Bayesian Statistics

With Applications to Computer Vision

Amy Fire

BMC Intermediate II Spring 2024

https://www.smbc-comics.com/comic/bayesian

Outline

- Part I: Bayesian Statistics
- Part II: Applications to Computer Vision

DISCLAIMER:

We’re going to talk in an incredibly simplified way to (hopefully) develop some intuition.

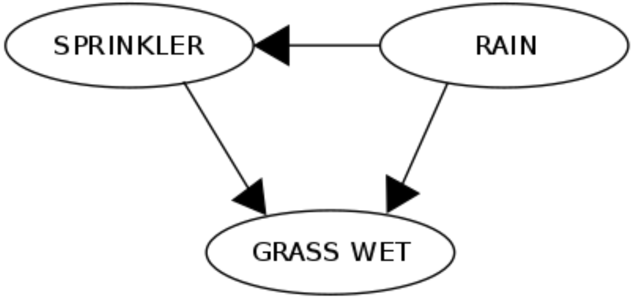
Part I: Bayesian Statistics

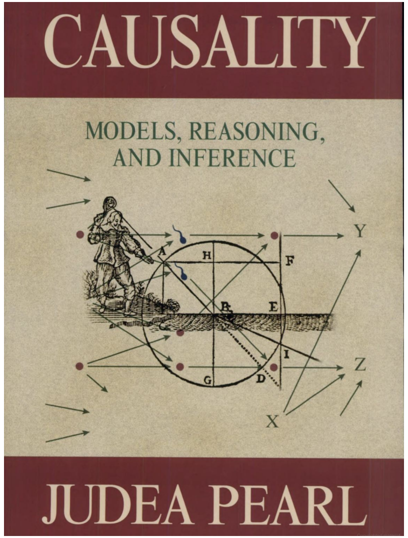
(sort of)

Pierre-Simon Laplace
1749 – 1827

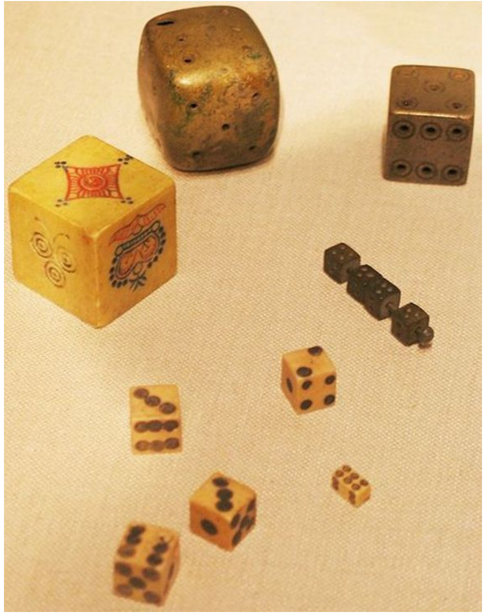
Motivation

- Chance of Rain: P(rain)
- P(rain) vs P(rain | grass wet)
- How does the knowledge of the grass being wet affect belief of other nodes?
- We want a way to update our prior model for P(rain) to incorporate the new data that it's wet outside to get a new posterior model P(rain | grass wet)





First:
The Quick Introduction to Probability



- Events
  - A **random variable** assumes any of several different numeric values as a result of some random event.
  - Can be **discrete** or **continuous**

- A **probability model** is the collection of all possible outcomes (the sample space) and their corresponding probabilities.

- **(Simplified) Probability Axioms**

1. $P(A) \geq 0$
2. $P(\Omega) = 1$
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

x				
P(X = x)				

Theoretical Probability

- If outcomes are equally likely,

$$P(A) = \frac{\#\ \text{outcomes in } A}{\#\ \text{possible outcomes}}$$

1. Rolling 20-sided die
   a. P(2)
   b. P(even)
   c. P(1 or 2)
   d. P(21)
   e. P(odd)

2. Tossing two coins
   a. P(no Heads)
   b. P(at least one H)
   c. P(one H, one T)
   d. P(1st is H)
   e. P(TT)

Antigen Covid Tests ~2020

Find:

