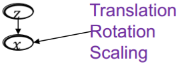
Notion of simplicity/complexity.

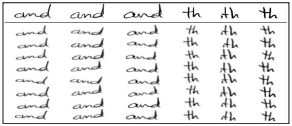
How do we define complexity?

Ability to Generalize

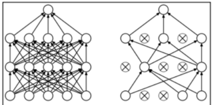
- ML algorithms need to perform well not just on training data but on new inputs as well

1. Parameter Norm Penalties (L^2- and L^1- regularization)
2. Data Set Augmentation



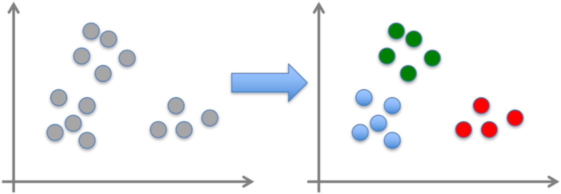


3. Early Stopping
4. Dropout


(a) Standard Neural Net (b) After applying dropout.

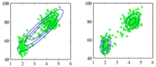
Unsupervised Learning

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

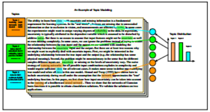


Unsupervised Learning

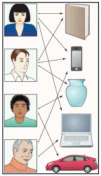
- Unlabeled data assuming underlying structure
  1. Clustering to find partition of data



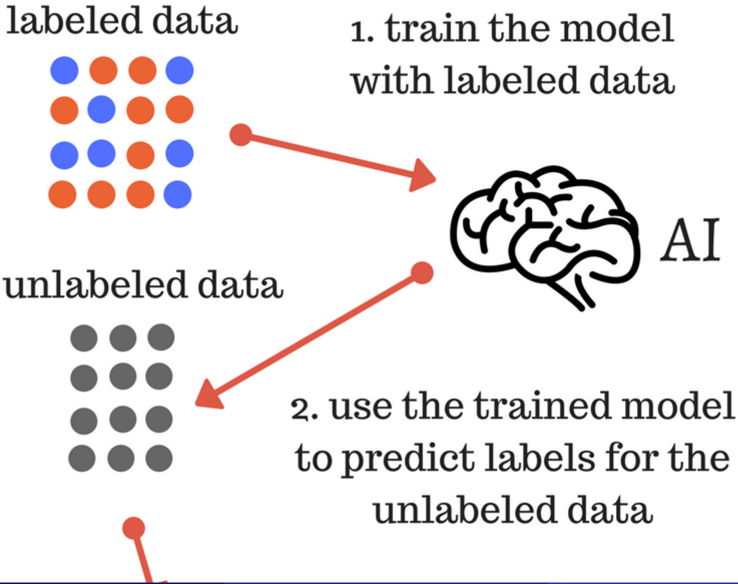
  2. Identify a low-dimensional manifold
     - PCA, Autoencoder
  3. Topic modeling
     - Topics are distributions over words
     - Document: a distribution across topics
       - Methods: SVD, Collaborative Filtering



  4. Recommendation Systems
  - Data links between users and items
  - Suggest other items to user
  - Solution: SVD, Collaborative Filtering

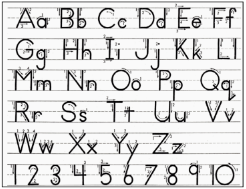


semi-supervised

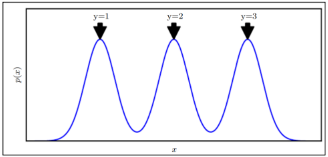


How semi-supervised can succeed

- Ex: density over $\boldsymbol{x}$ is a mixture over three components, one per value of $\mathbf{y}=$ cap/small/digit
- If components well-separated:
  - modeling $p(\mathbf{x})$ reveals where each component is
    - A single labeled example per class enough to learn $p(\mathbf{y}|\mathbf{x})$



