What is statistics?

- Machine Learning
  - Algorithms
  - Distribution free
  - Finite samples v.s. asymptotics
- AI
  - beyond humans/animals
  - math
- Algorithms & complexity
- Calculus
- Lin. Algebra
- Basic stats & probability theory

Learning

Supervised - Semi-supervised - Unsupervised

Cooperative - indifferent - adversarial teacher

Passive - active learner

Batch - Online learning

What is learning?

experience → expertise

- Learner : a human or a computer:
  interacts with the world she wants to learn about via data (observations / examples / samples)

  Data : finite initial segments
  $z^t := z_1, \ldots, z_t$ of an infinite data stream z^ω.

---

bait shyness

Generalization (inductive reasoning / inductive inference : inference that progresses from given examples to hitherto unknown examples and to general observational statements.

Pigeon superstition

Overfitting

Prior knowledge / inductive bias:

background assumptions that Restrict the space of possible outcomes.

Hypothesis: general statement about the world

Statistical hypothesis: probability dist. over a sample space.

Statistical model: set of statistical hypotheses.

Example: coin tossing. Learner sees z^t

model: Bernoulli random variable $\theta \in [0,1]$. Hyp: 'the coin is fair'.

→ statistical hypothesis: $z \overset{iid}{\sim} \text{Bern}(\frac{1}{2})$

Objectives of learning

1) Estimation
2) Predict (/classify)

3) Testing

Exploratory / Confirmatory research

'Exploratory research is about finding the question! In confirmatory research the validity of an existing hypothesis is tested.

Ex. Coin tossing $z^t \sim \text{Bern}(\theta)$

1) Estimate θ

2) Predict the next outcome

3) Test whether it is fair or not.

A method or rule for inference

Frequentist paradigm

Bayesian paradigm

Bayes theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

https://kiavmedicalvalue.github.io/Exercises
Lecture1.pdf