FUDAN FISS SUMMER SCHOOL INTRODUCTION TO AI

Probabilistic Reasoning and Decision Making

Day 1 – Intro, Probability review

Welcome to Introduction To AI

- A theory class focusing on Bayes learning and its variations

Elearning site: Fudan e-learning website

- Paul Cao— UC-San Diego
- Office hours: TBD
- yic242@eng.ucsd.edu
- TAs: Zhuoyao Gu

What is AI?

- Write down the first 5 things that come to mind when you think about Artificial Intelligence

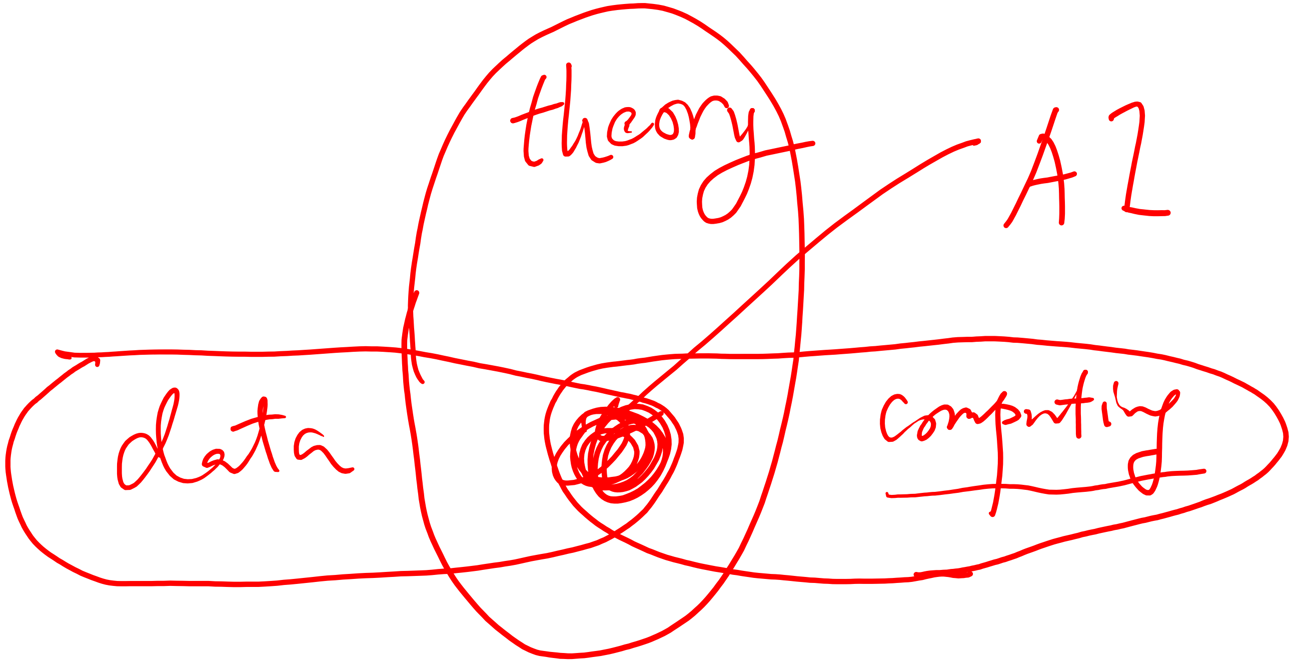
theory
NLP
NN (deep learning)
vision
classification
genetic algorithm / reinforcement learning

autonomous
vehicle
security
application

What is AI (ML) and where does this class fit in?

- First, a brief (video) history of AI…

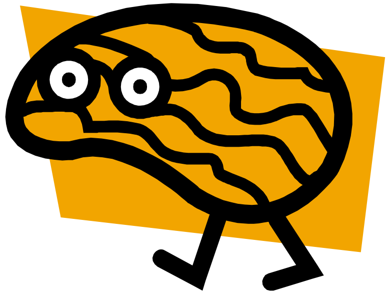
- https://www.youtube.com/watch?v=BFWt5Bxfcjo



When Is a Machine "Intelligent"?

“The Imitation Game”, A. Turing, 1950

When it acts like a human?



When Is a Machine "Intelligent"?

