DATA 1301
Introduction to Data Science

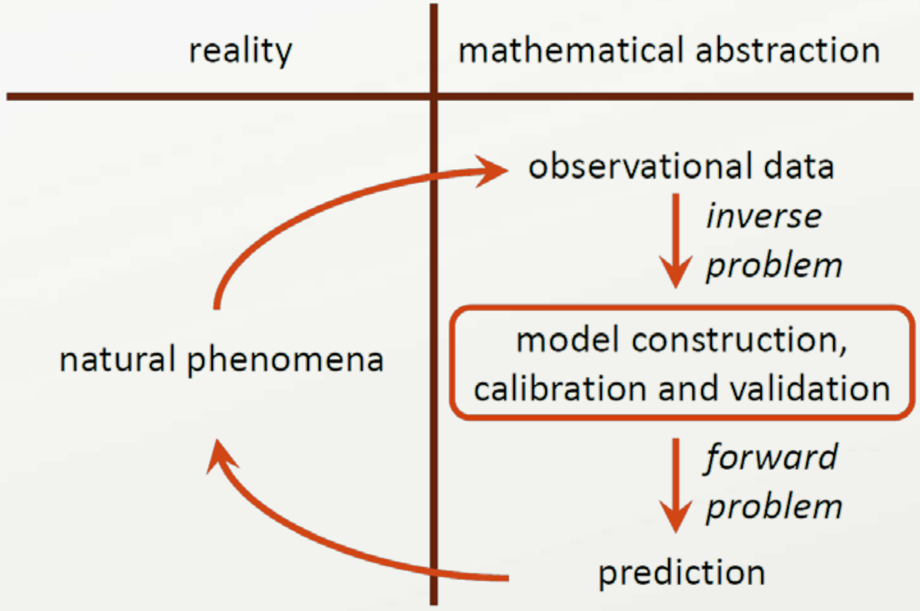
Amir Shahmoradi

Department of Physics / College of Science
Data Science Program / College of Science
The University of Texas
Arlington, Texas

The two classical pillars of science: Experiment and Theory

How do we make a scientific inference?

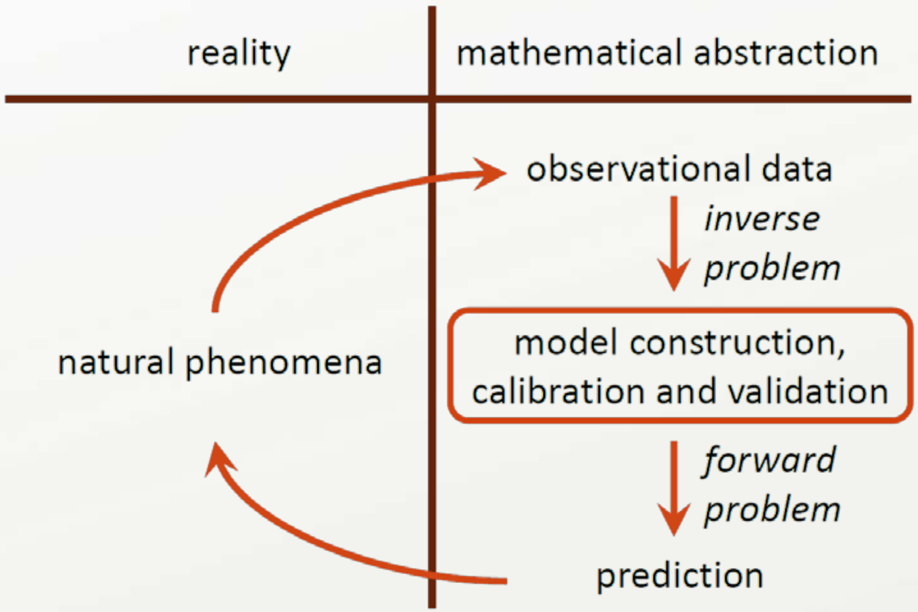
A very elementary depiction of the scientific method



