Defining AI Hypothesis

Dr. Animesh Chaturvedi

Assistant Professor: IIIT Dharwad

Post Doctorate: King's College London & The Alan Turing Institute

PhD: IIT Indore MTech: IIITDM Jabalpur

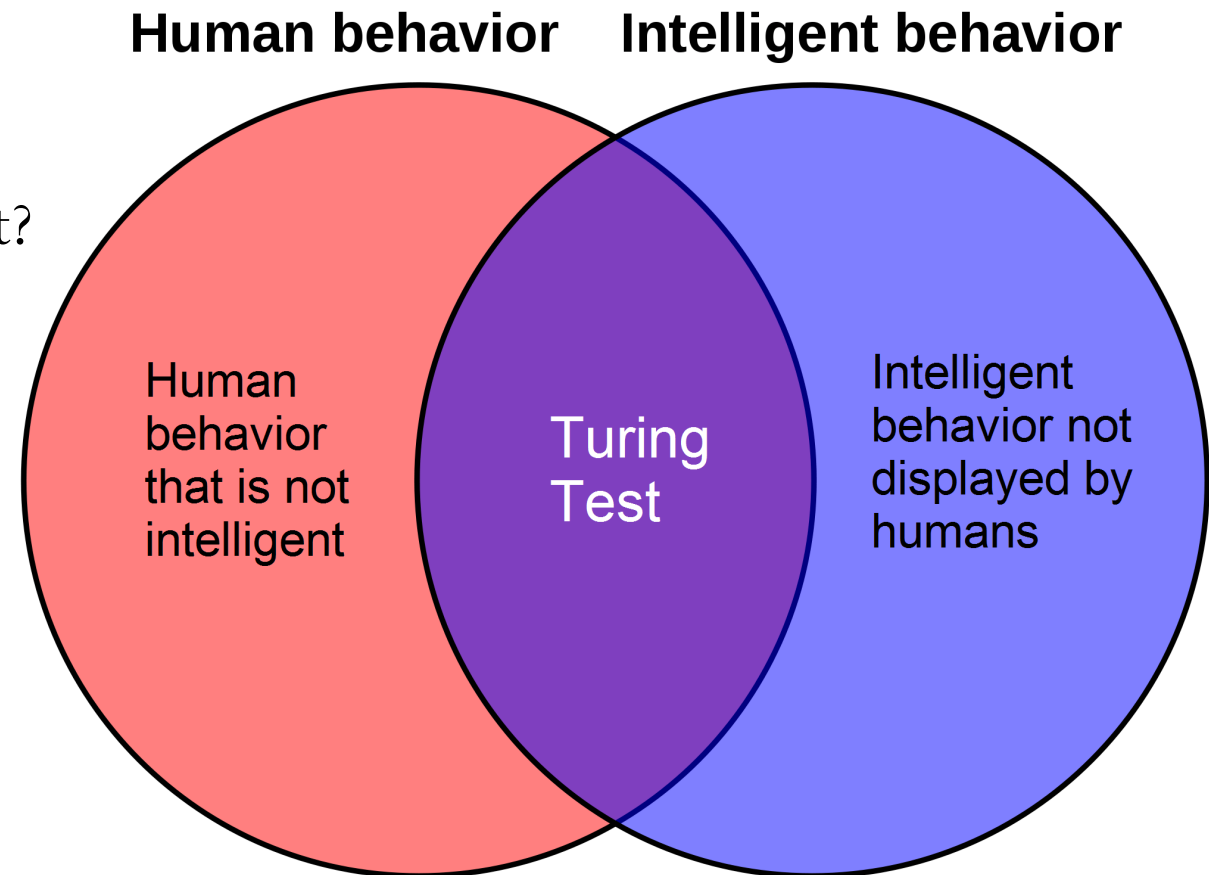# Humanly and Rationally (Thinking and Acting)

Definition of AI

- Physics:
  - Where did the *physical universe* come from?
  - What are **physical laws**?
- Biology:
  - Does the *biological life* evolved?
  - How does organs in **living organisms** works together?
- AI:
  - What is *intelligence* of a living being and non-living being?
  - capacity for learning, reasoning, understanding, and similar forms of mental activity

https://people.eecs.berkeley.edu/~russell/intro.html

What is Intelligence of Living being

- Ability to perceive and act in the world
- Reasoning: proving theorems, medical diagnosis
- Planning: take decisions
- Learning and Adaptation: recommend movies, learn traffic patterns
- Understanding: text, speech, visual scene

https://people.eecs.berkeley.edu/~russell/intro.html

Intelligence vs. humans

- Are humans intelligent?
  - replicating human behavior early hallmark of intelligence
- Are humans ~~always~~ intelligent?
  - not necessary ☺
- Can non-human behavior be intelligent?
  - possible ☺

