Machine Learning Overview

Sargur N. Srihari
University at Buffalo, State University of New York
USA

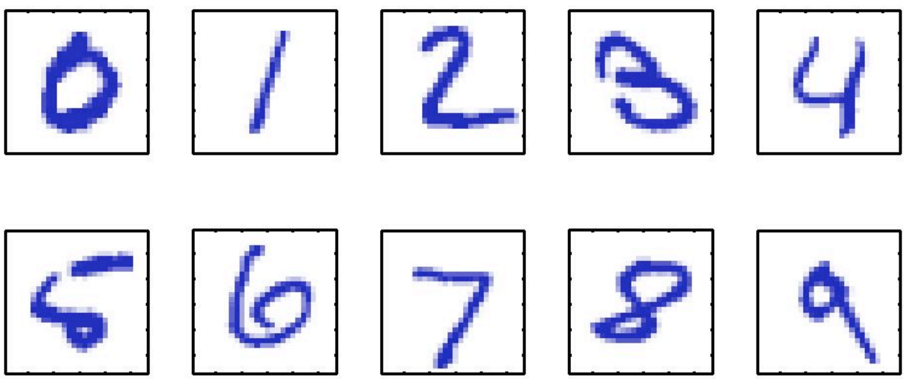
Outline

1. What is Machine Learning (ML)?
2. Types of Information Processing Problems Solved
    1. Regression
    2. Classification
    3. Clustering
    4. Modeling Uncertainty/Inference
3. New Developments
    1. Fully Bayesian Approach
4. Summary

What is Machine Learning?

- Programming computers to:
  - Perform tasks that humans perform well but difficult to specify algorithmically
- Principled way of building high performance information processing systems
  - search engines, information retrieval
  - adaptive user interfaces, personalized assistants (information systems)
  - scientific application (computational science)
  - engineering

Example Problem: Handwritten Digit Recognition



Wide variability of same numeral

- Handcrafted rules will result in large no of rules and exceptions
- Better to have a machine that learns from a large training set

ML History

- ML has origins in Computer Science
- PR has origins in Engineering
- They are different facets of the same field
- Methods around for over 50 years
- Revival of Bayesian methods
  - due to availability of computational methods

Some Successful Applications of Machine Learning

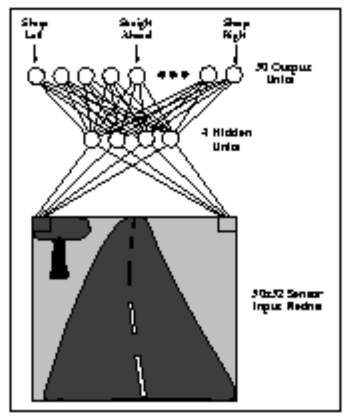
- Learning to recognize spoken words
  - Speaker-specific strategies for recognizing primitive sounds (phonemes) and words from speech signal
  - Neural networks and methods for learning HMMs for customizing to individual speakers, vocabularies and microphone characteristics

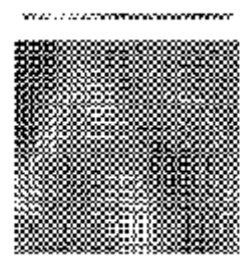
Table 1.1

Some Successful Applications of Machine Learning

- Learning to drive an autonomous vehicle
  - Train computer-controlled vehicles to steer correctly
  - Drive at 70 mph for 90 miles on public highways
  - Associate steering commands with image sequences

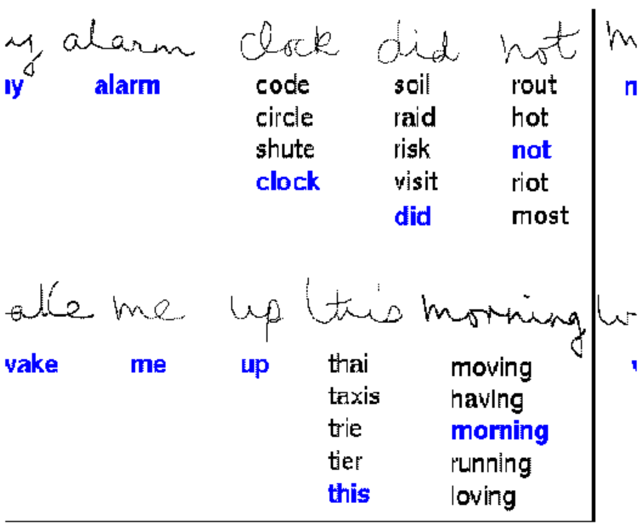
ALVINN [Pomerleau] drives 70 mph on highways