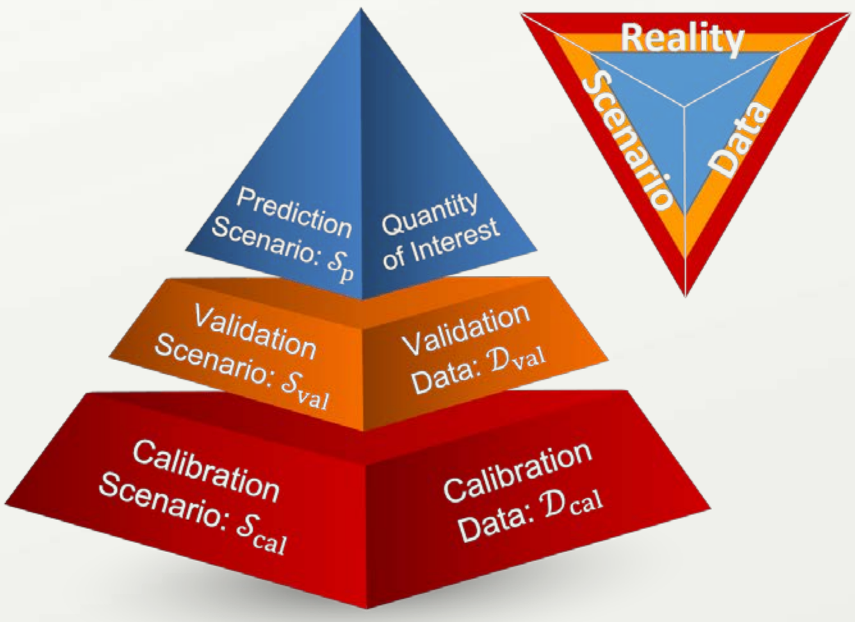
The two classical pillars of science: Experiment and Theory

How do we make a scientific inference?

