Pattern Recognition. Overview
Pooran Singh Negi
Generative model for discrete data
Concept learning
How to emulate this behavior in a machine

Pattern Recognition. Overview

Pooran Singh Negi

University Of Denver

September 19, 2017

Outline

1 Generative model for discrete data
- Concept learning
- How to emulate this behavior in a machine

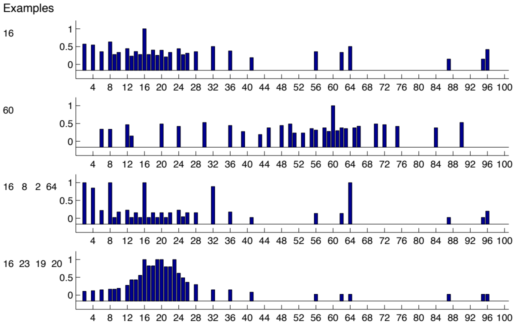
Bayesian concept learning

- Typical machine learning model require both positive and negative example
- Human learning abilities(only positive examples, small sample)

Example of concept learning. Number gameTenenbaum

- Teacher Choose some arithmetic concept C e.g. **prime number, multiple of 4**
- Student is given some positive example drawn at random from C in the range $[1 \cdots 100]$.
- Student is asked whether new example $x_{test} \in C$.

Empirical predictive distribution



Empirical predictive distribution averaged over 8 humans. First, second. third and fourth row **posterior predictive distribution** $p(x|\mathcal{D})$ **after seeing** $\mathcal{D} = \{16\}$ **,** $\{60\}$**,** $\{16, 8, 2, 64\}$ **and** $\{16, 23, 19, 20\}$**.**

Hypothesis Space

Consider possible hypothesis $\mathcal{H}$

- Even number
- odd number
- power of 2
- power of 2, plus 37
- power of 2 except 32

The subset of $\mathcal{H}$ consistent with data is called version space. After seeing some example, some rule $h \in \mathcal{H}$ is consistent with samples. How to combine them to predict $x_{test} \in C$.

Bayesian explanation

- Likelihood function. Size principle.
  Let assume example of sampled uniformly form the extension of concept. Then probability of sampling N examples from h is

$$p(h|\mathcal{D}) = \left[\frac{1}{size(h)}\right]^N$$

  Model favors simplest(smallest) hypothesis consistent with the data(**Occam's razer**)
- Prior. Intuition

Bayesian explanation

- Prior.
  - Intuition.
  - Helps in bringing background knowledge to the problem.
  - Also helps in learning from small sample size.

Let $\mathcal{D} = \{16, 8, 2, 64\}$. Given this data $p(\mathcal{D}|h'^{=poweroftwoexcept32})$ is more likely than $p(\mathcal{D}|h'^{=powerof2})$, although unnatural.

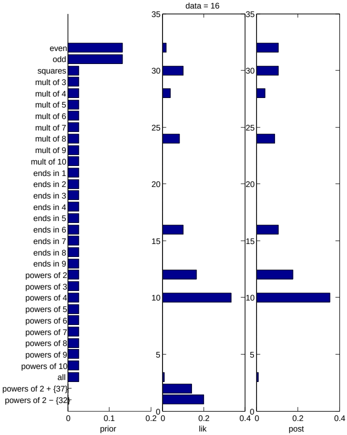
Which prior we should choose?

Bayesian explanation

Posterior

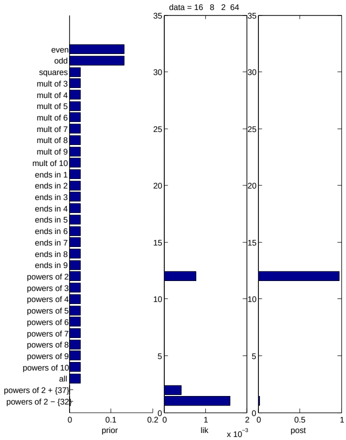
$$p(h|\mathcal{D}) = \frac{p(\mathcal{D}|h)p(h)}{\sum_{h' \in \mathcal{H} p(\mathcal{D}|h')p(h')}}$$

Bayesian explanation



Prior, likelihood and posterior for $\mathcal{D} = \{16\}$

Bayseian explanation



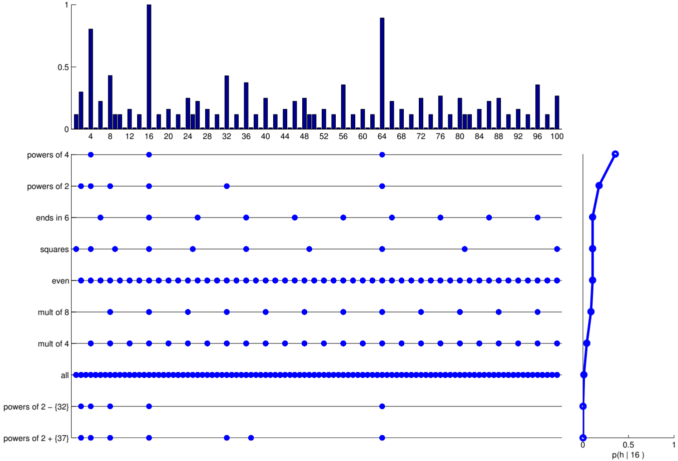
Prior, likelihood and posterior for $\mathcal{D} = \{16, 8, 2, 64\}$

When we have enough data posterior become peaked on single concept, called MAP (Maximum a posteriori estimatie) estimate.

$$\begin{aligned}
\hat(h)^{MAP} &= \text{argmax}_h p(h|\mathcal{D}) \\
&= \text{argmax}_h p((D)|h)p(h) \\
&= \text{argmax}_h [\log p(\mathcal{D}|h) + \log p(h)]
\end{aligned}$$

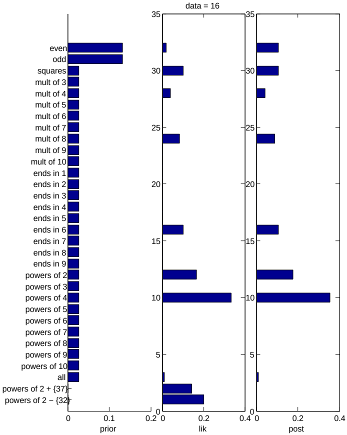
Posterior predictive distribution

Bayes model averaging $p(x_{new} \in C|\mathcal{D}) = \sum_h p(y = 1|x_{new})p(h|\mathcal{D})$



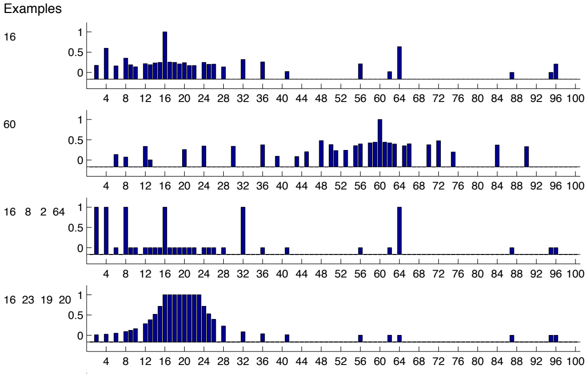
Posterior over hypothesis and corresponding predictive distribution after seeing $\mathcal{D} = \{16\}$

Bayesian explanation



Prior, likelihood and posterior for $\mathcal{D} = \{16\}$

Bayesian explanation



Predictive distribution for the model using full hypothesis space. Compare this to human experiment at the begining.

Thank you!