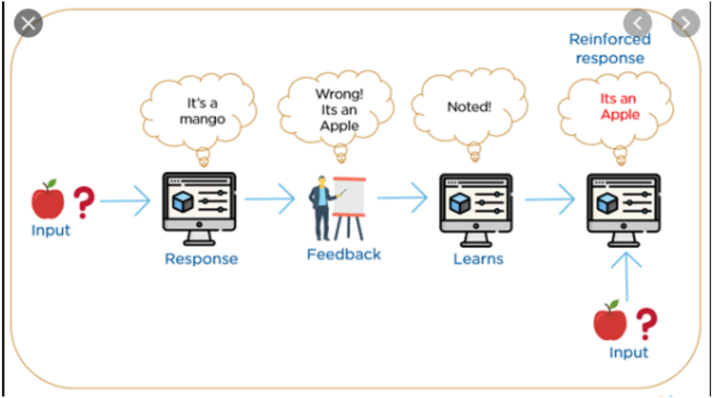
$\boldsymbol{x}$ = no. of black pixels



In this case $p(\mathbf{y}|\mathbf{x})$ is a univariate Gaussian for $\mathbf{y}=1,2,3$

21

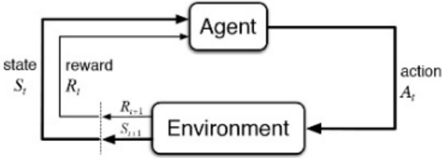
Reinforced Learning



Reinforced Learning

What is reinforcement learning?

- No explicit training data set.
- Nature provides reward for each of the learners actions.
- At each time,
  - Learner has a state and choses an action.
  - Nature responds with new state and a reward.
  - Learner learns from reward and makes better decisions.



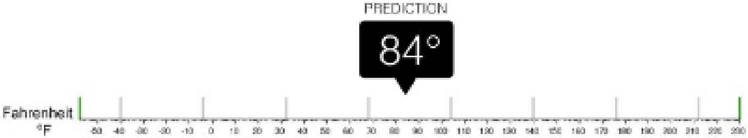
6

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
  – Robot in a maze
  – Balance a pole on your hand

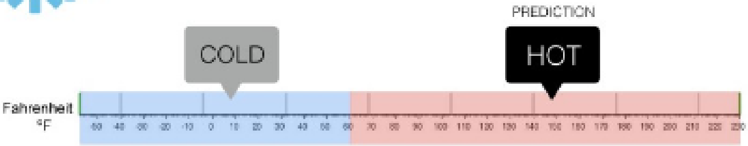
Regression

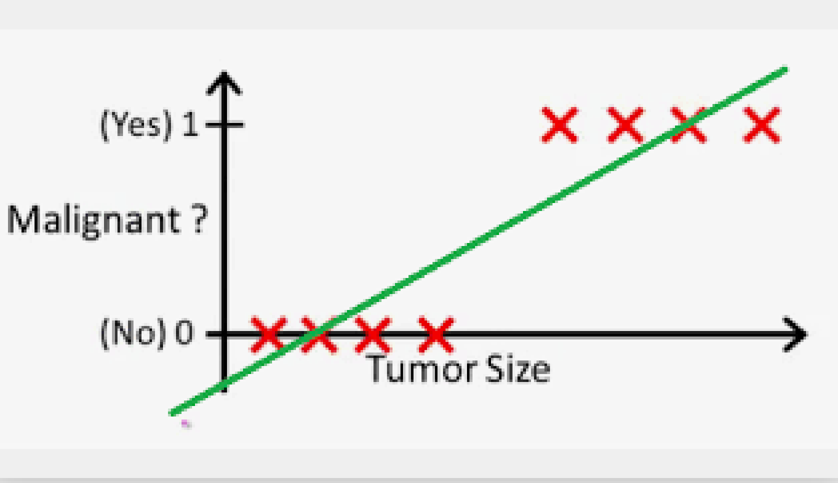
What is the temperature going to be tomorrow?



Classification

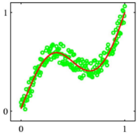
Will it be Cold or Hot tomorrow?




(Yes) 1
Malignant ?
(No) 0
Tumor Size

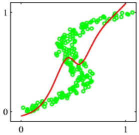
Regression

Problem data set



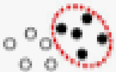
Red curve is result of fitting a two-layer neural network by minimizing squared error

Corresponding inverse problem by reversing x and t

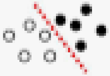


Very poor fit to data: GMMs used here

Generative vs. Discriminative



- Generative:
  - probabilistic "model" of each class
  - decision boundary:
    - where one model becomes more likely
  - natural use of unlabeled data



- Discriminative:
  - focus on the decision boundary
  - more powerful with lots of examples
  - not designed to use unlabeled data
  - only supervised tasks