A very elementary depiction of the scientific method



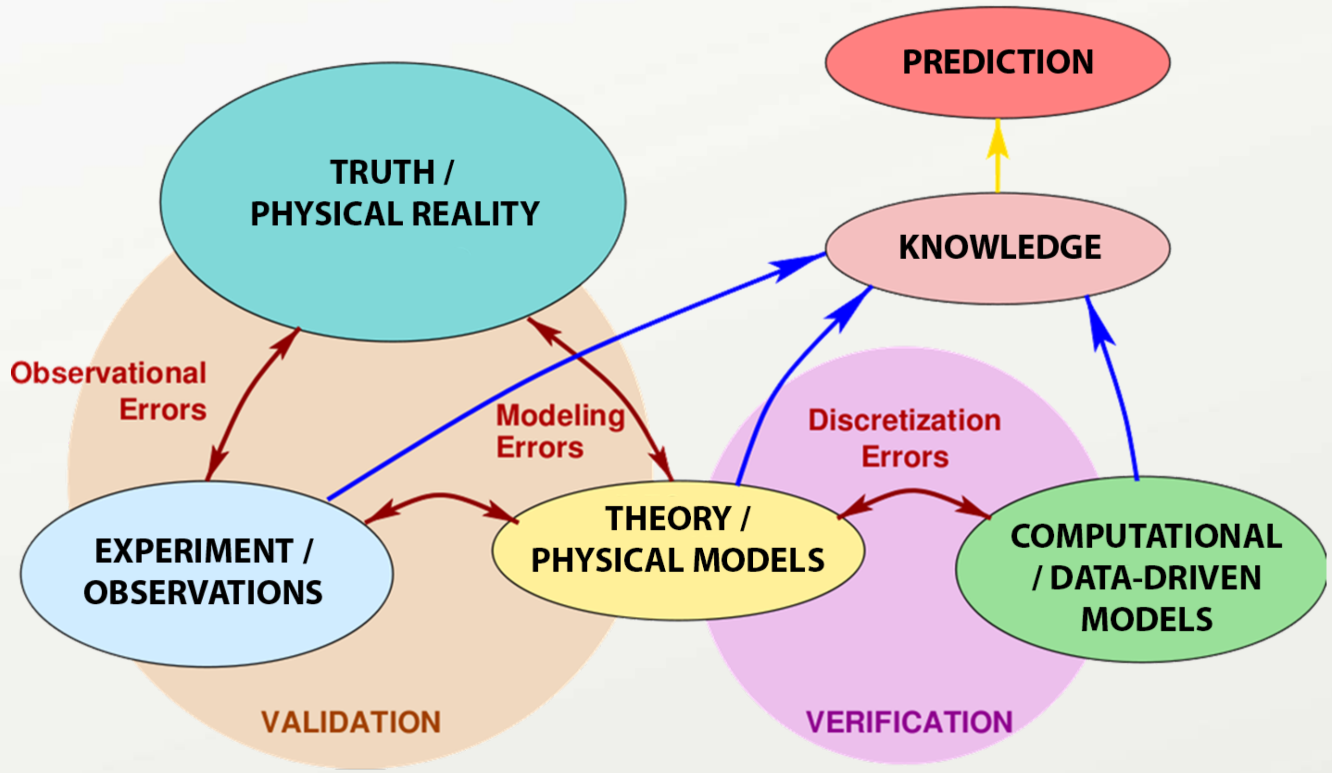
The prediction pyramid



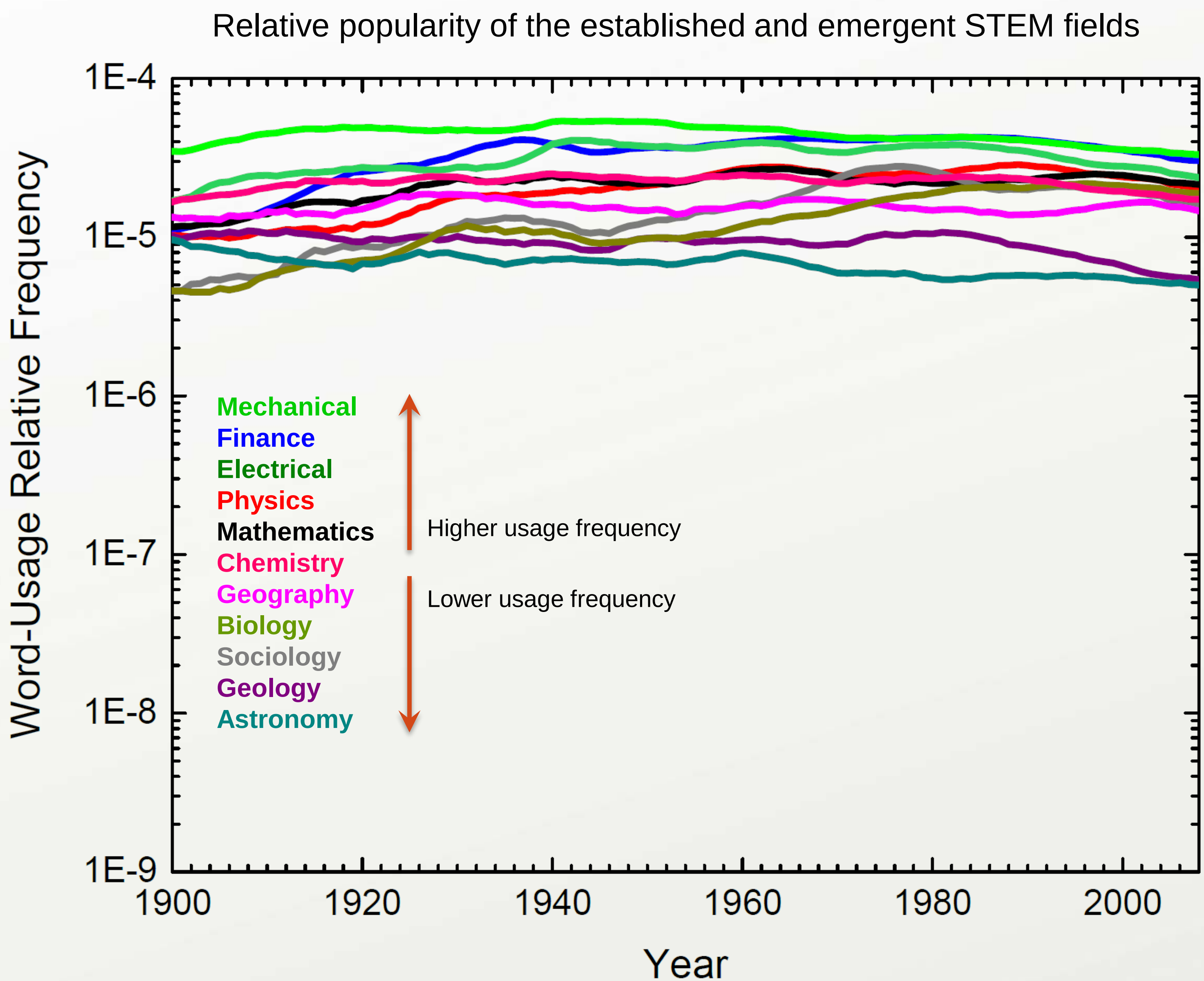
The three pillars of science: Experiment, Theory, Computation+Data

