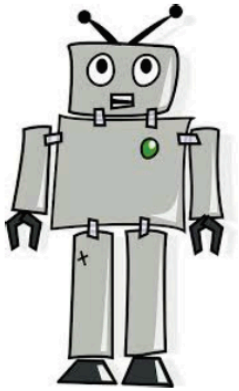
Intelligent Systems

Quan Thanh Tho

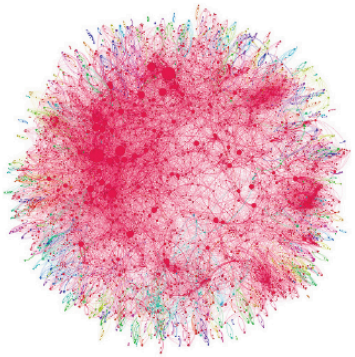
Contents

- From Data Science to Intelligent Systems
- The Course Outline
- Begin with the end in mind – Live Demo
- Exercises

Breakthrough of ICT

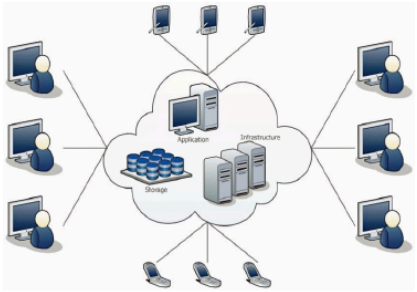


Artificial Intelligence



Big Data

Internet of Things



Cloud Computing

From Data Science to Inteliigent Systems

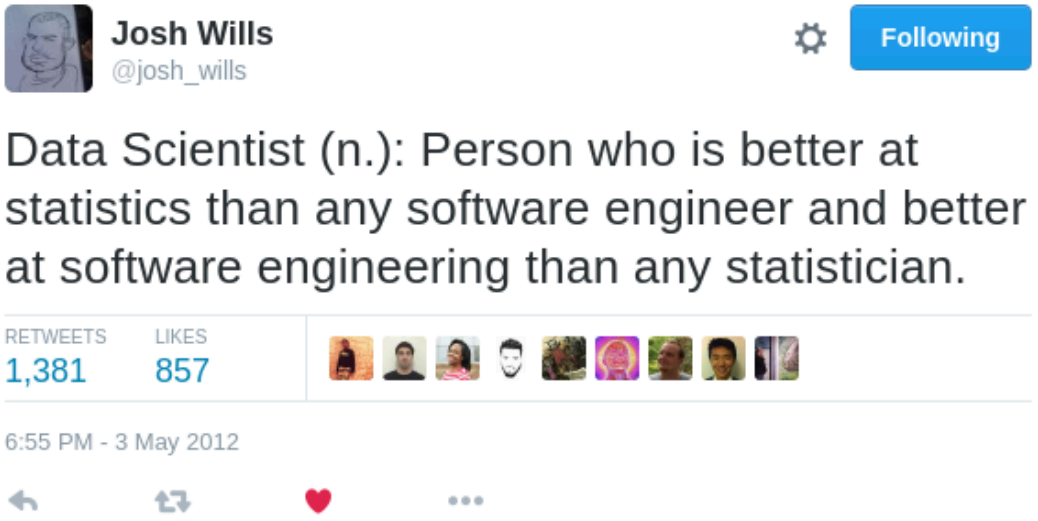
- Data Science and Data Scientists
- Intelligent Systems

Data Science

- everyone talks about it,
- nobody really knows how to do it,
- everyone thinks everyone else is doing it,
- so everyone claims they are doing it...