1. P(antigen positive)
2. P(antigen negative)
3. P(antigen negative AND PCR positive)

	No. of persons (row %)§			
Total no. of persons (column %)	Antigen-positive	Real-time RT-PCR–positive	Real-time RT-PCR–positive, antigen-negative	Real-time RT-PCR–negative, antigen-positive
3,419 (100)	161 (4.7)	299 (8.7)	142 (4.2)	4 (0.1)

https://www.cdc.gov/mmwr/volumes/70/wr/mm7003e3.htm#T1_down

Conditional Probability

$$P(A|B) := P(A \cap B)/P(B)$$

- Find
  1. P(PCR +)
  2. P(PCR + | ag +)
  3. P(PCR + | ag + , Symptomatic)
  4. P(PCR + | ag + , Asymptomatic)

- What do you notice?

Results and Performance	Real-time RT-PCR, no. of tests		
	Positive	Negative	Total
BinaxNOW antigen test result			
All participants (N = 3,419)			
Positive	157	4	**161**
Negative	142	3,116	**3,258**
Total	**299**	**3,120**	**3,419**
Symptomatic (≥1 symptom) (n = 827)			
Positive	113	0	**113**
Negative	63	651	**714**
Total	**176**	**651**	**827**
Asymptomatic (n = 2,592)			
Positive	44	4	**48**
Negative	79	2,465	**2,544**
Total	**123**	**2,469**	**2,592**

https://www.cdc.gov/mmwr/volumes/70/wr/mm7003e3.htm#T2_down

Bayes' Theorem

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}$$

- Let's prove it!

- We write:

$$f(y|x) \sim f(x|y)f(y)$$

[370]

quodque ſolum, certa nitri ſigna præbere, ſed plura concurrere debere, ut de vero nitro producto dubium non relinquatur.

LII. *An Eſſay towards ſolving a Problem in the Doctrine of Chances. By the late Rev. Mr.* Bayes, *F. R. S. communicated by Mr.* Price, *in a Letter to* John Canton, *A. M. F. R. S.*

Dear Sir,

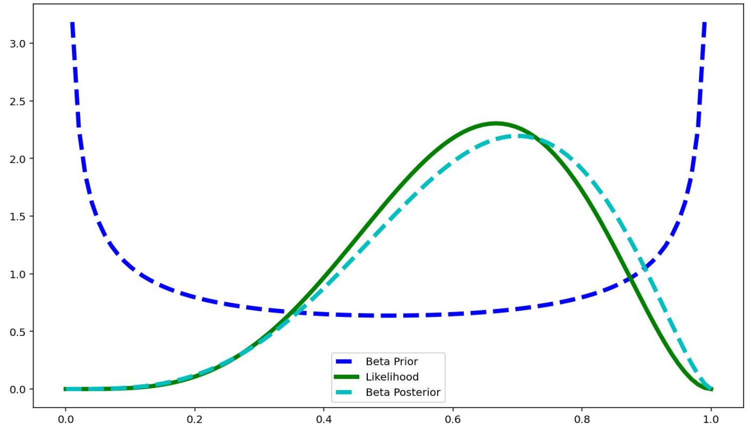
Read Dec. 23, 1763.

I Now ſend you an eſſay which I have found among the papers of our deceaſed friend Mr. Bayes, and which, in my opinion, has great merit, and well deſerves to be preſerved. Experimental philoſophy, you will find, is nearly intereſted in the ſubject of it; and on this account there

Bayes' Theorem – Names for the Different Parts

$$P(H|D) \sim P(D|H) * P(H)$$

Posterior Likelihood Prior



- Q1: What would be the effect of a "flat" prior?
    - MLE (maximum likelihood estimate) vs MAP (maximum a posteriori)
- Q2: What about a prior with point mass?

Coin Flipping Experiments

- Let’s consider coin flipping. Can we determine if a coin is fair?
- What is the probability model for a single coin flip?
  - Bernoulli
- What is the model for the likelihood: N flips with x heads?
  - Binomial
- So… What could work for a prior?

