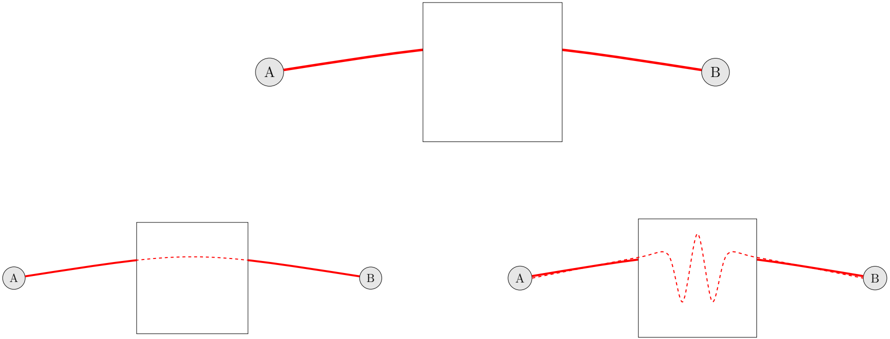
IIT Jodhpur

Biological Vision and Applications
Module 03-06: Occam's razor

Hiranmay Ghosh

Occam's Razor

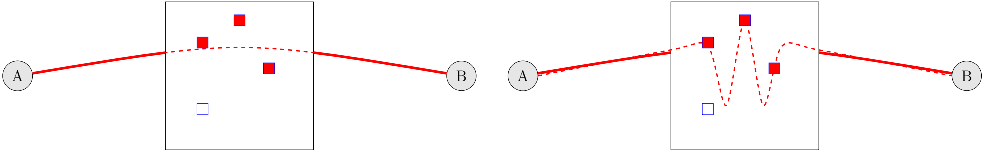
Human mind tends to choose the simplest explanation



- Intuitively, which of the possibilities will you choose ?

Occam's razor

... the observations should be explained

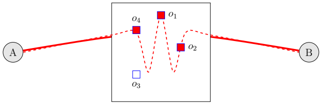


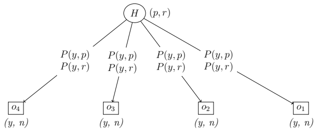
- Which of the possibilities do you choose now ?
- Inference is a tradeoff between complexity of model and goodness of fit
  - Goodness of fit: How well does the model explain the data

EdPuzzle – Occam's razor

Complexity of hypothesis & Goodness of fit

Complexity of hypothesis: $C(p)$ = No. of parameters required to define the curve





Goodness of fit: $GoF = \frac{P(p|D)}{P(r|D)}$

Complexity of explanation: $C(p \mid D) = -log_2(GoF)$

Complexity and belief

- Let $c(M)$ denote the complexity for a proposition M
- Prior belief in model M monotonically decreases with complexity
  - We assume an exponential model: $P(M) = 2^{-c(M)}$, $c(M) = -log_2 P(M)$

- $c(h_i) = -log_2 P(h_i)$: complexity of the hypothesis (prior)
- $c(d \mid h_i) = -log_2 P(d \mid h_i)$: complexity of evidence given the hypothesis (*inverse of goodness of fit*)
- $c(h_i \mid d) = -log_2 P(h_i \mid d)$: complexity of the inference

- Substituting in Baye's theorem $P(h_i \mid d) = \kappa.P(h_i).P(d \mid h_i)$
  - $c(h_i \mid d) = k + c(h_i) + c(d \mid h_i)$

Belief maximization $\equiv$ complexity minimization

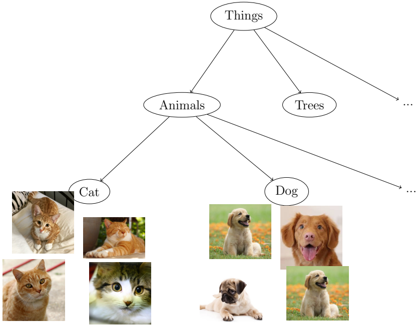
- Bayesian inference: $h^* = \text{argmax}_i \; P(h_i \mid d) = \text{argmax}_i \; P(h_i).P(d \mid h_i)$

- Equivalently: $h^* = \text{argmin}_i \; c(h_i \mid d) = \text{argmin}_i \; [c(h_i) + c(d \mid h_i)]$

- Human mind chooses the inference with least complexity
- Inference is a tradeoff between simplicity of the prior and goodness of fit

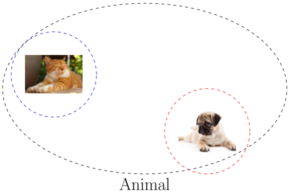
Taxonomy

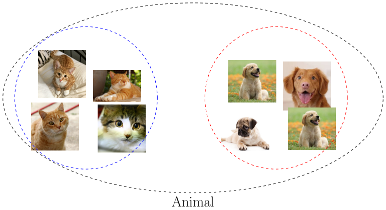
Organizing concepts in a hierarchy



- Learned top-down, or bottom-up ?