Handwriting Recognition

- **Task *T***
  - recognizing and classifying handwritten words within images
- **Performance measure *P***
  - percent of words correctly classified
- **Training experience *E***
  - a database of handwritten words with given classifications



The ML Approach

- Data Collection
  - Samples
- Model Selection
  - Probability distribution to model process
- Parameter Estimation
  - Values/distributions

- Search
  - Find optimal solution to problem

Generalization

(Training)

Decision

(Inference

OR

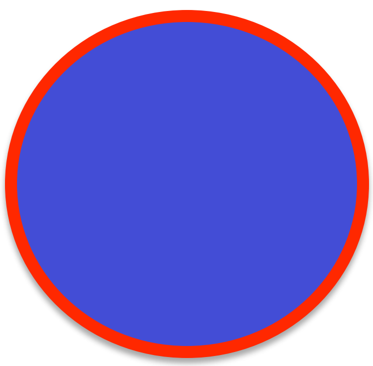
Testing)

Types of Problems machine learning is used for

1. Classification
2. Regression
3. Clustering (Data Mining)
4. Modeling/Inference

Example Classification Problem

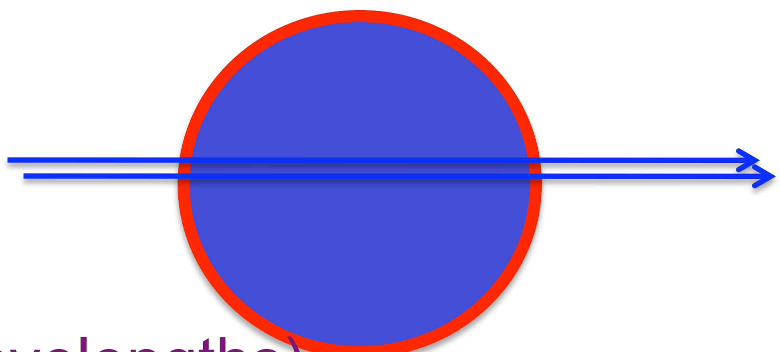
- Off-shore oil transfer pipelines
- Non-invasive measurement of *proportion* of
  - oil,water and gas
- Called Three-phase Oil/Water/Gas Flow



Dual-energy gamma densitometry

- Beam of gamma rays passed through pipe
- Attenuation in intensity provides information on density of material
- Single beam insufficient
  - Two degrees of freedom: fraction of oil, fraction of water

One beam of
Gamma rays
of two energies
(frequencies or wavelengths)

