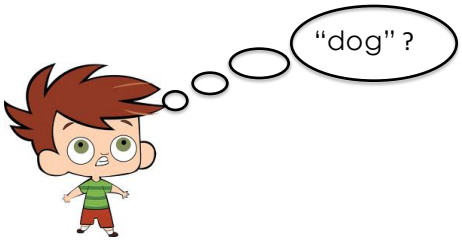
Bayesian Inference

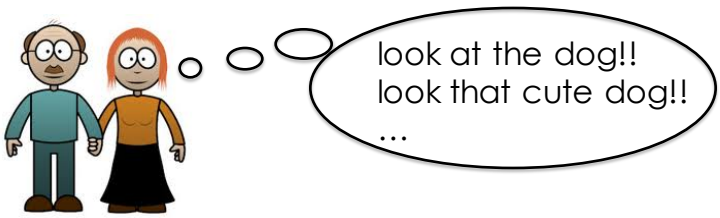
Karim Pichara
Computer Science Department
Pontificia Universidad Católica de Chile

Bayesian Concept Learning

How a child learns to understand the meaning of a word?

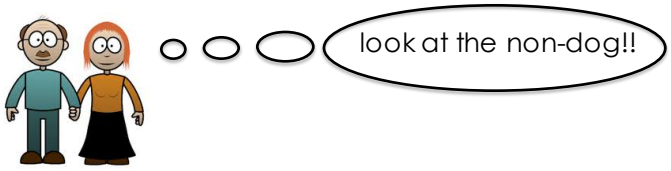


Parents teach pointing our positive examples:

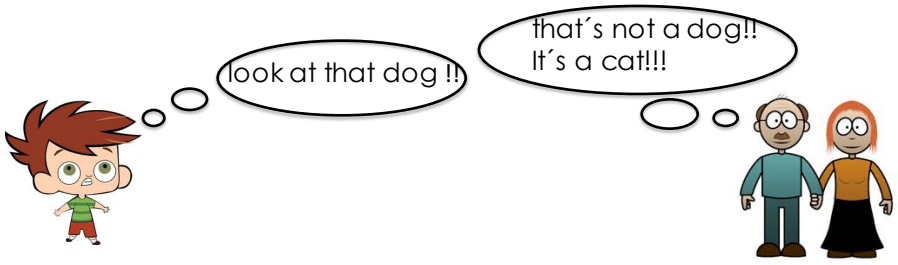


Bayesian Concept Learning

Parents do not provide negative examples:



Negative examples may be obtained during the learning process by Active Learning:



Bayesian Concept Learning

- Xu and Tenenbaum (2007) show that people can learn concepts from positive examples alone
- Consider the function f(x) = 1 if the example x belongs to a concept C and f(x) = 0 otherwise
- We aim to learn f only from positive examples (is not a regular binary classification problem)

The number game

- Suppose we need to guess the arithmetical concept (such as "a number between 1 and 10","prime numbers", etc.)
- Someone tell us a the set of numbers that belong to the concept C and we have to determine C
- Assume that all numbers are integers from 1 to 100
- Someone tell us that the number "16" belongs to C

17 ?, 6 ? , 32 ? may be

99 ? I don´t think so

Some numbers are more likely than others

The number game

17 ?, 6 ? , 32 ? may be

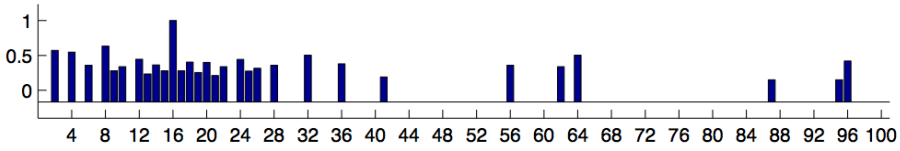
99 ? I don´t think so

Let´s represent these with probabilities ☺

$P(x \in C|D)$ probability that a number "x" belongs to C given data. Recall that C is unknown

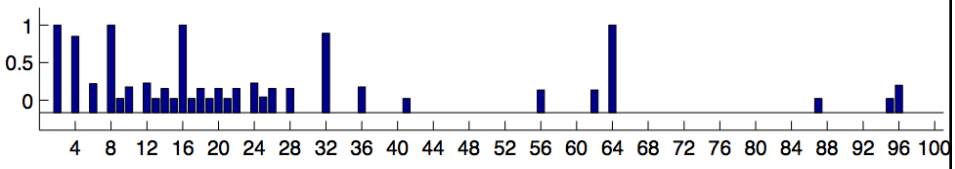
Posterior Predictive Distribution

$P(x \in C|16)$ (from a group of people)