Three major roles of computational models in contemporary science

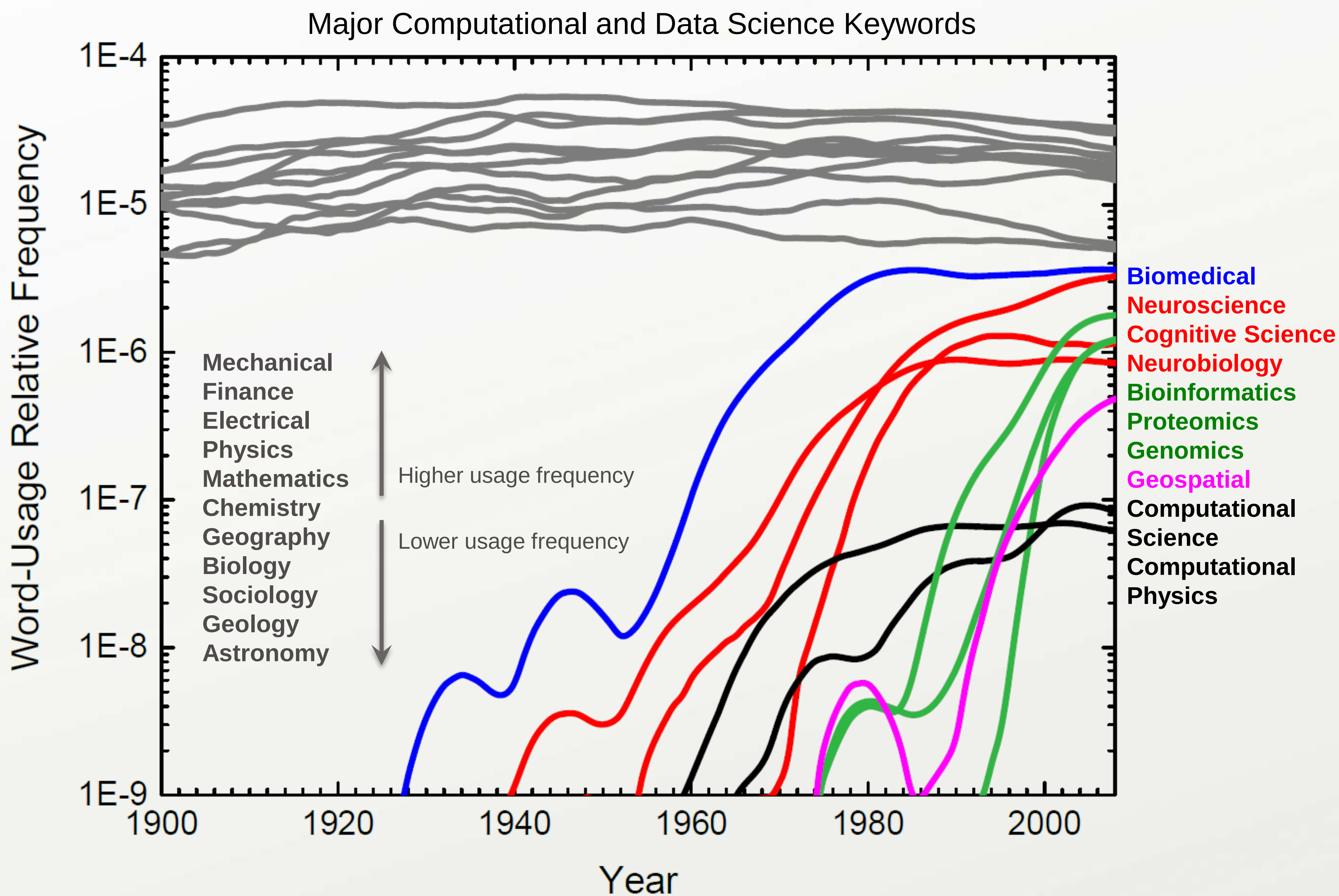
1. a workflow bridging data, hypothesis/theory, and prediction (**predictive computing**).
2. a substitute for experiment and observational data, where it is not available (**numerical simulation**).
3. a substitute for theory, where it is not available (**data-driven discovery via machine learning**, **deep learning, …**).