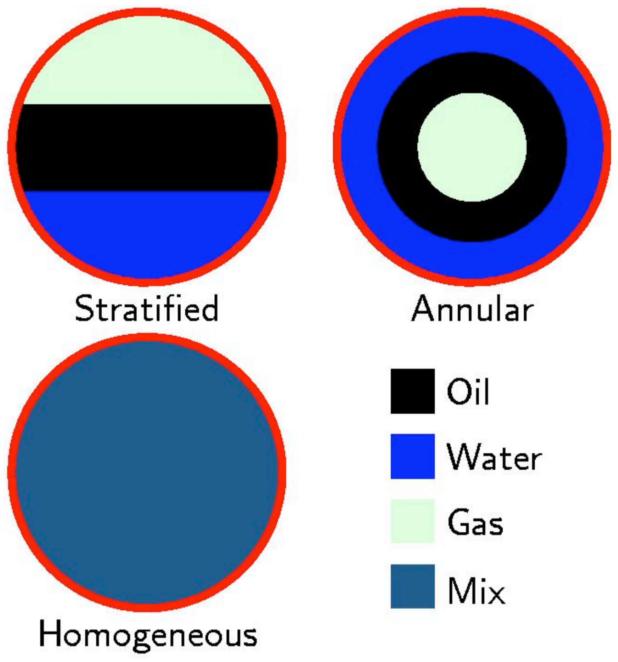


Detector

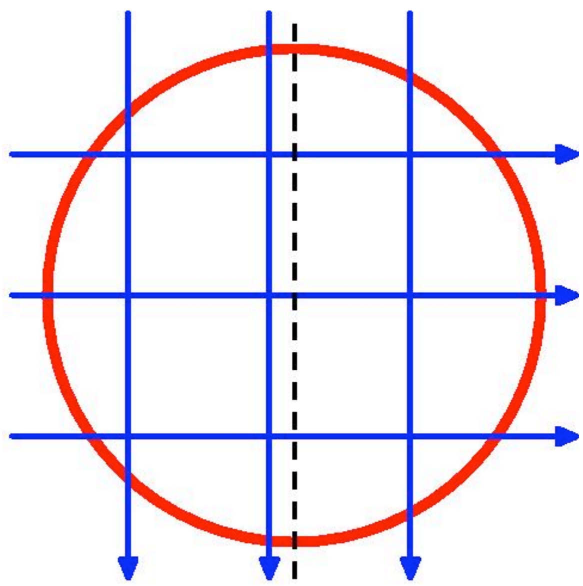
Complication due to Flow Velocity

1. Low Velocity: Stratified configuration
   – Oil floats on top of water, gas above oil
2. Medium Velocity: Annular configuration
   – Concentric cylinders of Water, oil, gas
3. High-Turbulence: Homogeneous
   – Intimately mixed

- Single beam is insufficient
  – Horizontal beam thru stratified indicates only oil



Multiple dual energy gamma densitometers

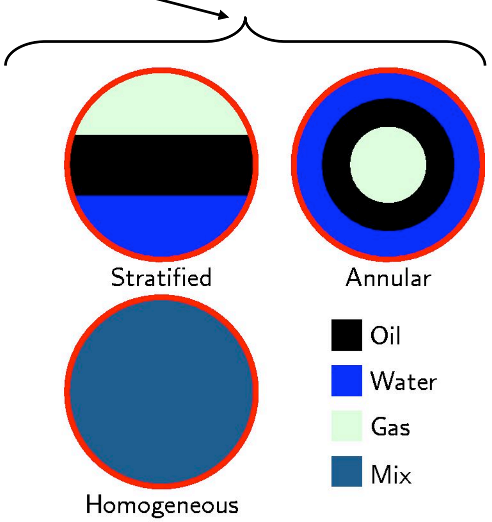


- Six Beams
- 12 measurements
  - attenuation

Prediction Problems

1. Predict Volume Fractions of oil/ water/gas
2. Predict geometric configuration of three phases

- Twelve Features
  - Fractions of oil and water along the paths
- Learn to classify from data