Human behavior **Intelligent behavior**

Human behavior that is not intelligent

Turing Test

Intelligent behavior not displayed by humans

Rationally vs Humanly

- **Artificial Intelligence is the study of how to make computers do things at the moment, and do better**

- **AI important elements:**

1. Systems that think like humans
2. Systems that act like humans
3. Systems that think rationally
4. Systems that act rationally

Human-like vs Rational
Thought vs Behavior

Human-like vs Rational

	Human-like	Rational
Thought vs Behavior (Thought)	``The exciting new effort to make **computers think** … *machines with minds*, in the full and literal sense'' (Haugeland, 1985)``The automation of activities that we associate with **human thinking**, activities such as decision-making, problem solving, learning …'' (Bellman, 1978)	``The study of **mental faculties** through the use of **computational models**'' (Charniak and McDermott, 1985)``The study of the computations that make it possible to **perceive, reason, and act**'' (Winston, 1992)
(Behavior)	``The art of creating machines that perform functions that require **intelligence when performed by people**'' (Kurzweil, 1990)``The study of how to make computers do things at which, at the moment, **people are better**'' (Rich and Knight, 1991)	``A field of study that seeks to **explain and emulate intelligent behavior** in terms of computational processes'' (Schalkoff, 1990)``The branch of computer science that is concerned with the **automation of intelligent behavior**'' (Luger and Stubblefield, 1993)

Thought	Systems that think like humans.	Systems that think rationally.
Behavior	Systems that act like humans	Systems that act rationally

https://people.eecs.berkeley.edu/~russell/intro.html

Intelligent Computer

- Possess the following capabilities:
  - **Natural Language Processing** to enable it to communicate successfully in English (or some other human language);
  - **Knowledge Representation** to store information provided before or during the interrogation;
  - **Automated Reasoning** to use the stored information to answer questions and to draw new conclusions;
  - **Machine Learning** to adapt to new circumstances and to detect and extrapolate patterns.
  - **Computer vision** to perceive objects, and
  - **Robotics** to move them about.

Thinking and Acting

- Two thoughts
  - Humanly: concerned with reasoning steps of human subjects solving the same problems.
  - Rationally: concerned with getting the right answers regardless of how humans might do it.

Thinking humanly

- Cognitive Science and Modelling:
  - Determining how humans think and workings of human minds
  - Hard to understand how humans think
  - Two ways
    - Introspection --trying to catch our own thoughts
    - Psychological experiments
- Do we want a machine that beats humans in chess or a machine that thinks like humans while beating humans in chess?
  - Deep Blue supposedly DOESN'T think like humans..
  - Express the theory as a computer program behavior matches human behavior
- Thinking like humans important in Cognitive Science applications
 - Intelligent tutoring
  - Expressing emotions in interfaces… HCI

https://people.eecs.berkeley.edu/~russell/intro.html

Thinking Rationally

- **Right Thinking:** "arguments structures that always gave correct conclusions given correct premises"

- **Syllogism:** "Socrates is a man; all men are mortal therefore Socrates is mortal."

- **Field of Logic:** Laws of thought were supposed to govern the operation of the mind.
  - Logical notation to find solution to a problem.
  - Finding solution to all kinds of things in the world and the relations between them.

- **Issue 1:** Not easy to take informal knowledge and state it in the formal terms required by Logical notation, particularly when the knowledge is less than 100% certain.

- **Issue 2:** There is a big difference between being able to solve a problem "in principle'" and doing so in practice. Means proposing algorithm and it's coding are different problems.