"The Imitation Game", A. Turing, 1950

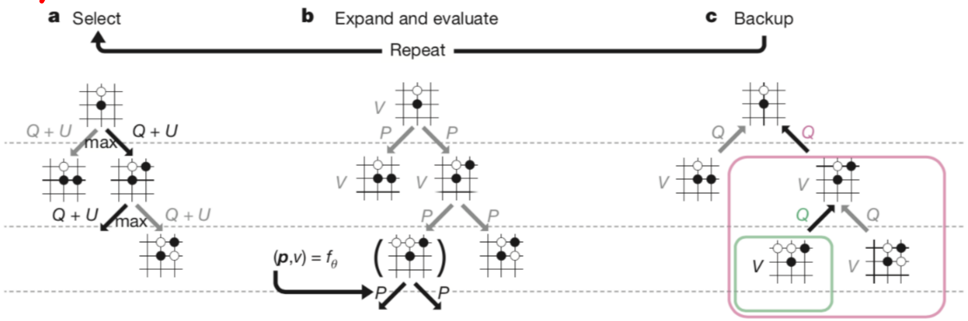
When it acts like a human?

AI fields

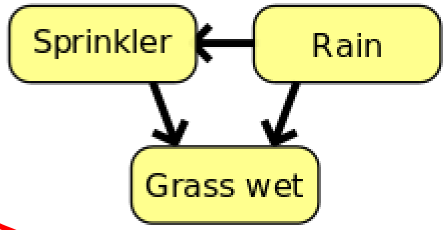
focus for this class

Optimization/Search

operations research

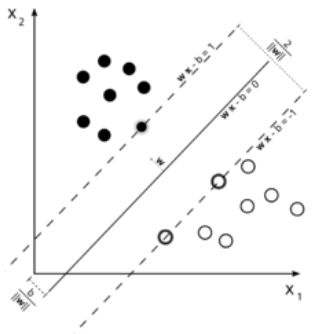


Probabilistic Reasoning

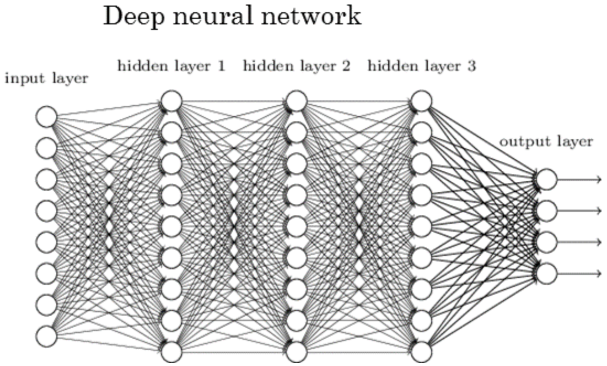


Machine Learning

Neural Networks



Data Mining



+Applications of AI: Databases, Natural Language Processing, Computer Vision, etc

Turing Prize 2011

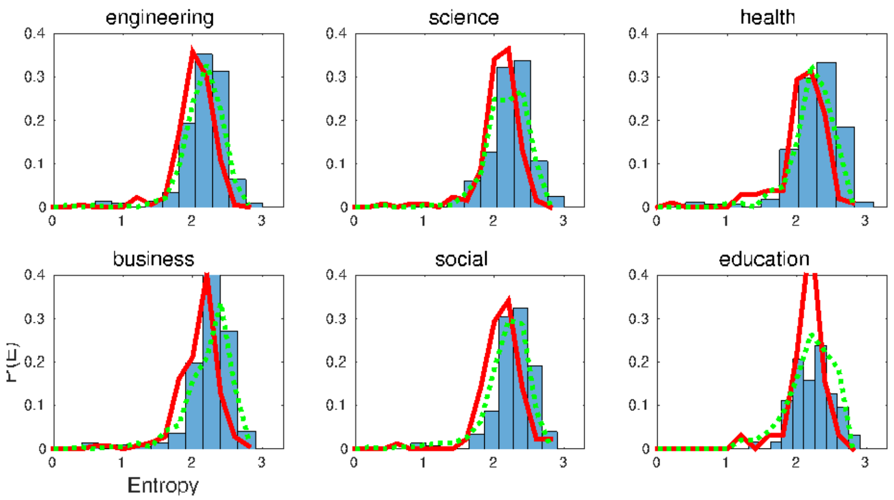
"Judea Pearl is credited with the invention of ***Bayesian networks***, a mathematical formalism for defining complex probability models, as well as the principal algorithms used for inference in these models. *This work not only revolutionized the field of AI but also became an important tool for many other branches of engineering and the natural sciences.*

Who uses probabilistic methods in AI and ML?