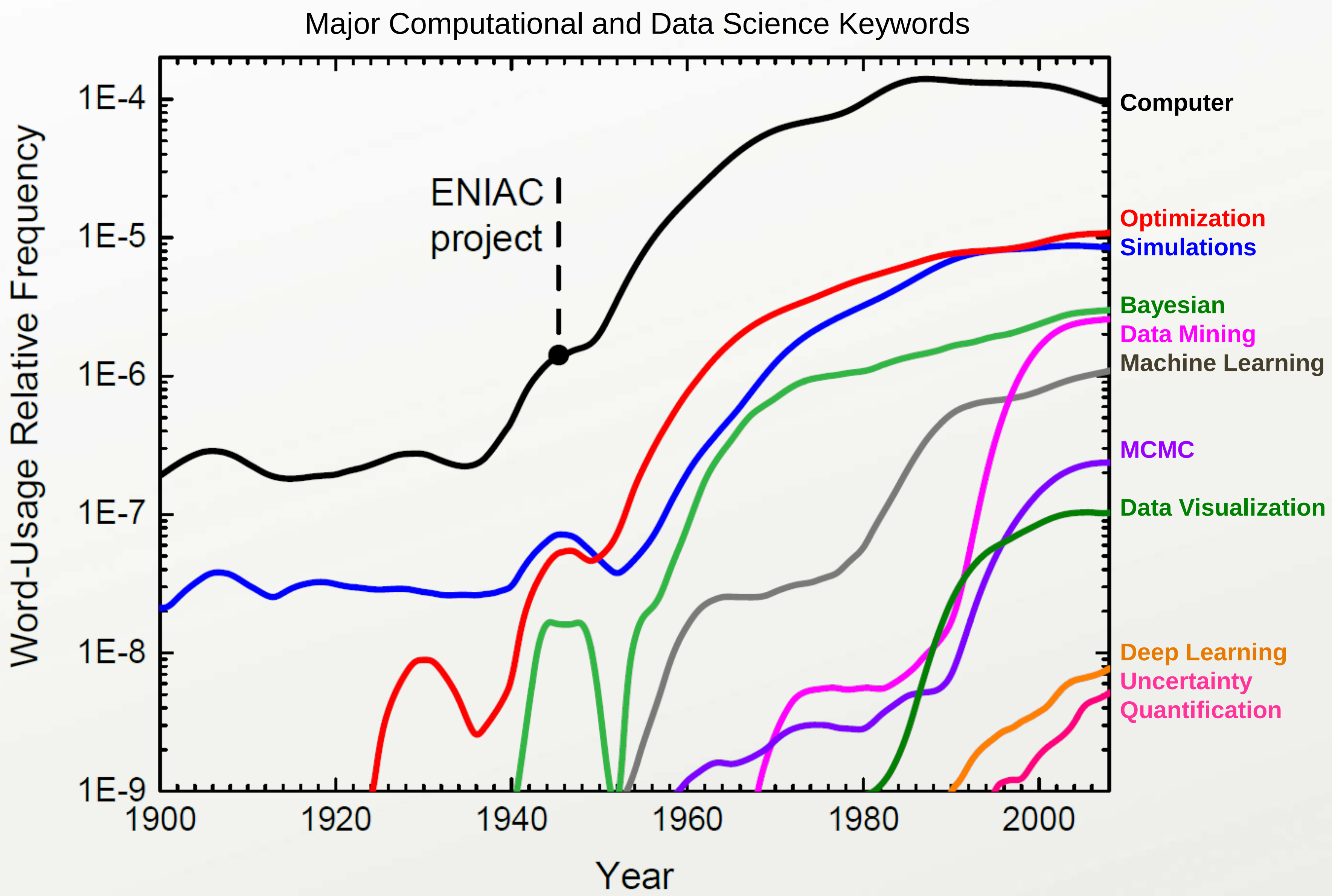
The emerging third pillar of science: Computational and Data Sciences



