2. Rationality

11 May 2019 11:42

Inductive bias: factors that lead a learner to favor 1 hypothesis over another
- necessary to generalise in a useful/rational way

Heuristic Bias: departures from normal rational theory; mostly violations of basic laws of probability
- focus on one aspect to ignore others

OVERVIEW

1) Framing effects
2) Representativeness
3) Availability
4) Base rate neglect

FRAMING EFFECTS

Bet 1
A. Win £240
B. 25% win £1000, 75% win nothing

Bet 2
C. Lose £750
D. 25% Lose nothing, 75% lose £1000

What might people be expected to do?

Expected Utility: $\sum_{i}^{n} P(O_i)\, U(O_i)$

$P(O)$ is the probability of the outcome,
$U(O_i)$ is the utility

Risk Aversion: We're willing to lose some money to reduce the risk of losing everything
- People chose A+D even though its strictly dominated by B+C

PROSPECT THEORY

1) We assign diminishing values to gains/losses - $U(+120) - U(+110) < U(+20) - U(10)$
2) Gains diminish more quickly; large losses more important - $U(100) < -U(-100)$
3) We overweigh improbable events - prone to lotteries

REPRESENTATIVNESS

Availability Heuristic: We estimate probabilities by recalling examply
Base Rates: bias in favor of explainations representative of the report.
→ more likely to favor conjunctions of outcomes than things in themselves
Rationality?: People choose options that give them information or minimise uncertainty

RATIONAL ANALYSIS

Ways of understanding Cognition
1) Bottom up from biology/chemistry
2) Mechanistic/Algorithmic explainations
3) What would a goal orientated rational solution look like?

↳ Andersons Elements of Rational analysis
1) Specify the goals of the cognitive system (minimise uncert., infer features, maximise utility)
2) Specify the environment (assumptions that entails; inductive biases, prior knowledge...)
3) Specify necessary computational limits (memory, time etc.)
4) Derive optimal set of behaviors from 1-3.

INDUCTIVE BIAS

"Soft" relative weightings. "Hard" rules/constraints, Implicit vs Explicit Bias
- Social pragmatic cues: label things their gazing at
- Whole object assumption: more likely to refer to whole objects than parts
- Taxanomic assumptions: words refer things in same category