Conjugate Prior to Binomial: Beta Distribution

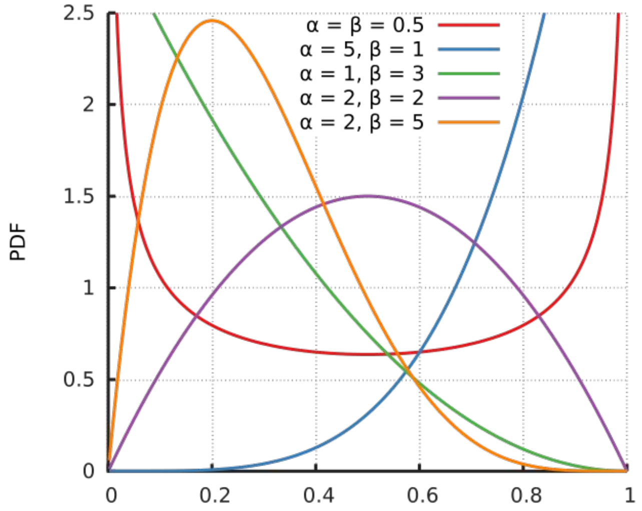
- A **<u>conjugate prior</u>** for a likelihood produces a posterior from the same family.
- The conjugate prior for the Binomial(N,X) or Bernoulli(theta) is the Beta distribution.

- **<u>The Beta(a,b) distribution</u>**

$$Beta(a,b) \sim c\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$E(\theta) = \frac{\alpha}{\alpha+\beta}$$

- Specify the posterior for Bernoulli!