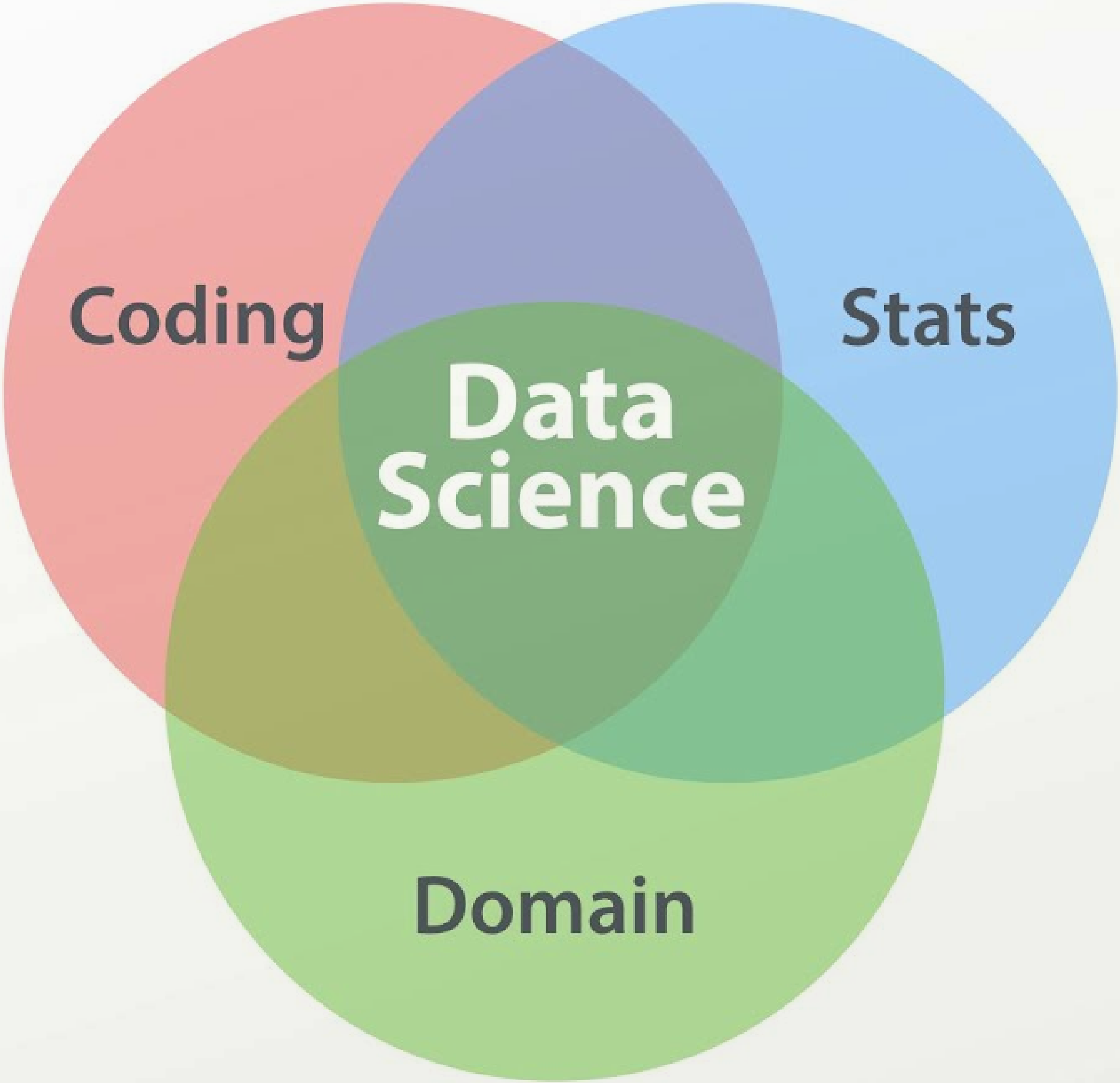
The emerging third pillar of science: Computational and Data Sciences



