IIT Jodhpur

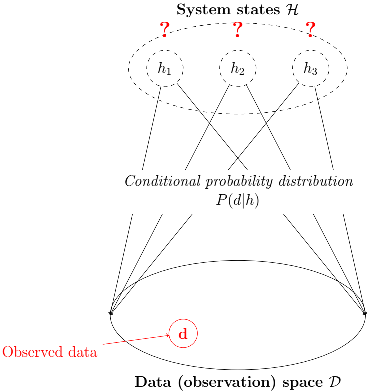
Biological Vision and Applications
Module 03-04: More on Bayesian reasoning

Hiranmay Ghosh

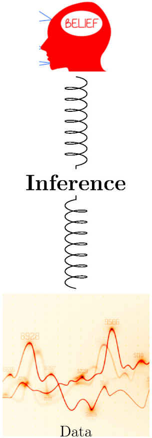
Bayesian Inference

Summary

- Hypothesis space: $\mathcal{H} = \{h_1, h_2 \dots h_m\}$
- Observable space: $\mathcal{D} = \{d_1, d_2 \dots d_n\}$
- Prior knowledge:
  - Prior probabilities: $P(h_1), P(h_2), \dots P(h_m)$
  - Conditional probabilities: $P(d_1 \mid h_1), P(d_2 \mid h_1) \dots P(d_n \mid h_m)$
- Observed data: $\mathrm{d} \in \mathcal{D}$
- Bayes formula: $P(h_i \mid \mathrm{d}) = \frac{P(\mathrm{d}|h_i).P(h_i)}{P(\mathrm{d})} = \frac{1}{\kappa}.P(\mathrm{d} \mid h_i).P(h_i)$
- Inference by best explanation (abduction):
  - $h^* = \mathrm{argmax}_{h_i \in \mathcal{H}} P(h_i \mid \mathrm{d})$



Prior knowledge and evidence



- Bayesian formula: $P(h_i \mid \mathrm{d}) = \frac{1}{\kappa}.P(\mathrm{d} \mid h_i).P(h_i)$
  - $P(h_i)$ represents prior belief in h_i
  - $P(\mathrm{d} \mid h_i)$ represents the evidential strength in support of h_i
  - Posterior belief $P(h_i \mid \mathrm{d})$ is the product of the two terms
- Bayesian inference is a synthesis of prior knowledge and evidence from observation
  - Key advantage over pure data-driven (machine learning) approach
  - Strong prior belief: Takes lots of evidence to offset it
  - Weak prior belief: Susceptible to noisy data

Odds

- Compare two hypotheses to find which one is more likely than the other
  - $\text{Odds}(h_i, h_j \mid d) = \frac{P(h_i|d)}{P(h_j|d)} = \frac{P(h_i)}{P(h_j)} \cdot \frac{P(d|h_i)}{P(d|h_j)}$
- Operating in logarithmic space makes the model additive
  - $\text{logOdds}(h_i, h_j \mid d) = (\log P(h_i) - \log P(h_j)) + (\log P(d \mid h_i) - \log P(d \mid h_j))$
- h_i is more likely than h_j iff $\text{Odds}(h_i, h_j \mid d) > 1$ or $\text{logOdds}(h_i, h_j \mid d) > 0$

$P(Banana) = 0.8, P(Apple) = 0.2$

		Fruits (A)				
		Banana	Apple	...	...	..
Color (B)	Red	0.1	0.6			
	Green	0.4	0.2			
	Yellow	0.5	0.2			
	Total	*1*	*1*			

- $P(B|Y) = 0.4 \times k$
- $P(A|Y) = 0.04 \times k$
- $\text{Odds}(B, A \mid Y) = \frac{0.4 \times k}{0.04 \times k} = 10$
- $\text{logOdds}(B, A \mid Y) = \log 10 = 1$

Composite data

- Data item d may be composite: $d = (d_1, d_2, \ldots, d_n)$
  - e.g. color, texture, shape
- Combinatorial explosion of data space makes modeling difficult
- Assuming conditional independence
  - $P(d \mid h) = P(d_1 \mid h).P(d_2 \mid h)\ldots.P(d_n \mid h)$
- $P(h_i \mid d) = k.P(h_i).\prod_{k=1}^{n} P(d_k \mid h)$

$$\boxed{\begin{array}{l} \mathrm{Odds}(h_i, h_j \mid d) = \frac{P(h_i)}{P(h_j)} \times \prod_{k=1}^{n} \frac{P(d_k|h_i)}{P(d_k|h_j)} \\ \\ \mathrm{logOdds}(h_i, h_j \mid d) = (P(h_i) - P(h_j)) + \sum_{k=1}^{n}(P(d_k \mid h_i) - P(d_k \mid h_j)) \end{array}}$$

Advantages of modeling with Elementary data items

- Easier to model the statistical dependency of a hypothesis h_i with an elementary data item d_k than the composite d
  - The data space combinatorially expands with number of elementary items
  - Data becomes sparse – there may not be any data available for some rare combinations

- Robust inference can be made with a subset of observations
  - Robust against missing observations
  - Generally, it is possible to use a few discriminatory data elements
  - Wrong observations have less impact
  - Incremental belief update

Example: Robust inference

- To recognize the object as a car, you need not consider all visual features of a car
  - Robust against occlusions, etc.

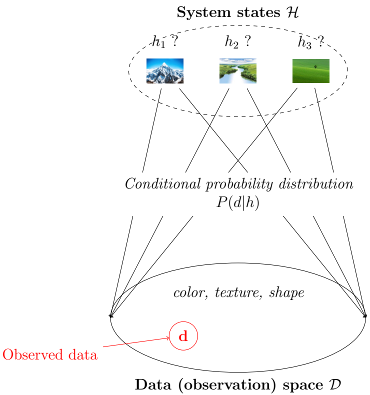
- Can it be done with deductive reasoning?

Incremental belief update

- $P(h_i \mid d) = k.P(h_i).P(d \mid h_i)$
- Assume that $d = d_1, d_2 \dots d_k$ represents a data stream
- After d_1 arrives
  - Posterior: $P(h_i \mid d_1) = k_1.P(h_i).P(d_1 \mid h_i)$
  - This posterior becomes the prior for the second observation
- After d_2 arrives
  - Posterior: $P(h_i \mid d_1, d_2) = k_2.P(h_i \mid d_1).P(d_2 \mid h_i) = k_{12}.P(h_i).P(d_1 \mid h_i).P(d_2 \mid h_i)$
  - This posterior becomes the prior for the third observation
- ...
- System updates it's belief incrementally
- In practice, it may be possible to infer even before all data arrives

Emergent knowledge

- We observe *d*
  - Visual patterns: color, texture, shape
- We infer *h*
  - Semantic concepts: mountain, river, greenery
- The inferred entities are of different kind than the observed entities
- New knowledge is created
- Paradigm applicable to higher layers of cognition also



Limitation of Bayesian reasoning

- We cannot infer an entity unless we have a model for it
  - Was that fruit really a kiwi ?
  - A way to cope up
    - Assume uniform probability distribution (0.33 for each color) to begin with
    - Learn (update probabilities) with experience
- Results are good only if
  - Model (priors, conditionals) is good
  - Data (observation) is good
  - Robust against imperfect models / noisy data

Quiz

Quiz 03-04

End of Module 03-04