- Power of the representation and reasoning systems

https://people.eecs.berkeley.edu/~russell/intro.html

Acting Humanly

- Loebner Prize
  - Every year in Boston
  - an annual contest based on the Turing Test
  - <mark>an annual competition in artificial intelligence that awards prizes to the computer programs considered by the judges to be the most human-like.</mark>
- Problems
  - <mark>Not reproducible, constructive, or mathematically analyzable</mark>
  - <mark>Make human-like errors</mark>
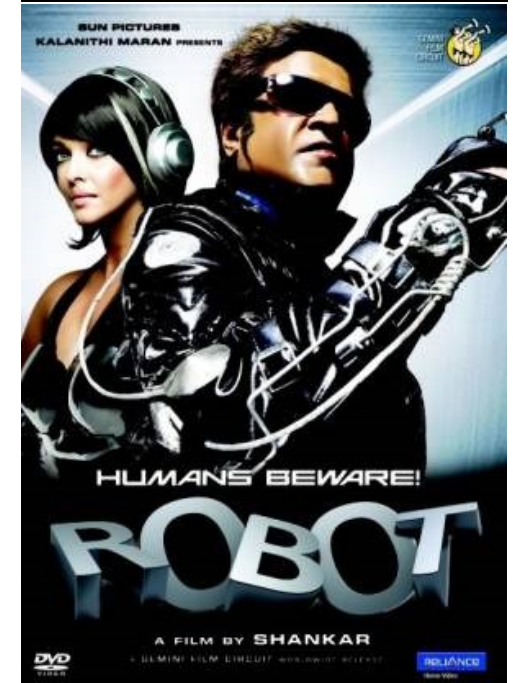
Acting Rationally

- Acting to achieve one's goals, given one's beliefs.

- Rational behavior: doing the right thing

- Need not always be deliberative
  - Reflexive

- Every art and every inquiry, and similarly every action and every pursuit is thought to aim at some good.
  - By Aristotle (Nicomachean ethics)

Hypothesis space:
Problem and Solution domain formulations

Machine Acting V/S Thinking?

- Weak AI Hypothesis vs. Strong AI Hypothesis
  - Weak Hyp: machines could act as if they are intelligent
  - Strong Hyp: machines that act intelligent have to think intelligently too
- Hypothesis:
  - "a supposition or proposed explanation made on the basis of limited evidence as a starting point for further investigation"
  - "a proposition made as a basis for reasoning, without any assumption of its truth"
- Machine learning:
  - "algorithms to build model using data for training, which makes machine capable enough to make predictions or decisions without being explicitly programmed"
- What is Machine learning Hypothesis?