The emerging third pillar of science: Computational and Data Sciences



Learning Data Science competencies via Word Embeddings



The two types of reasoning in Science: Deductive and plausible reasoning

This is the kind of reasoning we would like to use all the time; but, in almost all real-world situations confronting us we do not have the right kind of information to allow this kind of reasoning. Therefore we must often use weaker logical arguments.

Deductive reasoning

$$\frac{\text{if } A \text{ is true, then } B \text{ is true} \qquad A \text{ is true}}{\text{therefore, } B \text{ is true,}}$$

$$\frac{\text{if } A \text{ is true, then } B \text{ is true} \qquad B \text{ is false}}{\text{therefore, } A \text{ is false.}}$$

The two types of reasoning in Science: Deductive and plausible reasoning

Plausible reasoning

$$\frac{\begin{array}{c}\text{if } A \text{ is true, then } B \text{ is true} \\ B \text{ is true}\end{array}}{\text{therefore, } A \text{ becomes more plausible.}}$$

The evidence does not prove that A is true, but verification of one of its consequences does give us more confidence in A.

For example,

A ≡ it will start to rain by 10 am at the latest;

B ≡ the sky will become cloudy before 10 am.

The two types of reasoning in Science: Deductive and plausible reasoning

Plausible reasoning

$$\frac{\begin{array}{c}\text{if } A \text{ is true, then } B \text{ is true} \\ B \text{ is true}\end{array}}{\text{therefore, } A \text{ becomes more plausible.}}$$

The evidence does not prove that A is true, but verification of one of its consequences does give us more confidence in A.

For example,

A ≡ it will start to rain by 10 am at the latest;

B ≡ the sky will become cloudy before 10 am.

Observing clouds at 9:45 am does not give us a logical certainty that the rain will follow; nevertheless our common sense, obeying the weak syllogism, may induce us to change our plans and behave as if we believed that it will, if those clouds are sufficiently dark.

The two types of reasoning in Science: Deductive and plausible reasoning

logical implication and physical causation are NOT the same.

The premise ‘if A then B’ expresses B only as a logical consequence of A; and not necessarily a causal physical consequence, which could be effective only later. The rain at 10 am is not the physical cause of the clouds at 9:45 am. Nevertheless, the proper logical connection is not in the uncertain causal direction (clouds ⇒ rain), but rather (rain ⇒ clouds), which is certain, although noncausal.

We could have even weaker logical arguments, like,

$$\frac{\begin{array}{c}\text{If } A \text{ is true, then } B \text{ becomes more plausible}\\ B \text{ is true}\end{array}}{\text{therefore, } A \text{ becomes more plausible.}}$$

The Boolean algebra

To state these ideas more formally, we introduce some notation of the usual symbolic logic, or Boolean algebra, so called because George Boole (1854) introduced a notation like the following. Of course, the principles of deductive logic itself were well understood centuries before Boole, and, as we shall see, all the results that follow from Boolean algebra were contained already as special cases in the rules of plausible inference given more than a century before Bool, by Laplace (1812).