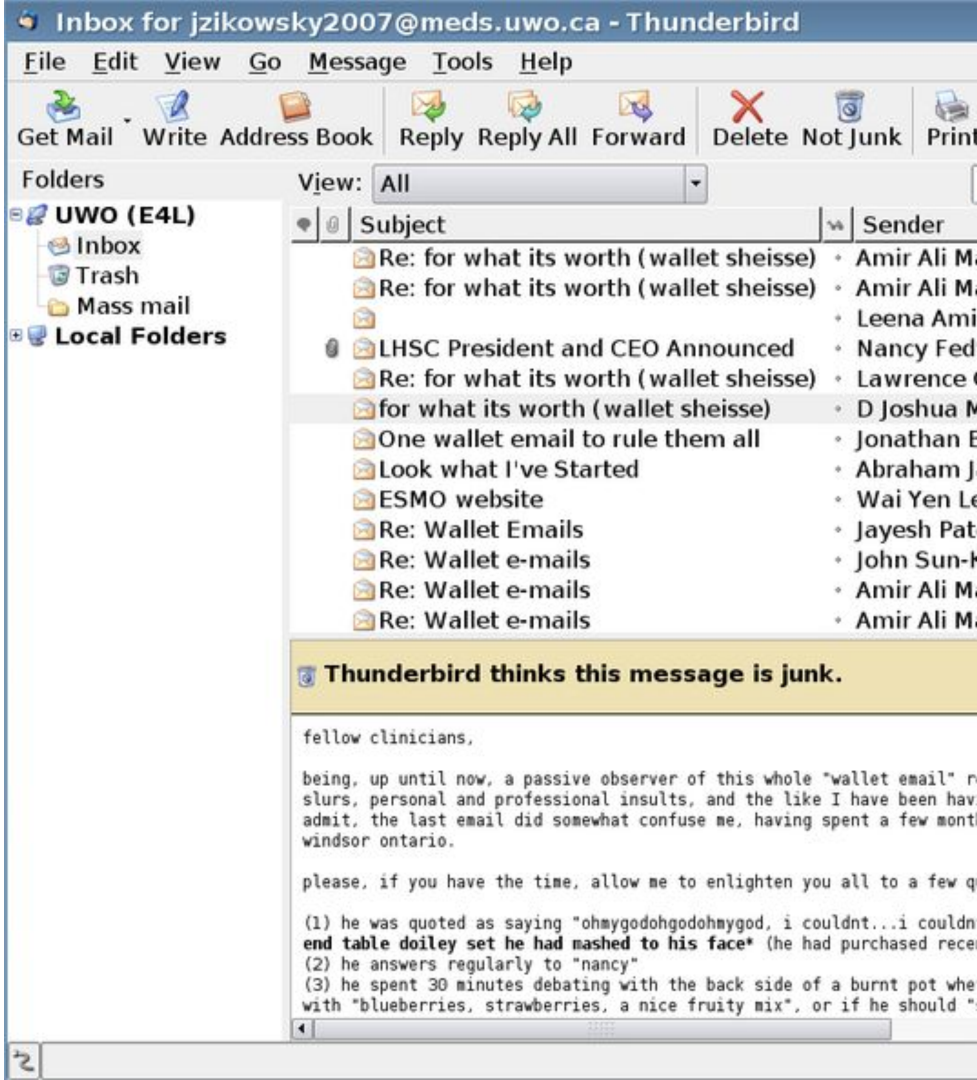


Back to Coin Flipping

- Let’s pick a prior and flip a coin one time.
- And again
- And again.
- Let’s add on 20 more flips!
- What happens?
- What do you notice about the more flips?

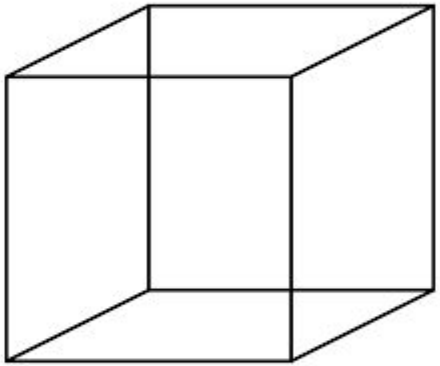
Classifying Spam: A Naive Bayes Classifier

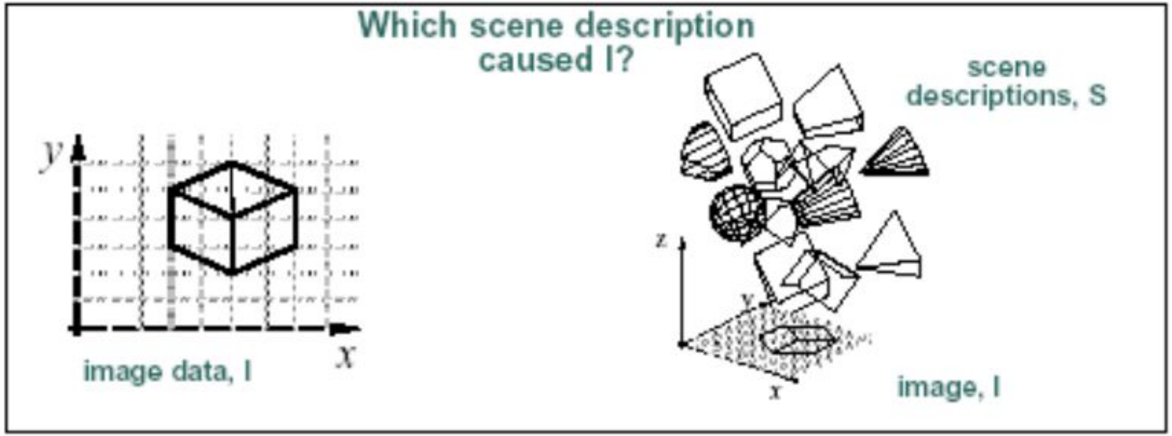
- Emails are Spam or Not Spam
- i-th word of a given document occurs in spam: $P(w_i|S)$
  - Assume independent
- $P(D|S) = \prod_i P(w_i|S)$
- What could we use for P(S)?
- What is P(S | D)?
- Classify Spam: P(S | D) and P(not S | D)



Part II: Computer Vision

What Do You See?