Machine Learning

- **Machine Learning do Prediction or Recommendation**:
  - **Given** examples of a function *(X, F(X))*
  - **Predict** function *F(X)* for new examples *X*
- Classification
  F(X) = Discrete
- Regression
  F(X) = Continuous
- Probability estimation
  F(X) = Probability*(X):*
- Feature Space: Properties that describe the problem •
- Given: <x, f(x)> for some unknown function f
- Learn: A hypothesis H, that approximates f
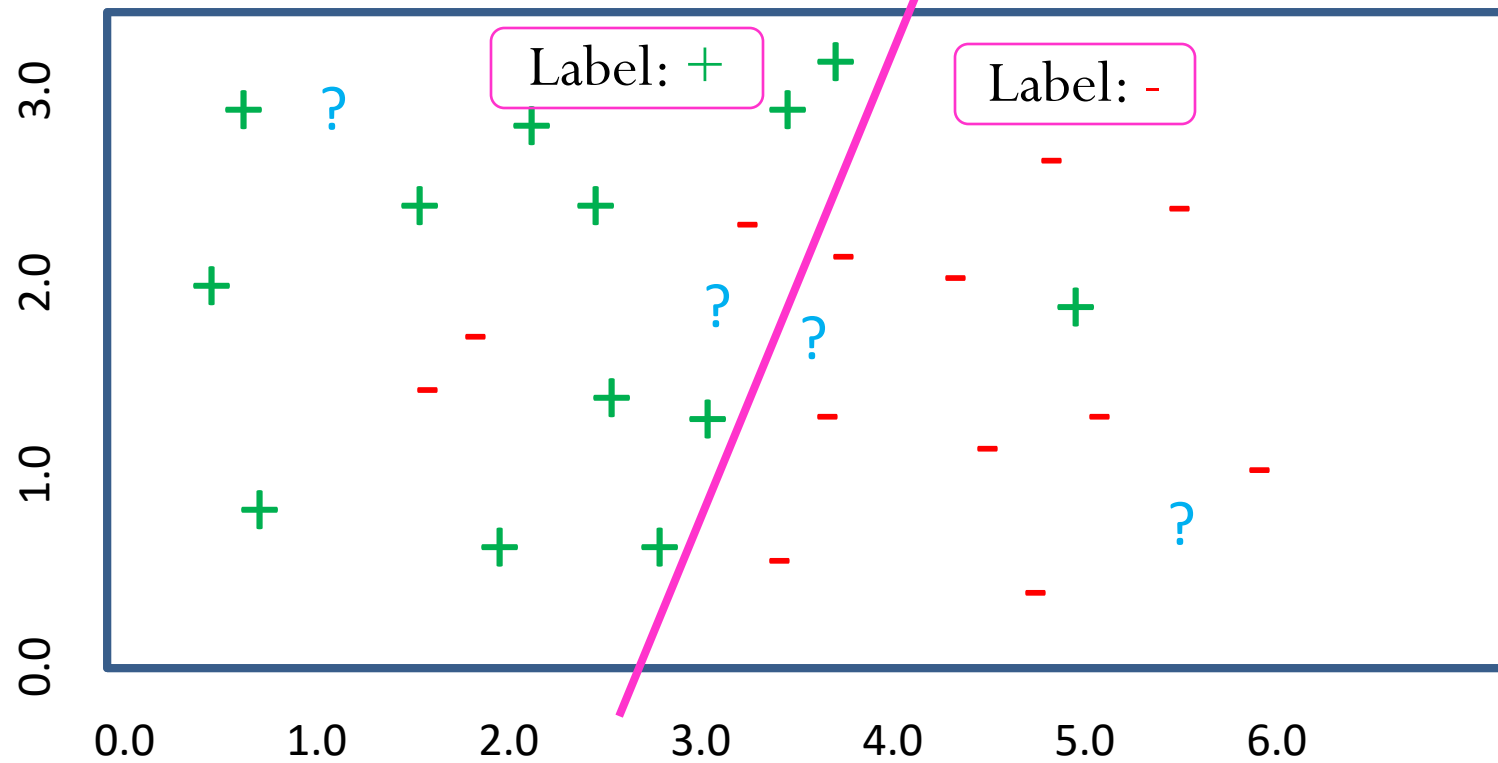
Terminology
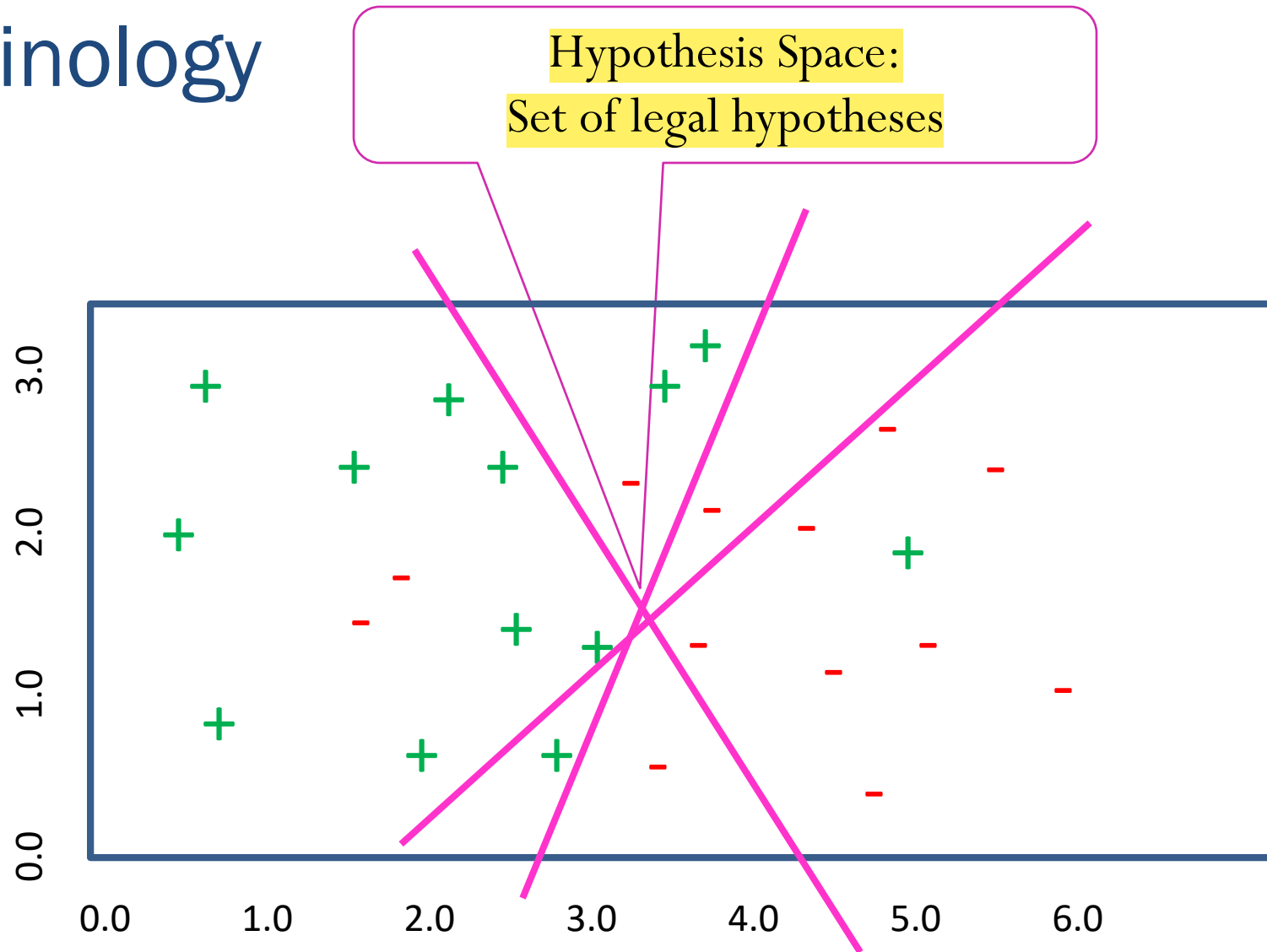
Inductive Bias