Generative classifiers

- Assume some functional form for **P(Y), P(X|Y)**
- Estimate parameters of **P(X|Y), P(Y)** directly from training data
- Use Bayes rule to calculate **P(Y |X)**

Discriminative Classifiers

- Assume some functional form for **P(Y|X)**
- Estimate parameters of **P(Y|X)** directly from training data

Examples:

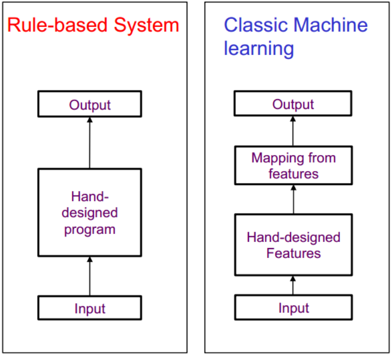
Generative classifiers

- Naïve Bayes
- Bayesian networks
- Markov random fields
- Hidden Markov Models (HMM)

Discriminative Classifiers

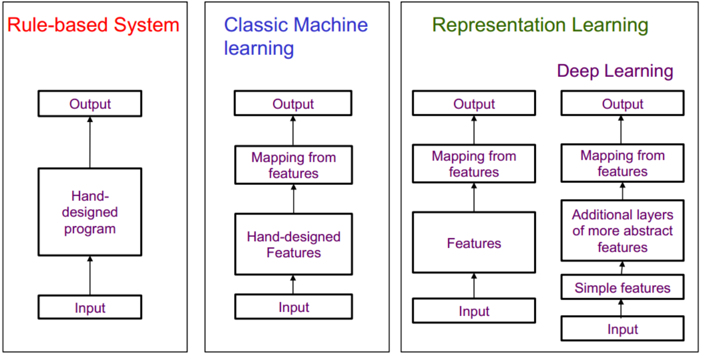
- Logistic regression
- Scalar Vector Machine
- Traditional neural networks
- Nearest neighbour
- Conditional Random Fields (CRF)s

Two paradigms in AI



Shaded boxes indicate components that can learn from data

Summary of AI Models



Shaded boxes indicate components that can learn from data

ML Steps

- Spaces: input space X, output space Y,
- Loss Function: $L : YxY \to \mathbb{R}$.
  - $L(y, \hat{y})$:
  - binary classification: 0-1 loss, $L(y, \hat{y}) = 1_{y \neq \hat{y}}$.
  - regression: $l(y, \hat{y}) = (y - \hat{y})^2$.
- Hypothesis Set: subset of functions out of which the learner selects his hypothesis.
  - depends on features.
  - represents prior knowledge about task.

Supervised Learning Set-Up

- Training Data: $S = ((x_1, y_1), (x_2, y_2), ..., (x_m, y_m)$; x_i: data; y_i: Label data
- Problem: Find hypothesis h with small generalization error (OPTIMIZATION),
  - Deterministic case: output label deterministic function of input, $y = f(x)$.
  - Stochastic case: output probabilistic function of input.

Notice any similarities with the way we solve our inverse problems?

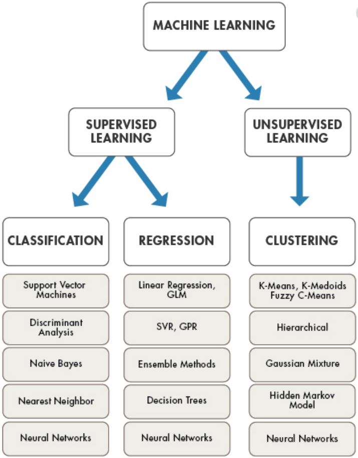
Machine Learning

- Forward Modeling operator is unknown!
- Goal: Find an operator f that can be applied to the data to estimate models.
- Find the operator by systematic examination of a series of observed data and their known answers. Learn from experience! TRAINING

$$\mathbf{m}_{est} = f\mathbf{d}$$

.

Summary


MACHINE LEARNING
SUPERVISED LEARNING
UNSUPERVISED LEARNING
CLASSIFICATION
REGRESSION
CLUSTERING
Support Vector Machines
Discriminant Analysis
Naive Bayes
Nearest Neighbor
Neural Networks
Linear Regression, GLM
SVR, GPR
Ensemble Methods
Decision Trees
Neural Networks
K-Means, K-Medoids Fuzzy C-Means
Hierarchical
Gaussian Mixture
Hidden Markov Model
Neural Networks