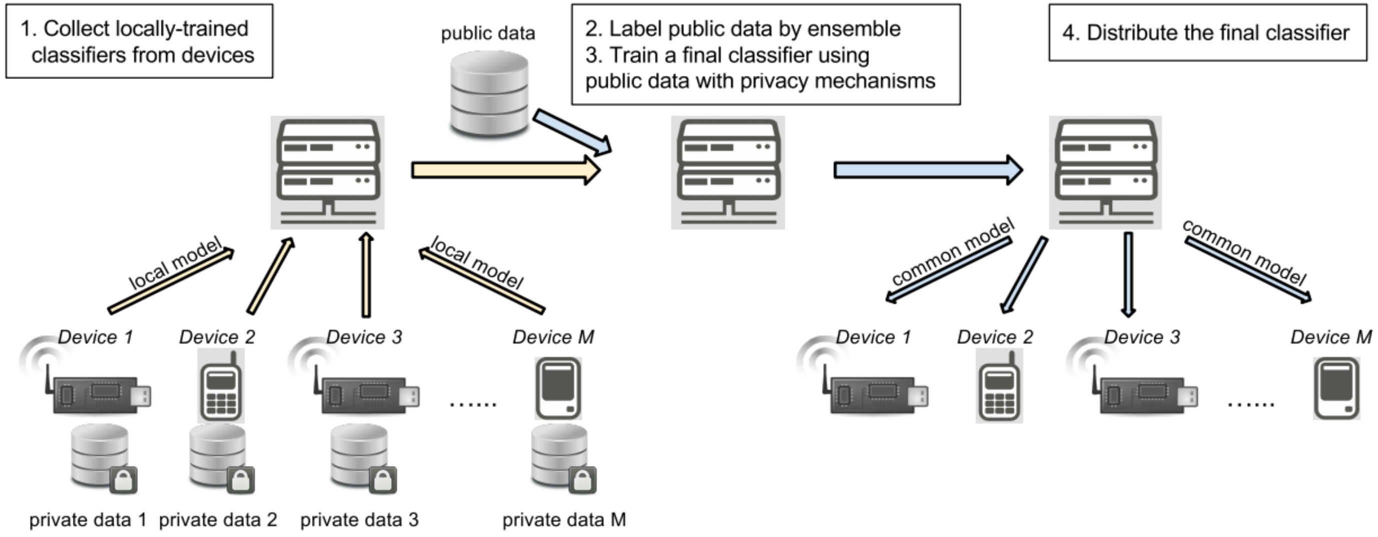
- Search & Ads – Google, Microsoft, Yahoo
- Sales & Recommendations – Amazon
- Social media – Facebook, Twitter, LinkedIn
- Gaming & HCI – XBox, Wii
- Forensics & signal analysis – FBI, NSA
- Data science & analytics

"Every company is a data company."

My Research Focus



- CS Education
- Distributed learning
- Modeling uncertainty



Prerequisites

- Elementary probability
  - *Random variables*
  - *Expected values*
- Linear algebra
  - *Matrix multiplication, inverses*
  - *Solving systems of linear equations*
- Calculus
  - *Computing derivatives*
  - *Computing maxima and minima*

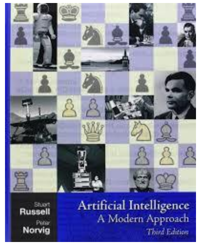
Prerequisites

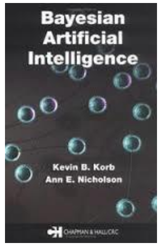
- Programming
  - *We will assign programming assignments*
  - *Also: basic data analysis and visualization.*
  - *Solutions accepted in any language!*
  - *Python, MATLAB, Java, C/C++, Perl, etc.*
  - *No hand-holding with compiling, debugging.*