- Need to make assumptions
  - Experience alone doesn't allow us to make conclusions about unseen data instances

- Two types of bias:
  - Restriction: Limit the hypothesis space (e.g., look at rules)   *give a priori-information*
  - Preference: Impose ordering on hypothesis space (e.g., more general, consistent with data)   *(some hypothesis are preferred over others)*

- Statistical hypothesis test: Comparing and Validating Machine Learning Algorithms

Hypothesis Testing in ML Machine learning models are chosen based on their mean performance, often calculated using k-fold cross-validation. The algorithm with the best mean performance is expected to be better than those algorithms with worse mean performance.

https://codinghero.ai/hypothesis-testing-in-ml-explained-to-kids/#:~:text=ML%20models%20use%20many%20statistical%20concepts%20and%20one,better%20than%20those%20algorithms%20with%20worse%20mean%20performance.

Slides of Artificial Intelligence by Jesse Davis, jdavis@cs.washington.edu https://courses.cs.washington.edu/courses/cse573/08au/slides/

Confusion Matrix

- True Positive (TP):
  - A test result that correctly indicates the presence of a condition or characteristic
- True Negative (TN):
  - A test result that correctly indicates the absence of a condition or characteristic
- False Positive (FP):
  - A test result which wrongly indicates that a particular condition or attribute is present
- False Negative (FN):
  - A test result which wrongly indicates that a particular condition or attribute is absent

$$P = TP + FN \qquad N = FP + TN$$

Confusion Matrix

		Predicted condition	
	Total population = P + N	Positive (PP)	Negative (PN)
Actual condition	Positive (P)	True positive (TP)	False negative (FN)
	Negative (N)	False positive (FP)	True negative (TN)

$$P = TP + FN \qquad N = FP + TN$$
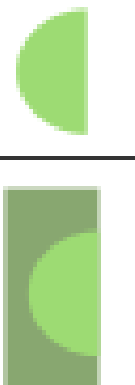
Precision and Recall

