The following slides are borrowed from UBC CS340.
Thanks Josh!

https://www.cs.ubc.ca/~murphyk/Teaching/CS340-Fall07/josh1.pdf
https://www.cs.ubc.ca/~murphyk/Teaching/CS340-Fall07/josh2.pdf

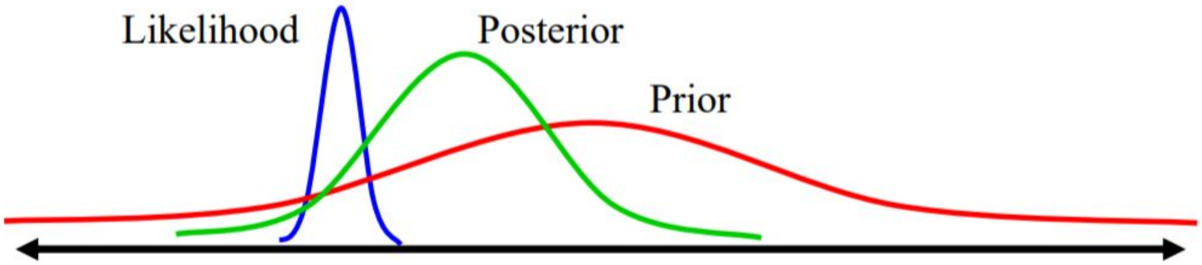
Bayesian belief updating

Posterior probability

Likelihood

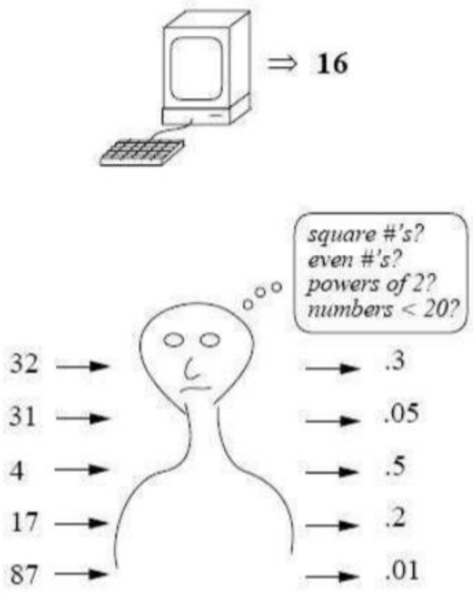
Prior probability

$$p(h \mid d) = \frac{p(d \mid h)\, p(h)}{\sum_{h' \in H} p(d \mid h')\, p(h')}$$

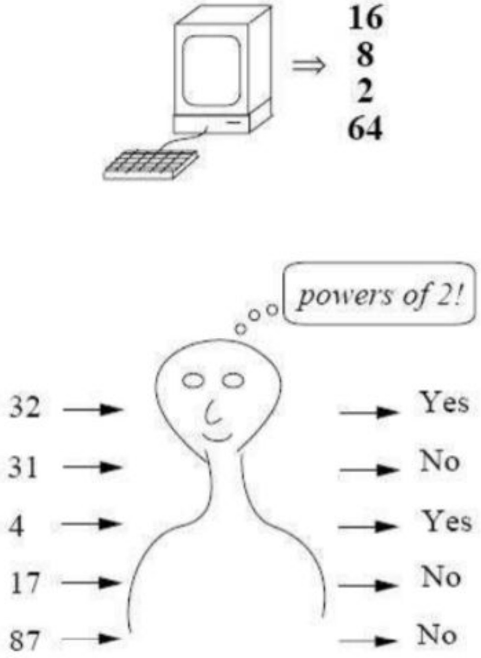


Number Game

1 random "yes" example:



4 random "yes" examples:



Bayesian model

- H: Hypothesis space of possible concepts:
- $X = \{x_1, \ldots, x_n\}$: n examples of a concept C.
- Evaluate hypotheses given data using Bayes' rule:

$$p(h \mid X) = \frac{p(X \mid h)\, p(h)}{\sum_{h' \in H} p(X \mid h')\, p(h')}$$

  - $p(h)$ ["prior"]: domain knowledge, pre-existing biases
  - $p(X|h)$ ["likelihood"]: statistical information in examples.
  - $p(h|X)$ ["posterior"]: degree of belief that h is the true extension of C.