Data Science

- everyone talks about it,
- nobody really knows how to do it,
- everyone thinks everyone else is doing it,
- so everyone claims they are doing it



Interchangeable Words

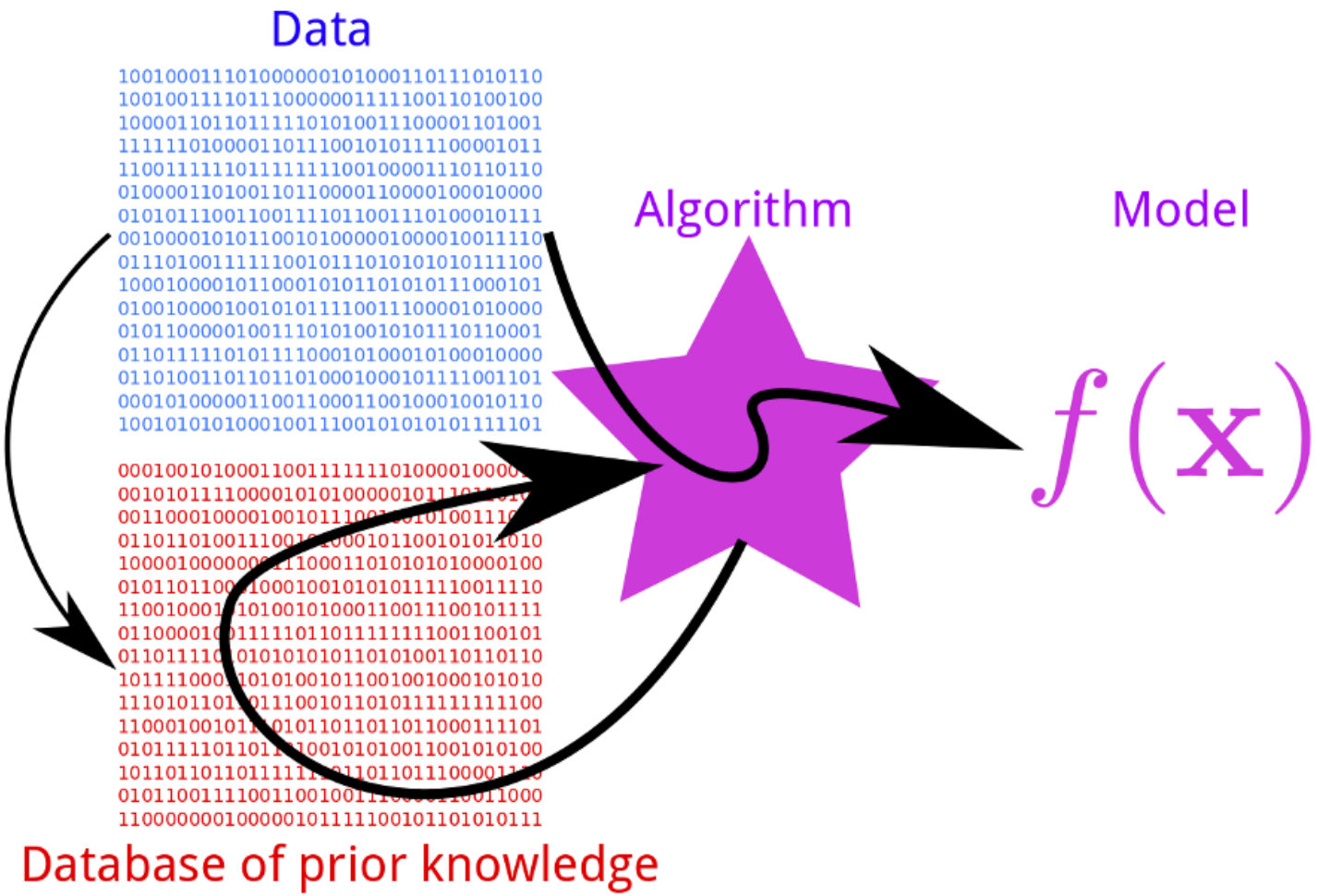
What is Data Science?