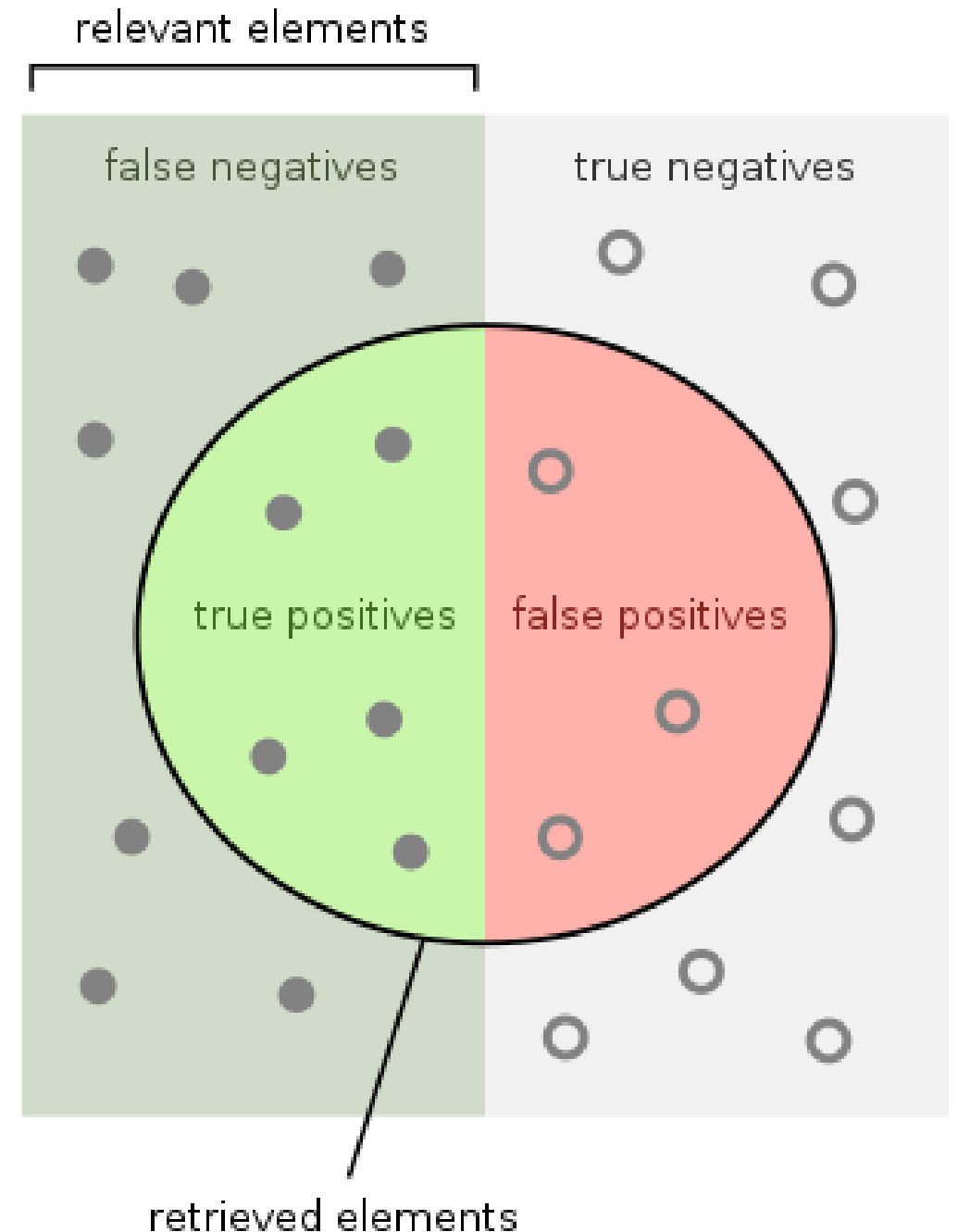How many **retrieved** items are relevant?

How many **relevant** items are retrieved?

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{TN}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}}$$

relevant elements

false negatives

true negatives

true positives

false positives

retrieved elements

Precision and Recall

- A set of retrieved documents
  - (e.g. the list of documents produced by a web search engine for a query)
- A set of relevant documents
  - (e.g. the list of all documents on the internet that are relevant for a certain topic)

$$\text{precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

$$\text{recall} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{relevant documents}\}|}$$

Accuracy and F-score

$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Sensitivity and Specificity

- **Recall** is also referred to as the **True Positive Rate** or **Sensitivity**