The symbol

$$AB,$$

called the **logical product** or the **conjunction**, denotes the proposition '**both A and B are true**'. Obviously, the order in which we state them does not matter; AB and BA say the same thing. The expression

$$A + B,$$

called the **logical sum** or **disjunction**, stands for '**at least one of the propositions, A, B is true**' and has the same meaning as B + A. These symbols are only a shorthand way of writing propositions, and do not stand for numerical values. Given two propositions A, B, it may happen that one is true if and only if the other is true; we then say that they have the same **truth value**.

The Boolean algebra

Evidently, then, it must be the most primitive axiom of plausible reasoning that **two propositions with the same truth value are equally plausible**. This might appear almost too trivial to mention, were it not for the fact that Boole himself (Boole, 1854, p. 286) fell into error on this point, by mistakenly identifying two propositions which were in fact different – and then failing to see any contradiction in their different plausibilities. Three years later, Boole (1857) gave a revised theory which supersedes that in his earlier book.

In Boolean algebra, the equal sign is used to denote equal truth value

$$A = B,$$

and the 'equations' of Boolean algebra thus consist of assertions that the proposition on the left-hand side has the same truth value as the one on the right-hand side. **The symbol '≡' means, as usual, 'equals by definition'**.

In denoting complicated propositions we use parentheses in the same way as in ordinary algebra, i.e. to indicate the order in which propositions are to be combined. For example,

$$AB + C \text{ denotes } (AB) + C; \text{ and not } A(B + C)$$

The Boolean algebra

The **denial** of a proposition is indicated by a bar:

$\bar{A} \equiv$ A is false. (1.8)

The relation between A and $\bar{A}$ is a reciprocal one:

A = $\bar{A}$ is false,

and it does not matter which proposition we denote by the barred and which by the unbarred letter. Note that some care is needed in the unambiguous use of the bar. For example, according to the above conventions,

$\overline{AB}$ = AB is false;

$\overline{AB}$ = both A and B are false.

However, these are quite different propositions; in fact, $\overline{AB}$ is not the logical product $\bar{A}\,\bar{B}$, but the logical sum:

$$\overline{AB} = \bar{A} + \bar{B}\ .$$

The Boolean algebra's fundamental identities

With these understandings, Boolean algebra is characterized by some rather trivial and obvious basic identities, which express the properties of

$$\text{Idempotence:} \quad \begin{cases} AA = A \\ A + A = A \end{cases}$$

$$\text{Commutativity:} \quad \begin{cases} AB = BA \\ A + B = B + A \end{cases}$$

$$\text{Associativity:} \quad \begin{cases} A(BC) = (AB)C = ABC \\ A + (B + C) = (A + B) + C = A + B + C \end{cases}$$

$$\text{Distributivity:} \quad \begin{cases} A(B + C) = AB + AC \\ A + (BC) = (A + B)(A + C) \end{cases}$$

$$\text{Duality:} \quad \begin{cases} \text{If } C = AB\text{, then } \overline{C} = \overline{A} + \overline{B} \\ \text{If } D = A + B\text{, then } \overline{D} = \overline{A}\,\overline{B} \end{cases}$$

but by their application one can prove any number of further relations, some highly nontrivial.

The Boolean algebra's fundamental identities

Implication

The proposition

$$A \Rightarrow B$$

to be read as 'A implies B', does not assert that either A or B is true; it means only that

$A\bar{B}$ is false,

or, the same thing,

$(A + B)$ is true.

This can be written also as the logical equation

$$A = AB.$$

That is, if A is true then B must be true; or, if B is false then A must be false.
On the other hand,
if A is false, $A \Rightarrow B$ says nothing about B, and
if B is true, $A \Rightarrow B$ says nothing about A.