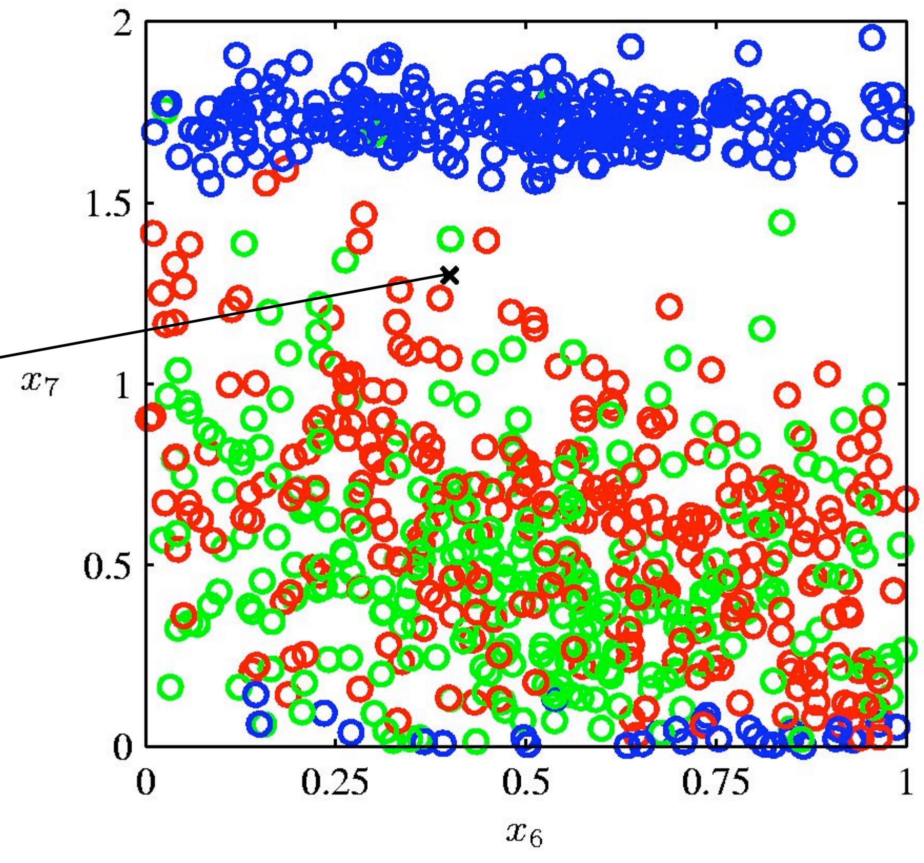


Feature Space

- Three classes (Stratified,Annular,Homogeneos)

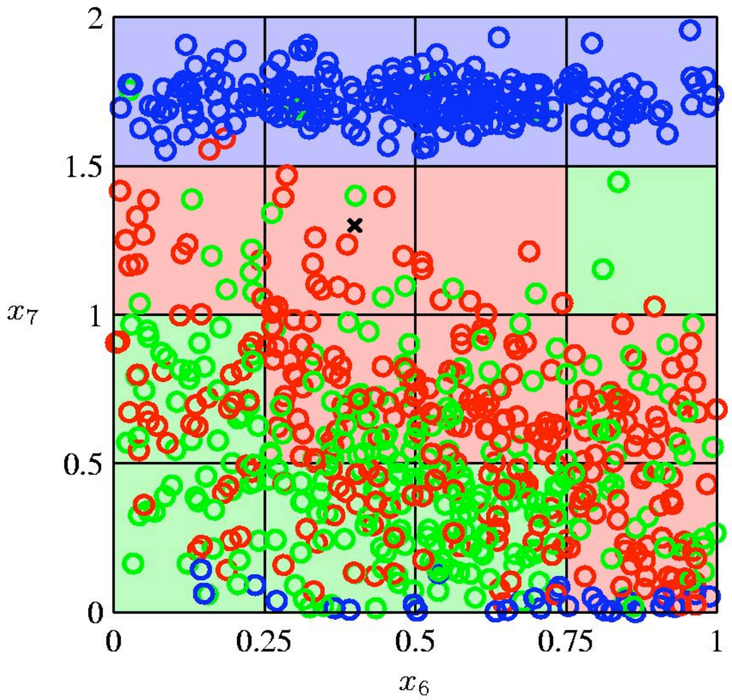
- Two variables shown
- 100 points

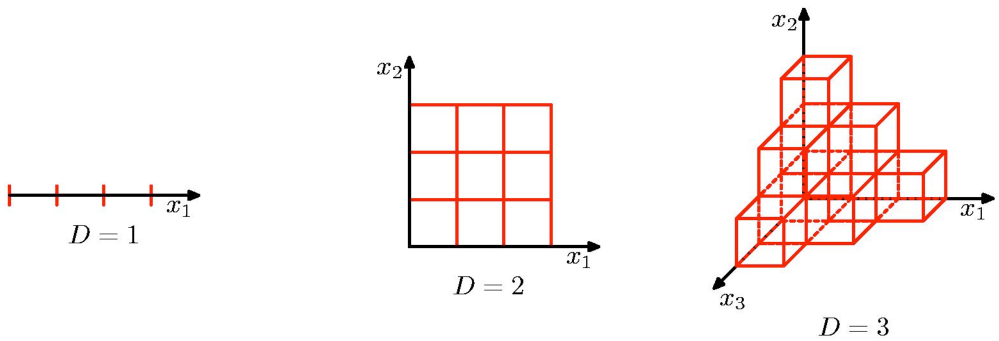
Which class should x belong to?