Data → Information

Data explodes → too much information

Analyzing & Extracting

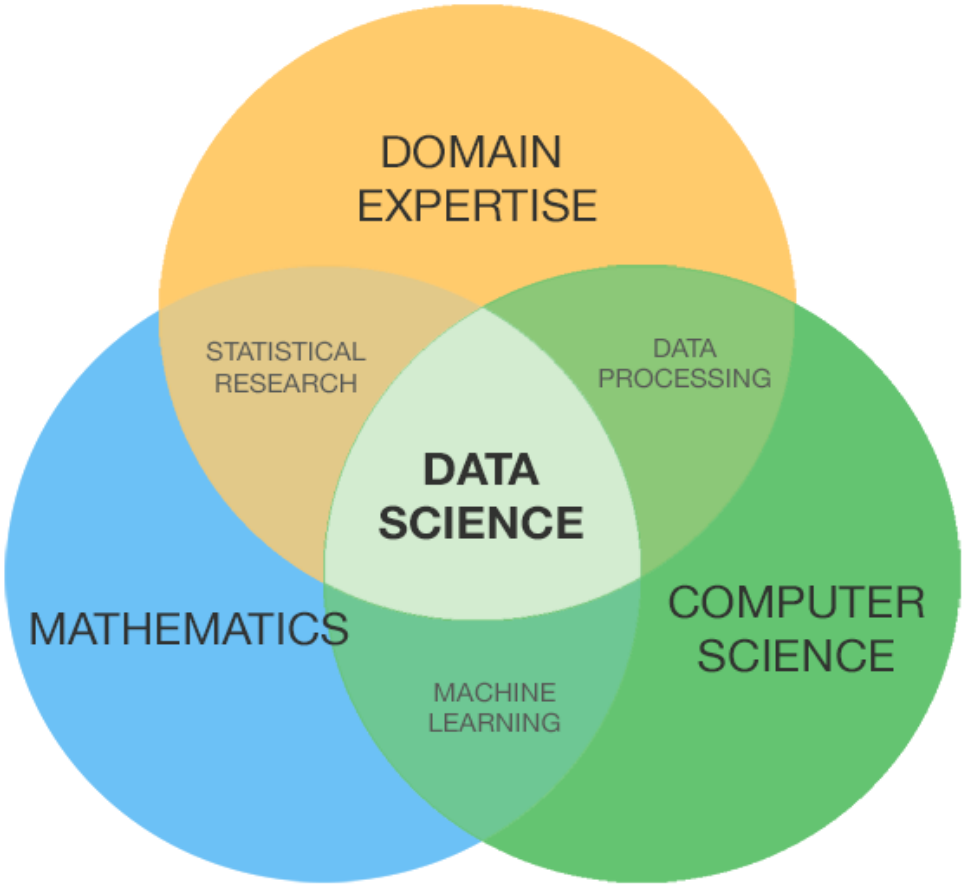
Math/Stats & Learning/Mining Algorithms



Richness of tools & APIs

Copyright (c) 2017, John von Neumann Institute - Viet Nam National University HCMC