How to Calculate??

Problem 1: Likelihood

- **Size principle**: Smaller hypotheses receive greater likelihood, and exponentially more so as *n* increases.

$$p(X \mid h) = \left[\frac{1}{\text{size}(h)}\right]^n \text{ if } x_1, \ldots, x_n \in h$$

$$= 0 \text{ if any } x_i \notin h$$

Kind of like a joint probability…
What if we don’t have a X, that is, no likelihood (P(X|h) = 1)...

- $X=\{20,40,60\}$
- $H1$ = multiples of 10 = $\{10,20,\ldots,100\}$
- $H2$ = even numbers = $\{2,4,\ldots,100\}$
- $H3$ = odd numbers = $\{1,3,\ldots,99\}$
- $P(X|H1) = 1/10 * 1/10 * 1/10$
- $p(X|H2) = 1/50 * 1/50 * 1/50$
- $P(X|H3) = 0$

Coding Sample...

```
def likelihood(x, h):
    '''
    x: a data point or a list of data points
    h: a hypothesis, the data type could be anything(int, string...)

    '''
    if h == 1 or h == 2: #even number or odd number
        prob = (1/50)**(len(x))
    elif h == 3: #square number (1,4,9,16,25,36,49,64,81,100)
        prob = (1/10)**(len(x))
    else:
        prob = 0 #for data points not in it...

    return prob
```

Answering question on Piazza, sum = 1?

- Since $p(\vec{x}|h)$ is a distribution over vectors of length n, we require that, for all h, $\sum_{\vec{x}} p(x|h) = 1$
- It is easy to see this is true,
e.g., for h=even numbers, n=2

$$\sum_{x_1=1}^{100}\sum_{x_2=1}^{100} p(x_1, x_2|h) = \sum_{x_1=1}^{100}\sum_{x_2=1}^{100} p(x_1|h)p(x_2|h) = \sum_{x_1 \in even}\sum_{x_2 \in even} \frac{1}{50}\frac{1}{50} = 1$$

- If x is fixed, we do not require $\sum p(X|h) = 1$

Problem 2: Posterior

$$p(h \mid X) = \frac{p(X \mid h)\, p(h)}{\sum_{h' \in H} p(X \mid h')\, p(h')}$$

Equal Prior P(h) = 1/6

Sudo Code...

```
def posterior(x,p(x|H)):
    '''
    x: a data point or a list of data points
    p(x|H): a list of probabilities containing p(x|H1), p(x|H2)...

    '''
    p(h|x)[0] = (p(x|H)[0]*1/6) / (sum(p(x|H)*1/6))
    p(h|x)[1] = (p(x|H)[1]*1/6) / (sum(p(x|H)*1/6))
    p(h|x)[2] = (p(x|H)[2]*1/6) / (sum(p(x|H)*1/6))

    #e.t.c...
    #return a list
    return p(h|x)
```

Generalizing to new objects...

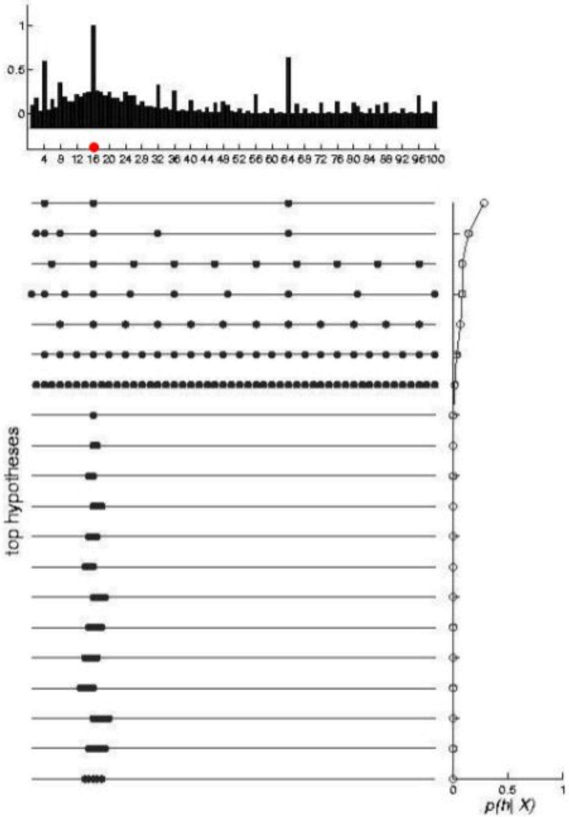
Compute the probability that C applies to some new object y by averaging the predictions of all hypotheses h, weighted by $p(h|X)$ (**Bayesian model averaging**):