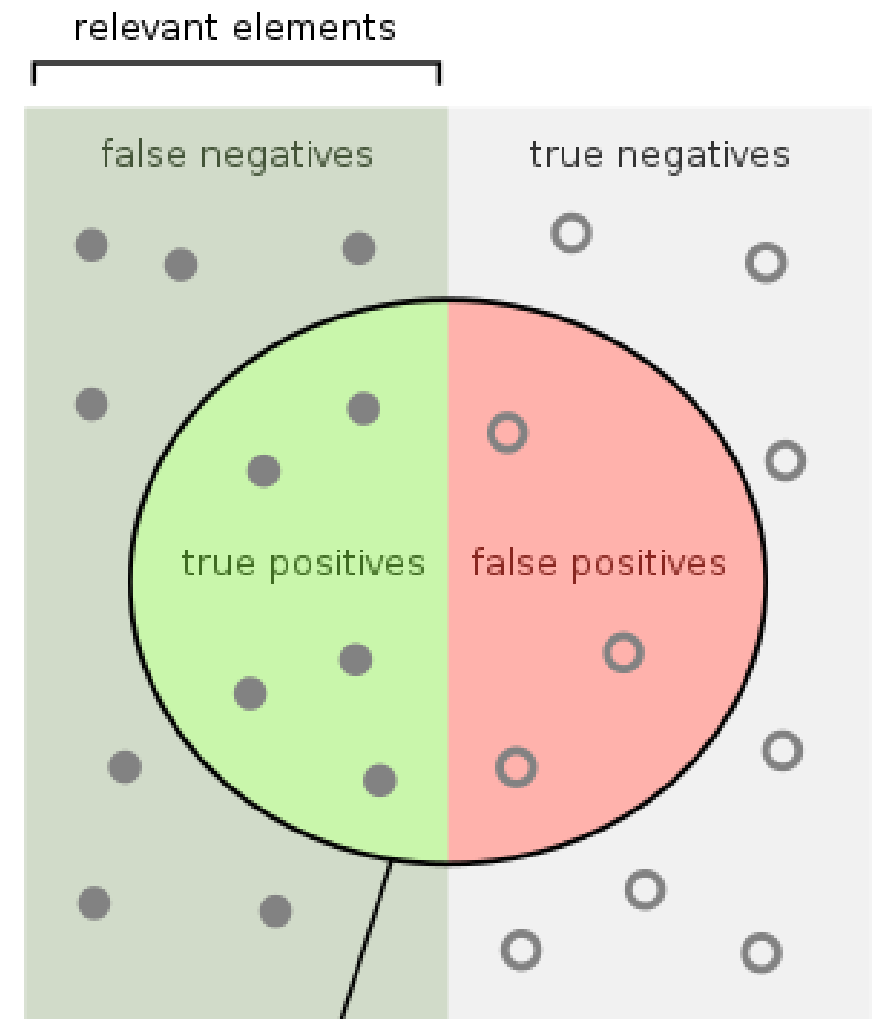- **Precision** is also referred to as **Positive Predictive Value (PPV)**

$$\text{PPV} = \frac{\text{Number of true positives}}{\text{Number of true positives} + \text{Number of false positives}} = \frac{\text{Number of true positives}}{\text{Number of positive calls}}$$

- **True negative rate** is also called **Specificity**

$$\text{True negative rate} = \frac{tn}{tn + fp}$$

References

- Slides of Artificial Intelligence by Jesse Davis, jdavis@cs.washington.edu https://courses.cs.washington.edu/courses/cse573/08au/slides/
- Stuart Russel, and Peter Norvig. "Artificial intelligence: A modern approach. Third edit." Upper Saddle River, New Jersey 7458 (2015). http://aima.cs.berkeley.edu/
- https://people.eecs.berkeley.edu/~russell/intro.html
- Wikipedia contents
  - https://en.wikipedia.org/wiki/Precision_and_recall
  - https://en.wikipedia.org/wiki/Confusion_matrix
- Images are from several sources e.g. movies, TV serials, internet, miscellaneous links, slides, blogs, etc.
- Artificial Intelligence, Prof. Mausam https://www.cse.iitd.ac.in/~mausam/courses/col333/autumn2019/

ขอบคุณ
Thai

Grazie
Italian

תודה רבה
Hebrew

Gracias
Spanish

Спасибо
Russian

English

Thank You

Obrigado
Portuguese

شكراً
Arabic

https://sites.google.com/site/animeshchaturvedi07

Merci
French

多謝
Traditional
Chinese

Danke
German

धन्यवाद
Hindi

நன்றி
Tamil
Tamil

多谢
Simplified
Chinese

ありがとうございました
Japanese

감사합니다
Korean