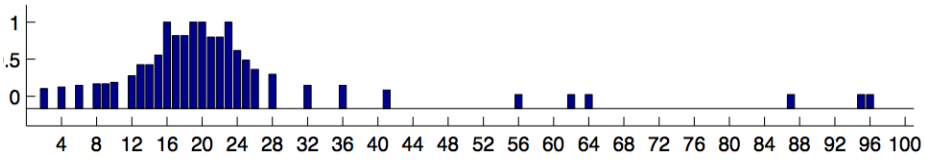


The number game

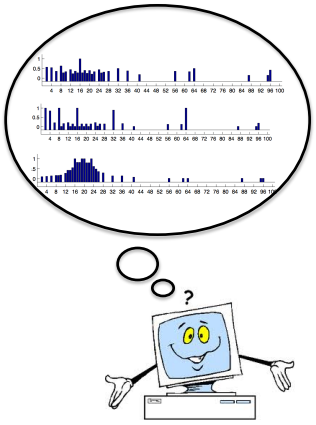
$P(x \in C|16, 8, 2, 64)$



$P(x \in C|16, 23, 19, 20)$



How we emulate this behavior in a machine?



Likelihood

- Suppose again the data D = { 16, 8 , 2 , 64 }
- Consider the two concepts: "powers of two" and "even numbers"
- What will you choose?
- Why (both concepts match the examples)? The key intuition is that we want to avoid suspicious coincidences If the true concept was "even numbers", how come we only saw numbers that happened to be powers of two?
- The likelihood is the probability of the data given the model:
- P(16,8,2,64 | "powers of two") = $(1/6)^4$
- P(16,8,2,64 | "even numbers") = $(1/50)^4$
- $(1/6)^4 > (1/50)^4$

Priors

- Suppose again the data D = {16,8,2,64}
- The concept "powers of two except 32" is more likely than "powers of two" following the likelihood
- What we choose?

Priors

- Our intuition says that "powers of two except 32" seems to be "unnatural"
- We can capture such intuitions by assigning low prior probabilities to unnatural concepts
- Priors are subjective ☹
- We can include background information to make the prior more objective ☺

Posterior

- The posterior is the likelihood times the prior (normalized)

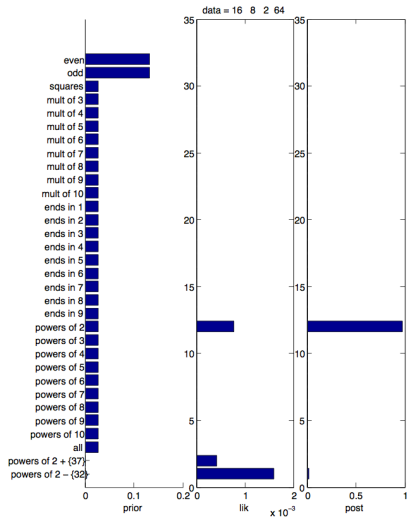
likelihood

Prior

$$P(h|D) = \frac{P(D|h)P(h)}{\sum_h P(D|h)P(h)}$$

hypothesis consistent with data

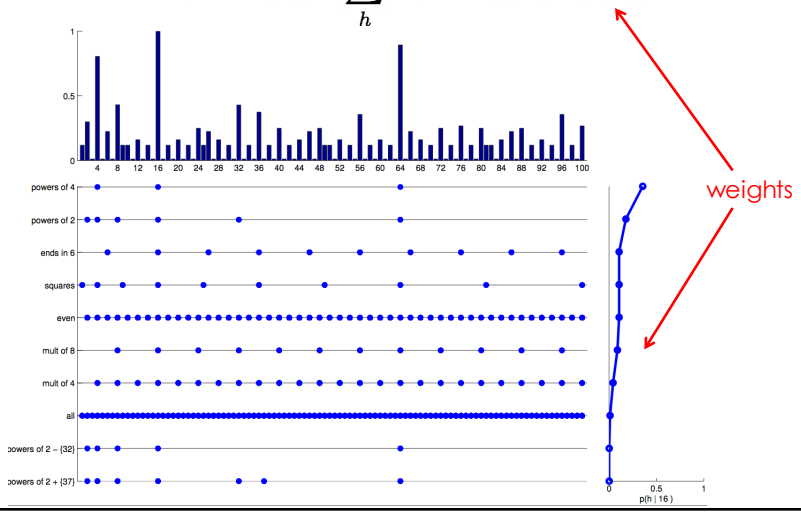
Version space: Set the hypothesis consistent with data, in this example, the possible concepts consistent with 16,8,2,64.



Bayesian Model Averaging: Posterior Predictive

To predict the probability of a number under a concept we should get a weighted average among all the consistent hypothesis:

$$P(x \in C|D) = \sum_h P(x \in h|x,h)P(h|D)$$



Some Intuitions

- We are trying to take decisions under uncertainty
- With Bayesian Inference we are able to manage the uncertainty by using a distribution over the possible models, without directly using any of them
- Instead of estimating the model and then make decisions, we pass through all the possible models using the distribution over them (posterior predictive)
- This is good because:
  - We obtain a distribution as a final decision
  - Unlikely models are discarded automatically

The beta binomial model

θ ? (Probability of heads?)