Taxonomy Learning

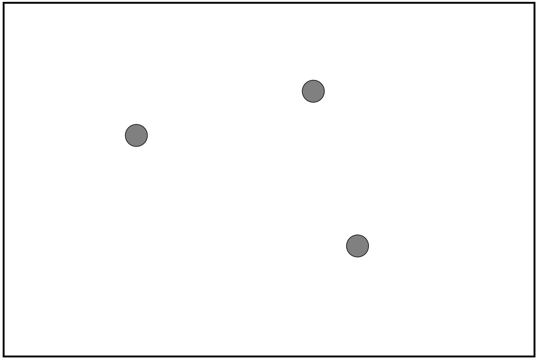




- Taxonomy is a tradeoff between complexity of hypothesis (number of classes) and goodness of fit

Example

Sparse data



Feature space

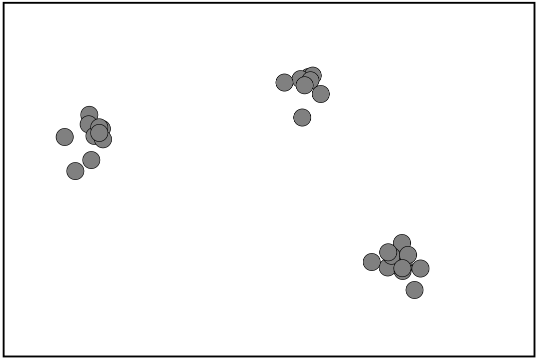
Hypothesis 1: One class
Simple hypothesis
Poor goodness of fit

Hypothesis 2: Three classes
Complex hypothesis
Better goodness of fit

- Intuitively, which one is more acceptable ?

Example

Dense data



Feature space

Hypothesis 1: One class
Simple hypothesis
Poor goodness of fit

Hypothesis 2: Three classes
Complex hypothesis
Better goodness of fit

- Intuitively, which one is more acceptable ?

Bayesian approach to taxonomy learning

- To minimize: $c(h_i \mid d) = c(h_i) + c(d \mid h_i)$
- A hypothesis h_i is characterized by
  - N_i: Number of classes proposed in the hypothesis
  - where each class is characterized by (n_i, A_i)
    - n_i: number of data points
    - A_i: area (tightest fit to all data points)
- Complexity of hypothesis (prior): $c(h_i) = k_1.N_i$

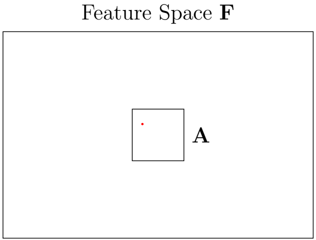
Bayesian approach to taxonomy learning

... cont'd.

Probability of a random point to fall in $\mathbf{A} = \frac{A}{F}$
Probability of n random point to fall in $\mathbf{A} = (\frac{A}{F})^n$

Goodness of fit: $GF \propto 1/(A^n)$

Data complexity: $\mathbf{log}\frac{1}{GF} = k + n.\mathbf{log}A$



- There are N_i classes in hypothesis h_i, each characterized by (n_i, A_i)
  - Complexity of evidence: $c(d \mid h_i) = k_2 + \sum_{i=1}^{N_i} n_i.logA_i$

Bayesian approach to taxonomy learning

... cont'd.

- Complexity of the posterior belief for hypothesis h_i:
  - Complexity of prior + Complexity of evidence
  - $c(h_i \mid d) = k_1.N + k_2 + \sum_{i=1}^{N} n_i.logA_i$

- Inference: $h^* = \text{argmin}_i c(h_i \mid d) = \text{argmin}_i (k_1.N + k_2 + \sum_{i=1}^{N} n_i.logA_i)$

Let us work out one numerical example for better understanding (quiz)

You can refer to this slide while attending the quiz

Quiz

Quiz 03-06

End of Module 03-06