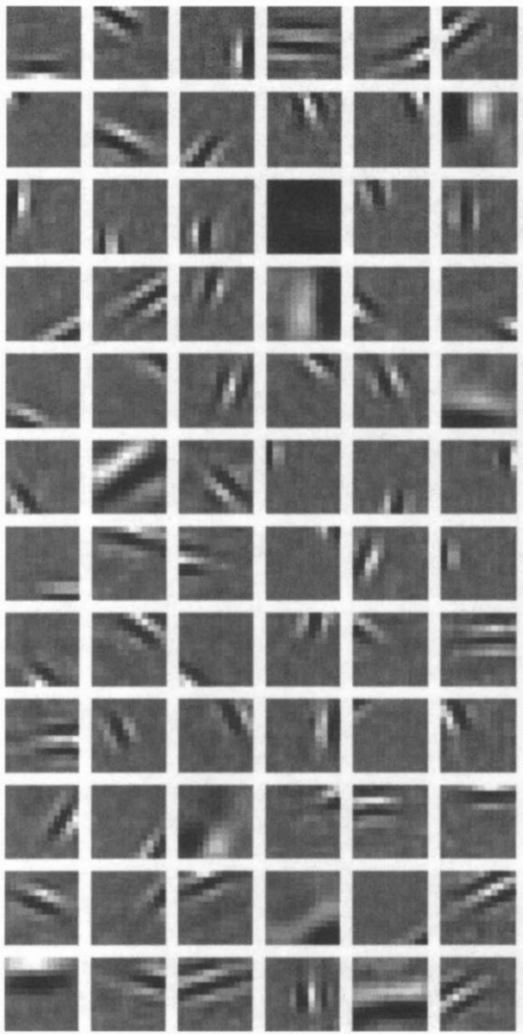




https://www.cs.jhu.edu/~ayuille/JHUcourses/VisionAsBayesianInference2023/1/Lecture1.pdf

Representing Images

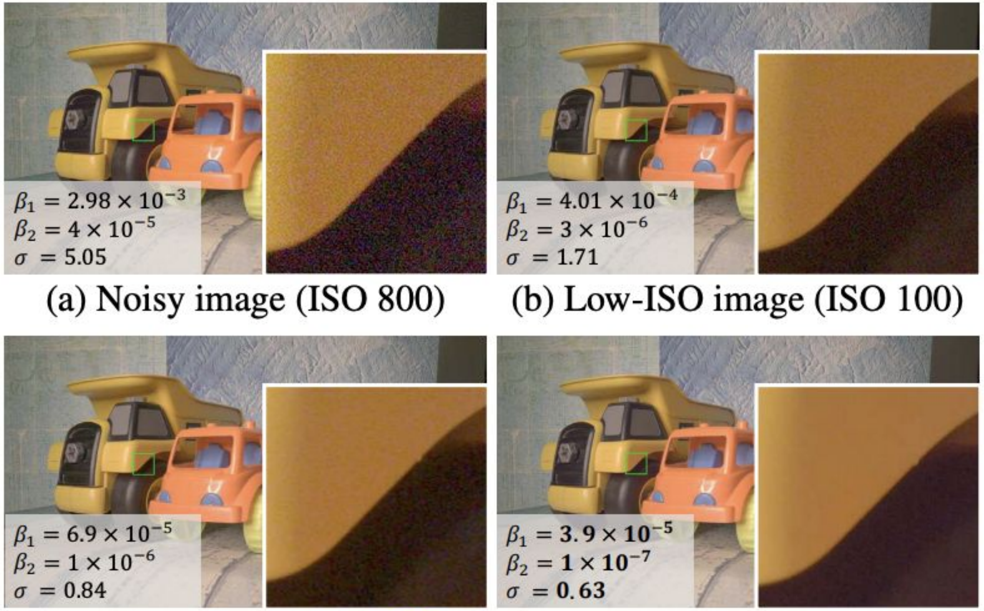
- Images are complex and ambiguous
- Single 1024x1024 image 0-255 photons
  - # Possible images = (1024x1024)^256
- But: # Naturally occurring images is much much less
  - There are regularities in images and the world
- Given Image I, State of the World W
- Discriminative models: P(W | I)
- Generative models: P(I | W) P(W)
- Probability Distribution defined on a structured representation
  - Use prior and likelihood
  - Large question: how do you compute?



Olshausen, Field. Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1 ? 1997

Image Denoising

- Place a prior distribution over clean images
  - Represents our beliefs about what clean images look like.
- Likelihood
  - P(Noisy | Clean) ~ Normal
  - Assumption: noise is additive Gaussian noise
- Use Bayesian inference to estimate the posterior distribution over clean images given the observed noisy image and the prior distribution.