Data Science today



Bring Science to Data

https://www.youtube.com/watch?v=vbb-AjiXyh0

Applications

https://www.youtube.com/watch?v=kmXwvyUTwP8

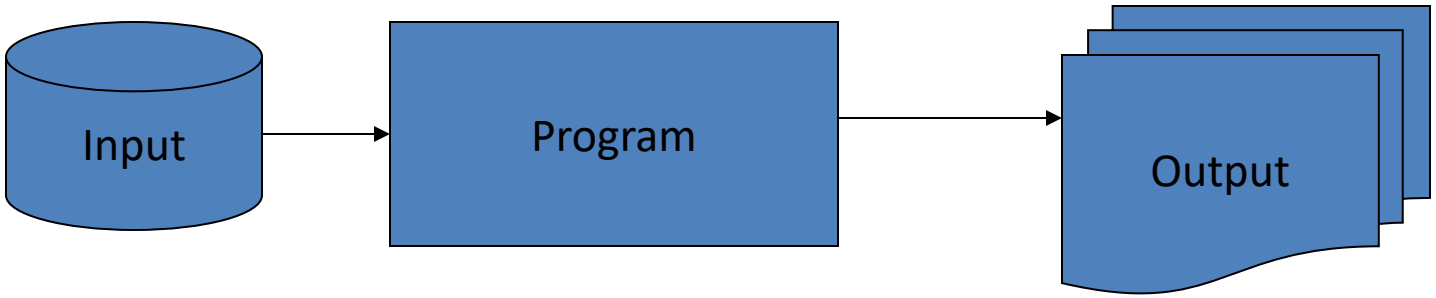
Business Analysis

Financial Services

Analyzing social networks
→Digital marketing/e-commerce

Intelligent Systems

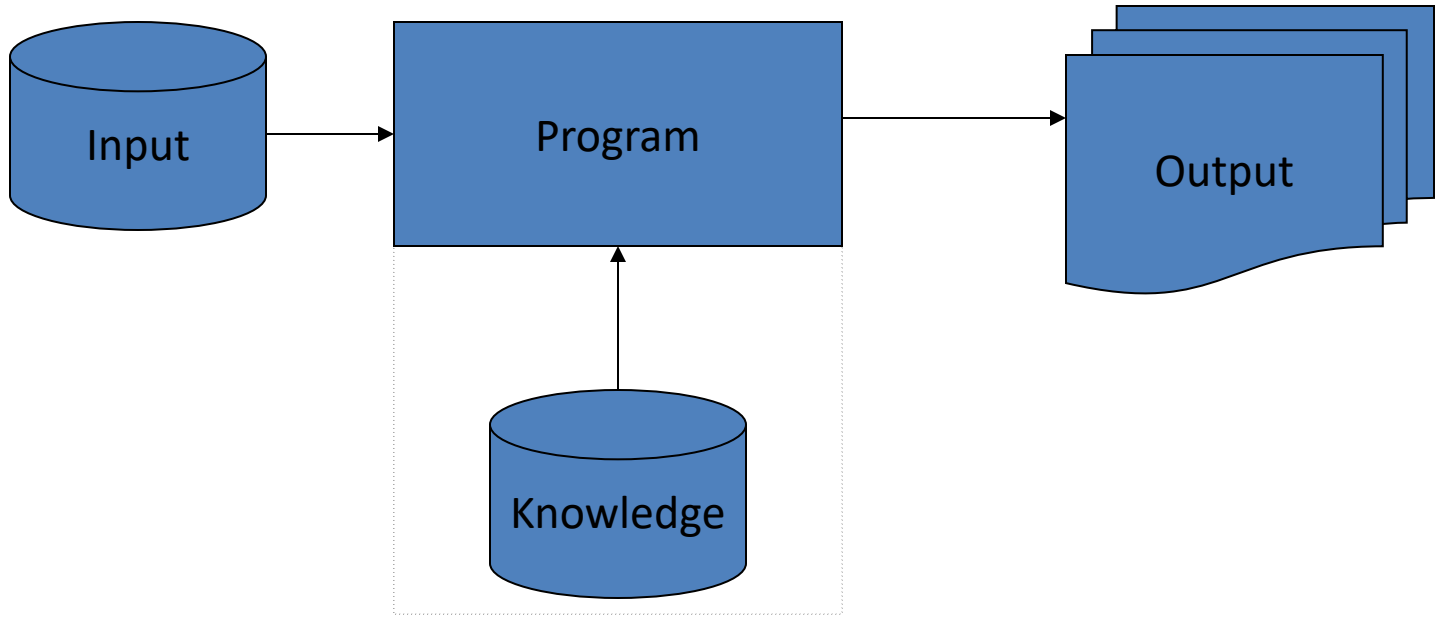
- Information Systems:



  - Same outputs produced for same inputs
  - “How many good students (>8.0) in this class?”

Intelligent Systems

- Information Systems:



Intelligent Systems

- Information Systems:
  - Different outputs produced of the same inputs, depending on the integrated knowledge
  - “How many students will pass the next exam?”