Cell-based Classification

- Naïve approach of cell based voting will fail
  - exponential growth of cells with dimensionality
  - 12 dimensions discretized into 6 gives 3 million cells
- Hardly any points in each cell





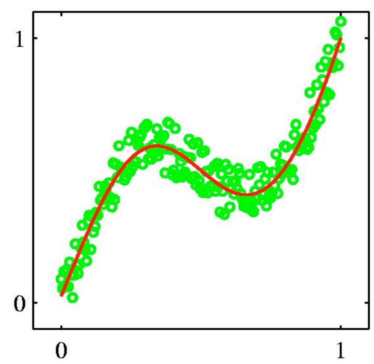
Popular Statistical Models

- Generative
  - Naïve Bayes
  - Mixtures of multinomials
  - Mixtures of Gaussians
  - Hidden Markov Models (HMM)
  - Bayesian networks
  - Markov random fields

- Discriminative
  - Logistic regression
  - SVMs
  - Traditional neural networks
  - Nearest neighbor
  - Conditional Random Fields (CRF)

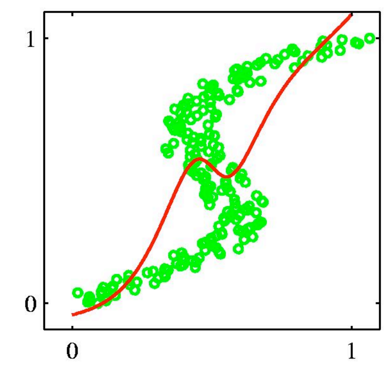
Regression Problems

Forward problem
data set



Red curve is result of
fitting a two-layer
neural network
by minimizing
sum-of-squared
error

Corresponding inverse
problem by reversing
x and t

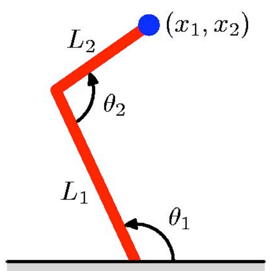


Very poor fit
to data:
GMMs used here

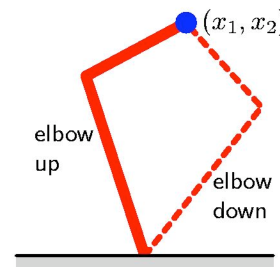
Forward and Inverse Problems

- Kinematics of a robot arm

Forward problem:
Find end effector position given joint angles
Has a unique solution



Inverse kinematics: two solutions:
Elbow-up and elbow-down



- Forward problems correspond to causality in a physical system
  have a unique solution
  e.g., symptoms caused by disease
- If forward problem is a many-to-one mapping, inverse has multiple solutions

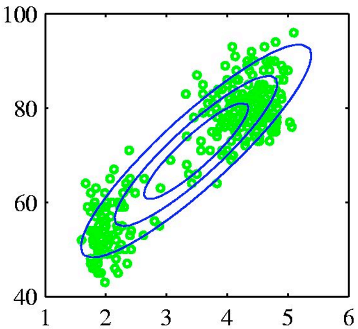
Clustering

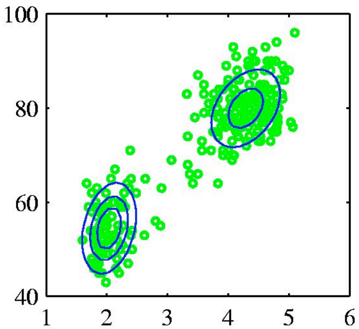
- Old Faithful (Hydrothermal Geyser in Yellowstone)
  - 272 observations
  - Duration (mins, horiz axis) *vs* Time to next eruption (vertical axis)
  - Simple Gaussian unable to capture structure
  - Linear superposition of two Gaussians is better
- Gaussian has limitations in modeling real data sets
- Gaussian Mixture Models give very complex densities

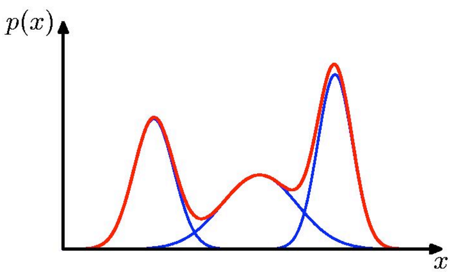
$$p(\mathrm{x}) = \sum_{k=1}^{K} \pi_k N(x \mid \mu_k, \Sigma_k)$$

π_k are mixing coefficients that sum to one

- One –dimension
  - Three Gaussians in blue
  - Sum in red







Estimation for Gaussian Mixtures

- Log likelihood function is

$$\ln p(X \mid \pi, \mu, \Sigma) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k N\left(\mathrm{x}_n \mid \mu_k, \Sigma_k\right) \right\}$$