We need data (many flips)

Likelihood

We usually expect that the probability of heads is close to 0.5

Prior

We need distributions!!

The beta binomial model

- Binomial distribution: Probability of getting N_1 successes (lets say heads) in N trials

$$N_1 \sim Bin(N, \theta) \qquad P(N_1 = k) = \binom{N}{k} \theta^k (1-\theta)^{N-k}$$

  - Here the random variable is the number of heads

- Bernoulli distribution: Probability of getting a success in one trial

$$X_i \sim Ber(\theta) \qquad P(X_i|\theta) = \theta^{X_i}(1-\theta)^{1-X_i}$$

- Beta distribution:

Distribution over a distribution!

$$Beta(\theta|\alpha_1, \alpha_2) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \theta^{\alpha_1 - 1}(1-\theta)^{\alpha_2 - 1}$$

The beta binomial model

- Let $X_i \sim Ber(\theta)$ where $X_i = 1$ represent "heads" and $X_i = 0$ represent "tails". $P(X_i|\theta) = \theta^{X_i}(1-\theta)^{1-X_i}$
- $\theta \in [0,1]$ is the probability of "heads"
- Likelihood: $P(D|\theta) = \theta^{N_1}(1-\theta)^{N_0}$

number of heads in N trials: $N_1 = \sum_{i=1}^{N} \mathbb{I}(x_i = 1)$

number of tails in N trials: $N_0 = \sum_{i=1}^{N} \mathbb{I}(x_i = 0)$

$N = N_0 + N_1$

suficient statistics

$s(D)$ such that $P(\theta|D) = P(\theta|s(D))$

The beta binomial model: Priors

$N_1 \sim Bin(N, \theta)$ $\quad N_1 = \sum_{i=1}^{N} \mathbb{I}(x_i = 1)$

Probability to get N1 heads in N Bernoulli trials

$P(N_1 = k) = \binom{N}{k} \theta^k (1-\theta)^{N-k}$

We need a prior which has support on [0,1]. Beta distribution is a convenient prior:

$$P(\theta) \propto \theta^{\alpha_1 - 1}(1-\theta)^{\alpha_2 - 1}$$

The posterior (likelihood times the priors) is friendly (also Beta)

$$P(\theta|D) \propto P(D|\theta)P(\theta) = \theta^{N_1}(1-\theta)^{N_0}\theta^{\alpha_1-1}(1-\theta)^{\alpha_2-1} = \theta^{N_1+\alpha_1-1}(1-\theta)^{N_0+\alpha_2-1}$$

conjugate prior

The beta binomial model: Priors

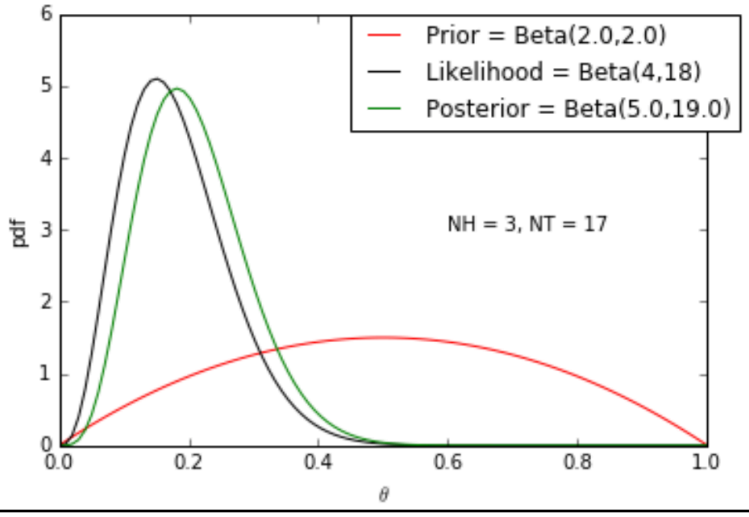
- Beta is the conjugate prior of the Bernoulli

$$Beta(\theta|\alpha_1, \alpha_2) \propto \theta^{\alpha_1 - 1}(1-\theta)^{\alpha_2 - 1}$$

hyperparameters

Example

Updating a Beta(2, 2) prior with a Binomial likelihood with sufficient statistics $N_H = 3$, $N_T = 17$ to yield a Beta(5,19) posterior.



Example

Updating a Beta(5, 2) prior with a Binomial likelihood with sufficient statistics $N_H = 11$, $N_T = 13$ to yield a Beta(16,15) posterior.