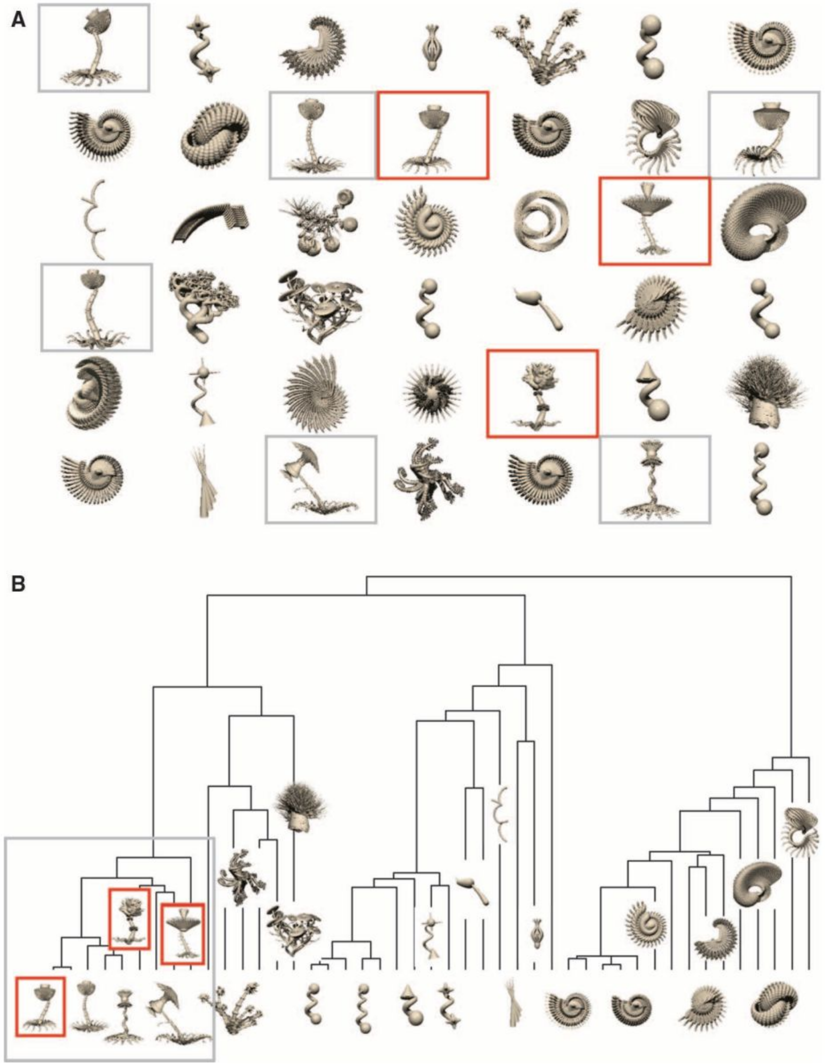
(a) Noisy image (ISO 800) (b) Low-ISO image (ISO 100)

(c) Ground truth using [25] (d) Our ground truth

SIDD (Smartphone Image Denoising Dataset)