Non-CSE majors are welcome but you need to be able to write basic programs

Readings vs lectures

- Readings
  - *No truly required texts.*
  - *Some handouts.*





- Lectures
  - *Designed to be self-contained.*
  - *Important for homework assignments.*
  - *Emphasis on mathematical development.*
  - *Will be interactive*
  - *We will also watch some videos together from various AI researchers*

Collaboration Policy

- What is allowed:
  - *You may ask TA or instructor* for help.
- What is not allowed:
  - *Copying from current or former students.*
  - *Uploading current materials to archives.*

Academic dishonesty will result in F for this class

Examples of AI problems that you will be able to solve with what you learn in this course

1. Bayesian Network
   ① inference (net is known)
   ② learn param of net (structure is known, param unknown)

   MLE   EM

2. HMM (sequence of events)
   ① inference (chain is known)
   ?? ② learn param of HMM

Core themes of modern AI

- Probabilistic modeling of uncertainty

Core themes of modern AI

- Probabilistic modeling of uncertainty
- Learning as optimization

Core themes of modern AI

- Probabilistic modeling of uncertainty : Bayes Net (one example)
- Learning as optimization
- Knowledge as prediction

encoded in model

chance involved

Model ($\alpha, \beta, \gamma, \ldots$)

learn param

α, β, γ

learn structure (harder)

$y = ax + b$

param

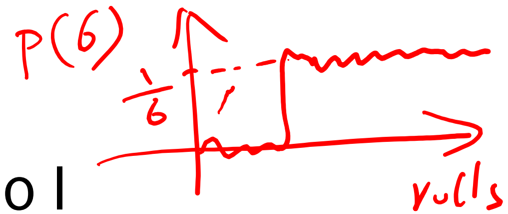
Probability in AI

Frequentist

$P(\text{rain}) = 1\%$ (long term frequency)

$P(\text{roll a die} = 6) = \boxed{\frac{1}{6}}$

Probability Theory == "How knowledge affects belief" (Poole and Mackworth)



What is the probability that it is raining out?

How much do I believe that it is raining out?

affected by data (observations)

Viewing probability as measuring belief (rather than frequency of events) is known as the Bayesian view of probability (as opposed to the frequentist view).

Variables and values

Discrete "random variables", denoted with capital letters: e.g., X

Domain of possible values for a variable, denoted with lowercase letters: e.g. $\{x_1, x_2, x_3, \dots, x_n\}$

Example: Weather W ; $\{w_1 = sunny, w_2 = cloudy\}$

result of a coin toss, T : $\{t_1 = head, t_2 = tail\}$

Unconditional (prior) Probability

X : rolling a die

R.V.

event

$$P(X = x)$$

$P(X=6)$

e.g., What is the probability that the weather is sunny?

$$P(W = w_1) = 95\%$$

$P(w_1) = 95\%$

Axioms of probability

$P(T=1 \text{ or } T=2)$

$= P(T=1) + P(T=2) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6}$

$1 \geq P(X=x) \geq 0$: My belief that X would take a value of x is always between 0 and 1 inclusive.

$\sum_{i=1}^{n} P(X = \underline{x_i}) = 1$: my belief of all possible states sum to 1

a state of x

$\sum_{i=0}^{1} P(W=W_i)$ $P(W = \underbrace{sunny}_{W_0}) + P(W = \underbrace{cloudy}_{W_1}) = 1$

$\sum_{i=1}^{6} P(T=i) = P(T=1) + P(T=2) + \cdots + P(T=6) = 1$

$P(X = x_i \text{ or } X = x_j) = P(X = x_i) + P(X = x_j) \text{ if } x_i \neq x_j$

assume : x_i & x_j are mutually exclusive (x_i & x_j won't happen at the same time)

$P(W = \text{sunny or } W = \text{cloudy}) = P(\text{sunny}) + P(\text{cloudy}) = 1$

Extension of Probability Axioms

We normally care about a group of events

$$P(\varphi) = \sum_{w \in \varphi} P(w)$$

$$P(T \text{ is odd}) = \sum_{x_i \in \{1,3,5\}} P(x_i)$$

T & location are independent

W & math dependant

Conditional Probability

$P(T=6) = \frac{1}{6}$ (unconditional)