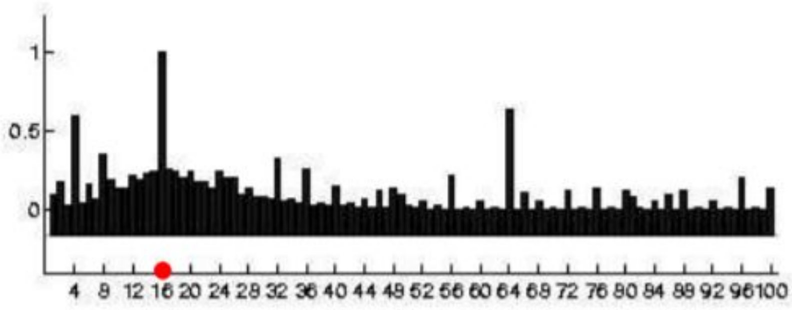
$$p(y \in C \mid X) = \sum_{h \in H} \underbrace{p(y \in C \mid h)}_{=\begin{cases} 1 \text{ if } y \in h \\ 0 \text{ if } y \notin h \end{cases}} \; p(h \mid X)$$

$$= \sum_{h \supset \{y, X\}} p(h \mid X)$$

Result looks like this...

Examples:
16

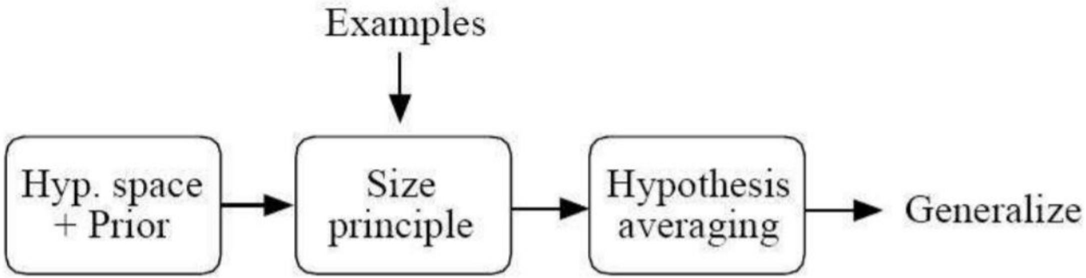




No Data? No Likelihood.

P(1 belongs to the concept) = P(odd)
P(16 belongs to the concept) = P(even) + P(square) + P(power of 2)