The course outline

- The classical (but practical) machine learning approaches
- The (legendary) deep learning approaches
  - NLP
  - Image Processing
  - Voice Processing

Machine Learning

- The goal of machine learning is to build computer systems that can adapt and learn from their experience (Tom Dietterich, 1999).

- *Given*
  - $\{\boldsymbol{x}_i\}, \boldsymbol{x}_i$ is description of an object in some space, $i = 1, 2, \ldots, n$.
  - y_i is some property of $\boldsymbol{x}_i$ viewed as its label, $y_i \in \{C_1, C_2, \ldots, C_K\}$ or $y_i \in \mathbb{R}$
  - $(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_n, y_n)$

- *Find*

  Function $p(y|\boldsymbol{x})$ for label data and $p(\boldsymbol{x})$ for unlabeled data

(Source: Eric Xing lecture notes)

Five Tribes of Machine Learning

Tribes	Origins	Master Algorithms
Symbolists	Logic, philosophy	Inverse deduction
Evolutionaries	Evolutionary biology	Genetic programming
Connectionists	Neuroscience	Backpropagation
Bayesians	Statistics	Probabilistic inference
Analogizers	Psychology	Kernel machines

Machine Learning Approaches

- Performance Evaluation
- Decision Tree and Random Forest
- k-NN and KD-Tree
- Bayes Classifier
- SVM
- Clustering
- Perceptron in Neural Networks

Begin with the end in mind

- What is the outcome?

Begin with the end in mind

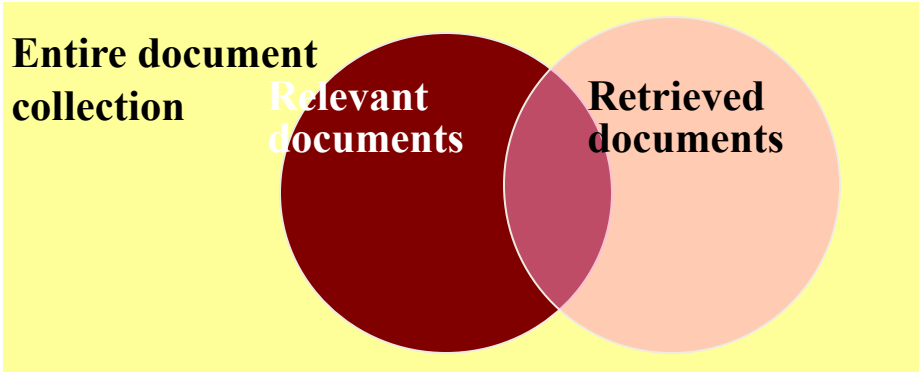
- What is the outcome? >>> How to evaluate the outcome?

Something to demo

Human Labeled Corpora (Gold Standard)

- Start with a corpus of documents.
- Collect a set of queries for this corpus.
- Have one or more human experts exhaustively label the relevant documents for each query.
- Typically assumes binary relevance judgments.
- Requires considerable human effort for large document/query corpora.

Precision and Recall



	retrieved	not retrieved
irrelevant	retrieved & irrelevant	Not retrieved & irrelevant
relevant	retrieved & relevant	not retrieved but relevant

$$recall = \frac{Number\ of\ relevant\ documents\ retrieved}{Total\ number\ of\ relevant\ documents}$$

$$precision = \frac{Number\ of\ relevant\ documents\ retrieved}{Total\ number\ of\ documents\ retrieved}$$