$P(T=6 \mid \text{roll in Australia}) = ?$

A: $\frac{1}{6}$ B: $> \frac{1}{6}$ C: $< \frac{1}{6}$

$P(W = \text{rain at SD}) = 7\%$

$P(W = \text{rain} \mid \text{time} = \text{Jan})$

A: 7%   B: >7%   C: <7%

$$P(X = x_i | Y = y_j)$$

"What is my belief that $X = x_i$ if I **already know** $Y = y_j$"

only thing that we know

$P(W = \text{cloudy}) \approx 10\%$

$P(W = \text{cloudy} \mid \text{temp} = 130°F) < 10\%$

- Usually, knowing Y gives you information about X, i.e. changes your belief in X. In this case X and Y are said to be *dependent*.

- When we look at conditional probability, we really say we only already know Y = yj.
  - Cavity example

Independence

$$P(X = x_i | \underline{Y = y_j}) = P(X = x_i)$$ $(\forall x_i, y_j)$

Sometimes knowing Y does not change your belief in X. In this case, X and Y are said to be independent.

$$P(W = w_i | Y = y_j) = P(W = w_i)$$

day of the week & weather independent

For which variable Y is the above statement most likely true?

A. Y = The weather yesterday
B. Y = The day of the week
C. Y = The temperature

More independence

Consider two students Roberto and Sabrina, who both took the same test. Define the following random variables:

R = Roberto aced the test $\begin{cases} 0 : \text{didn't ace (no)} \\ 1 : \text{yes} \end{cases}$

S = Sabrina aced the test

What is the most logical relationship between $P(R = 1)$ and $P(R = 1|S = 1)$?

A. $P(R = 1) = P(R = 1|S = 1)$

B. $P(R = 1) > P(R = 1|S = 1)$

C. $P(R = 1) < P(R = 1|S = 1)$

Conditional independence

$T = \begin{cases} 1 : \text{easy} \\ 0 : \text{isn't easy} \end{cases}$

What if you also know the test was easy (variable T)?

A. $P(R = 1|T = 1) = P(R = 1|T = 1, S = 1)$

B. $P(R = 1|T = 1) > P(R = 1|T = 1, S = 1)$

C. $P(R = 1|T = 1) < P(R = 1|T = 1, S = 1)$

Conditional independence

$P(X = x_i \mid Y = y_j, Z = z_k)$

$= P(X = x_i \mid Z = z_k)$

What if you also know the test was easy (variable T)?

then X & Y are conditionally independent given Z.

A. $P(R = 1|T = 1) = P(R = 1|T = 1, S = 1)$

B. $P(R = 1|T = 1) > P(R = 1|T = 1, S = 1)$

C. $P(R = 1|T = 1) < P(R = 1|T = 1, S = 1)$

R and S are conditionally independent give T. I.e. if you already know T, knowing S does not give you additional information about R.

$P(X = x_i) = P(X = x_i \mid Y = y_j)$

X & Y are (marginally) independent

More independence

Consider two events:

B = A burglar breaks into your apartment

E = An earthquake occurs

Are these events independent or dependent? (i.e. does knowing that one happened change your belief in the other?)

A. They are independent because knowing that one happened does not change your belief that the other happened.
B. They are dependent, because knowing that one happened changes your belief that the other happened.

Conditional dependence

$$P(B = 1) = P(B = 1|E = 1) = P(B = 1|E = 0)$$

Now consider a third event:

A = Your alarm goes off

Which of the following relationships best models beliefs about the world?

A. $P(B = 1|A = 1) = P(B = 1|A = 1, E = 1)$

B. $P(B = 1|A = 1) > P(B = 1|A = 1, E = 1)$

C. $P(B = 1|A = 1) < P(B = 1|A = 1, E = 1)$

Axioms of Conditional probability

- Which of the following axioms hold for conditional probabilities?

A. $P(X = x_i | Y = y_j) \geq 0$

B. $\sum_i P(X = x_i | Y = y_j) = 1$

C. $\sum_j P(X = x_i | Y = y_j) = 1$

D. A and B only

E. A, B and C

AI Ethics

- https://www.youtube.com/watch?v=8g15qeE-LoM