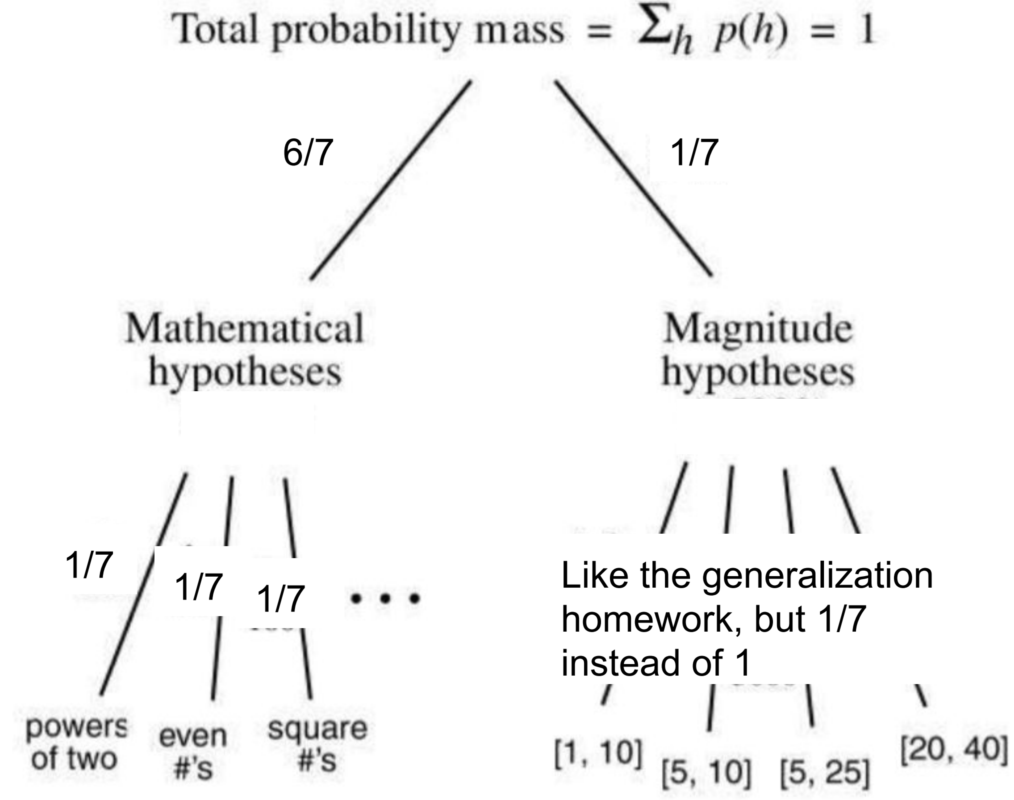
Summary of the Bayesian approach



1. Constrained hypothesis space H
2. Prior p(h)
3. Likelihood p(X|h)
4. Hypothesis (model) averaging:

$$p(y \in C\ X) = \sum_{h} p(y \in C|h)p(h|X)$$

Problem 3



Like the generalization homework, but 1/7 instead of 1

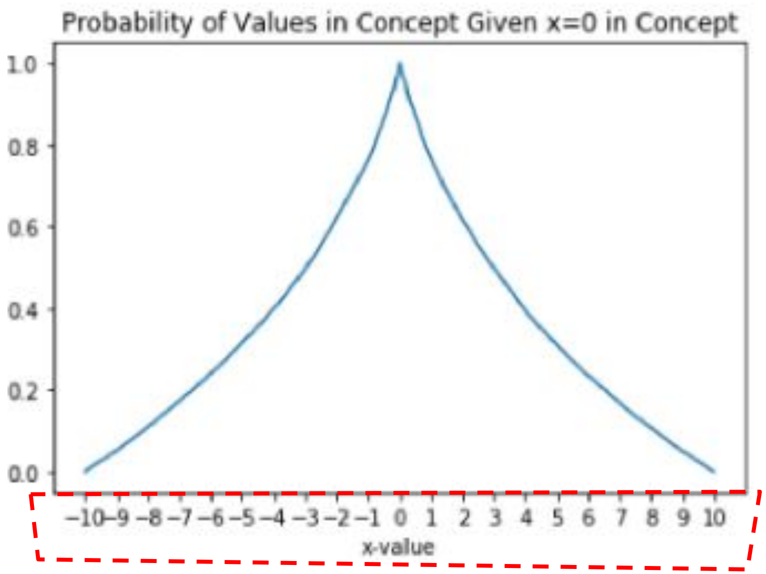
- Region is…(question on Piazza)

[1], [1,2], [1,2,3], ...[1,...100], [2], [2,3], ... [2,100], [3], [3,4], ...
[3,...100]... [100]

- How to create such region…

Two FOR loops for two boundaries like you did in that generalization problem. Anything occurs to you?

Prior = 1/7 * 1/(how many regions? 5050.)



Add this probability to all other 6 probability...

How many regions if 0 to 100?