Cognitive scientists: Bayesian statistics models human thought

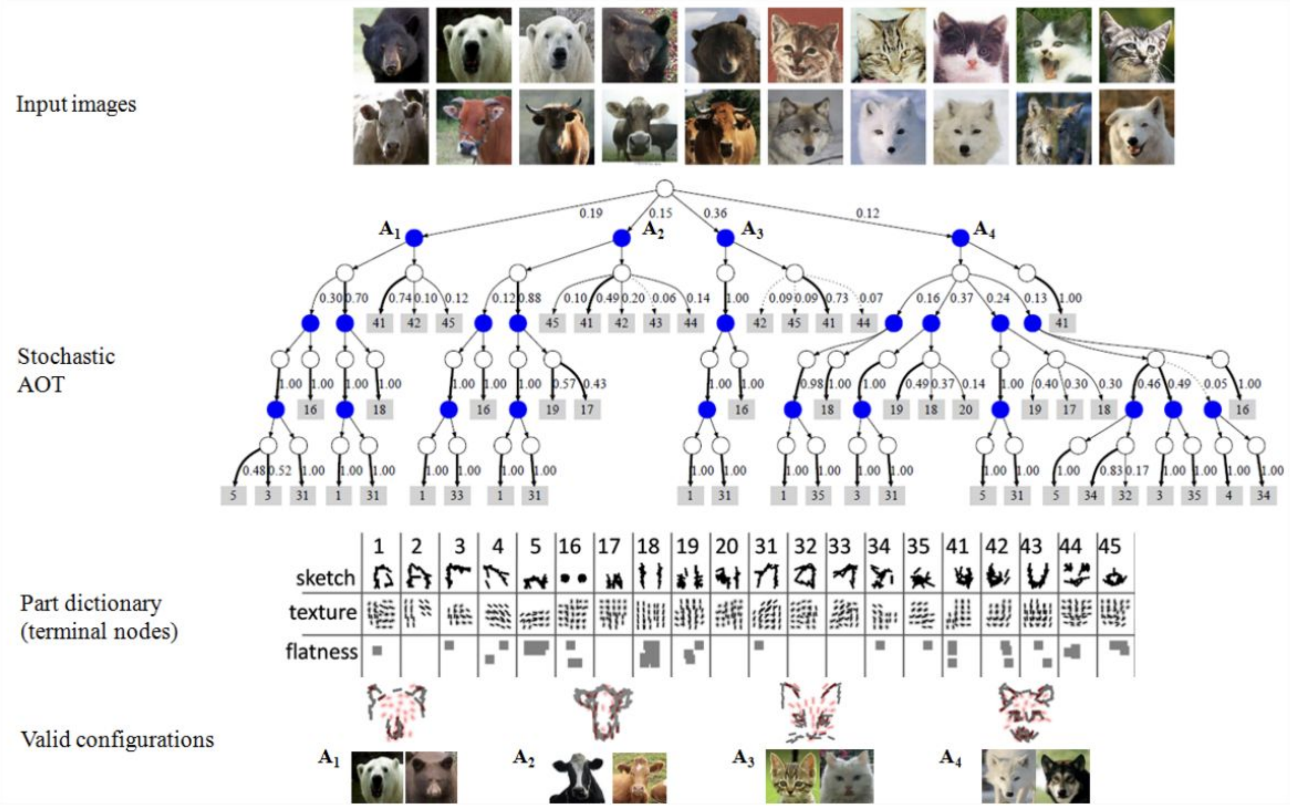
- Learning names for categories can be modeled with Bayesian inference
- Assume a hierarchical tree structure
- Priors are placed on branch length
- Likelihoods favor lower branches
- Posterior probabilities favor generalizing across the lowest branch that spans all observed examples (grey)



How to Grow a Mind: Statistics, Structure, and Abstraction
Joshua B. Tenenbaum, et al. 2011

Learning Concepts from Images

- Prior
  - Structural configuration and
  - Complexity
- Likelihood on features
- Unsupervised learning of hierarchical reconfigurable image templates



Zhangzhang Si and Song-Chun Zhu. Learning AND-OR Templates for ObjectRecognition and Detection.