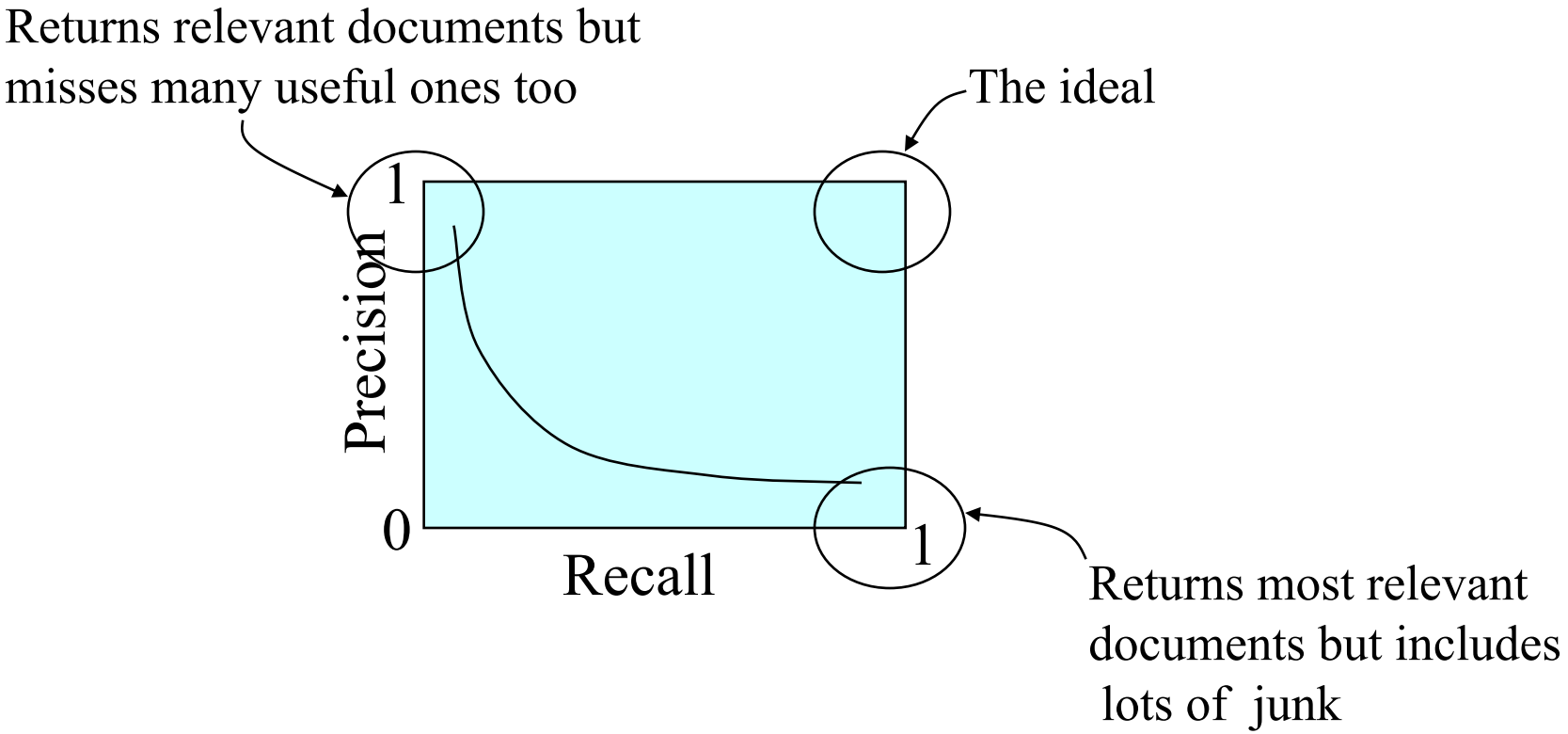
Precision and Recall

- Precision
  - The ability to retrieve top-ranked documents that are mostly relevant.
- Recall
  - The ability of the search to find ***all*** of the relevant items in the corpus.

Determining Recall is Difficult

- Total number of relevant items is sometimes not available:
  - Sample across the database and perform relevance judgment on these items.
  - Apply different retrieval algorithms to the same database for the same query. The aggregate of relevant items is taken as the total relevant set.

Trade-off between Recall and Precision



Class Exercises