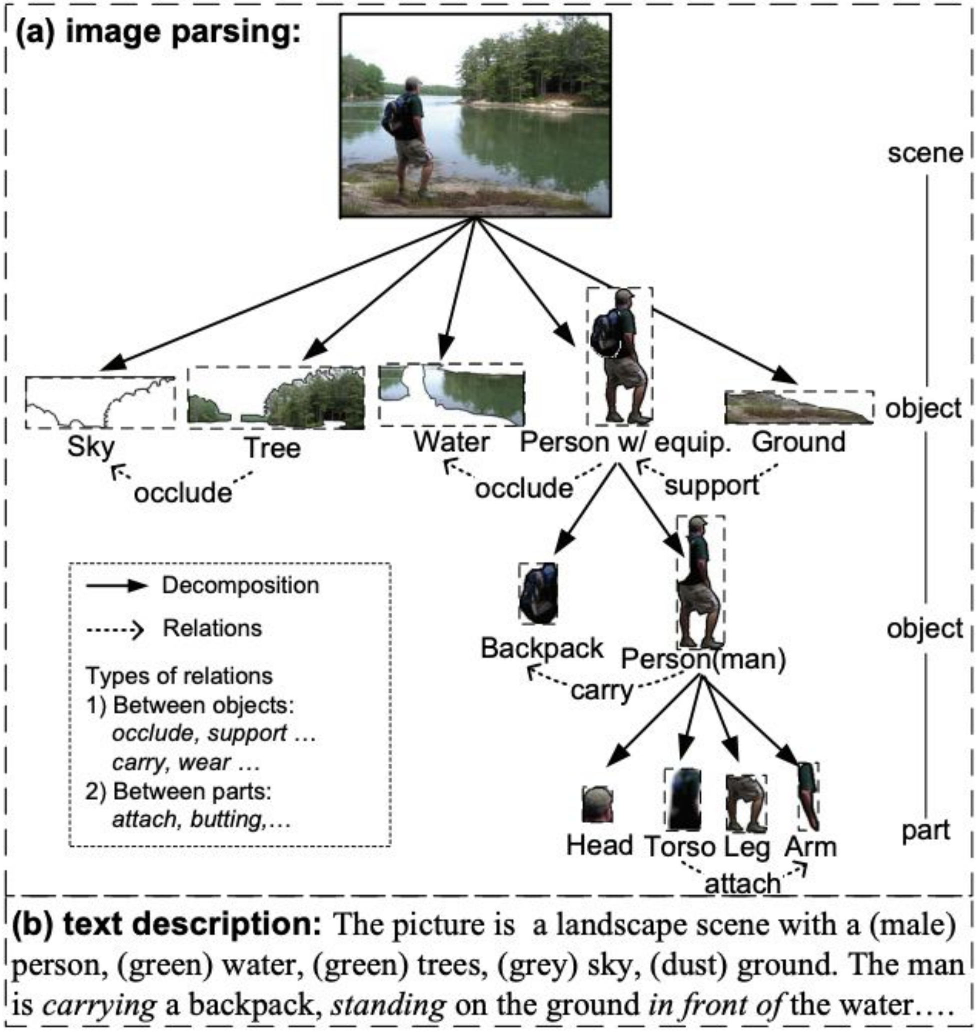
Object Segmentation/Recognition: Microsoft’s COCO Dataset

- The Tasks
  - Segmentation
  - Recognition in context
- What can prior beliefs tell us about the the presence and location of objects in the image (e.g., object sizes, shapes, and typical locations)?
- How can we use Bayesian inference to update our beliefs about the presence and location of objects based on observed image features?

https://cocodataset.org/#explore

Image Parsing

- And-or Graph (AoG) visual knowledge representation, which provides a graphical representation serving as prior knowledge for representing diverse visual patterns and provides top-down hypotheses during the image parsing
- Objects that have a high prior probability of being on together are grouped together with “positive” edges, while objects that have low prior probability of being on together are grouped by negative edges.



Yao, et al. I2T: Image Parsing to Text Description

Pedestrian or not pedestrian

- If you were creating an autonomous vehicle, what type of prior would you place on detecting pedestrians?
- A Note on Error
    - Type I Error: Detect pedestrian when there is none. False Alarm.
    - Type II Error: We fail to detect pedestrian when there is actually a pedestrian. Big Problem.

Washington Post

Tesla Autopilot crashes on cross traffic - The Washington Post

A Post analysis reveals that people have died or been gravely injured in crashes where Tesla's software should not have been enabled in the...

Dec 10, 2023

NBC News

Self-driving car runs over pedestrian hit by SF human driver

A hit-and-run driver struck a pedestrian Monday night, tossing her into the path of a Cruise self-driving car that then drove over her,...

Oct 3, 2023

Futurism

Tesla Driver Says He's Not Sure If He Killed a Pedestrian Because He Was on Autopilot

A Tesla driver, who at first denied having killed a woman with his Tesla in a hit-and-run, is now claiming he must've been using Autopilot.

Feb 12, 2024

InsideEVs

Tesla Crashes Into Tesla, Then Crashes Into Another Tesla

Also, another Tesla was involved but didn't get mixed up in the crash.

Feb 19, 2024

Daily Express

Tesla 'on autopilot' mows down pedestrian before crashing outside migrant centre

Medics were at the scene of the accident outside a migrant center after a Tesla ran over a pedestrian before slamming into multiple cars.

Feb 2, 2024

Overview of Bayesian Statistics

- The Good
  - Small sample inference is the same as large sample
  - It reduces to MLE
  - It makes use of prior information
  - It is interpretable (we didn’t talk much about the frequentist way, but trust me!).
  - It is convenient for a wide variety of models

- The Challenges
  - Choosing the prior – this is really subjective
  - Computational complexity – this is real!