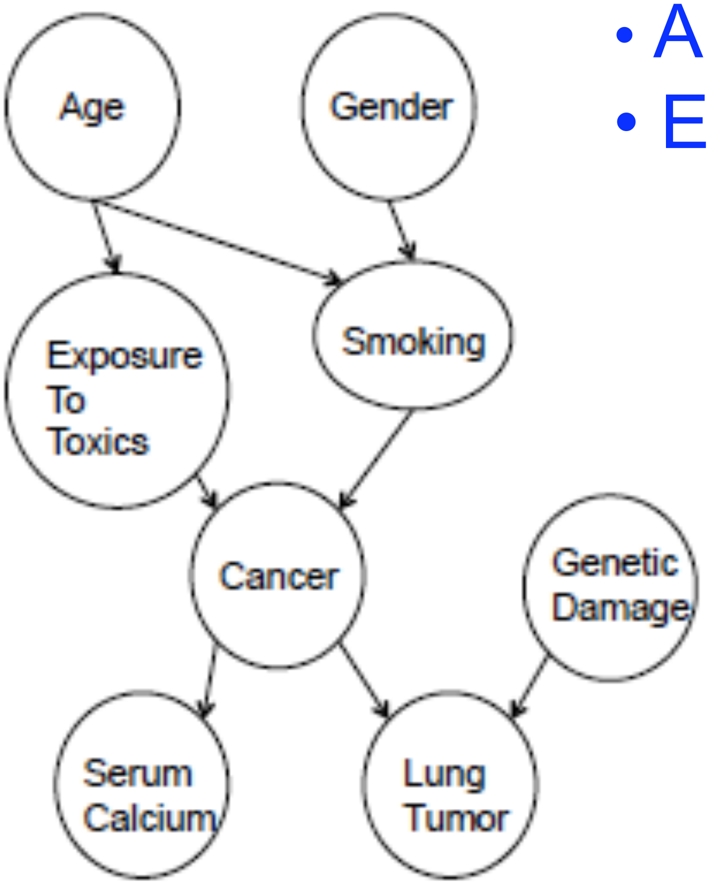
- No closed form solution
- Use either iterative numerical optimization techniques or *Expectation Maximization*

Bayesian Representation of Uncertainty

- Not just frequency of random, repeatable event
- It is a quantification of uncertainty
- Example: Whether Arctic ice cap will disappear by end of century
  - We have some idea of how quickly polar ice is melting
  - Revise it on the basis of fresh evidence (satellite observations)
  - Assessment will affect actions we take (to reduce greenhouse gases)
- Handled by general Bayesian interpretation

- Use of probability to represent uncertainty is not an ad-hoc choice
- If numerical values represent degrees of belief,
  - then simple axioms for manipulating degrees of belief leads to sum and product rules of probability

Modeling Uncertainty



- A Causal Bayesian Network
- Example of Inference:

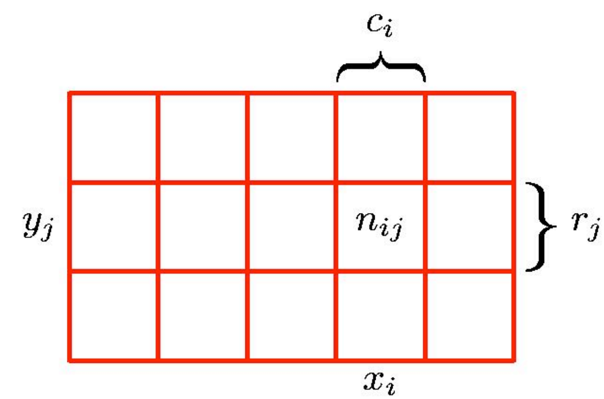
  Cancer is independent of Age and Gender given exposure to Toxics and Smoking

The Fully Bayesian Approach

- Bayes Rule
- Bayesian Probabilities
- Concept of Conjugacy
- Monte Carlo Sampling

Rules of Probability

- Given random variables X and Y
- Sum Rule gives Marginal Probability



$$p(X = x_i) = \sum_{j=1}^{L} p(X = x_i, Y = y_j) = \frac{c_i}{N}$$

- Product Rule: joint probability in terms of conditional and marginal

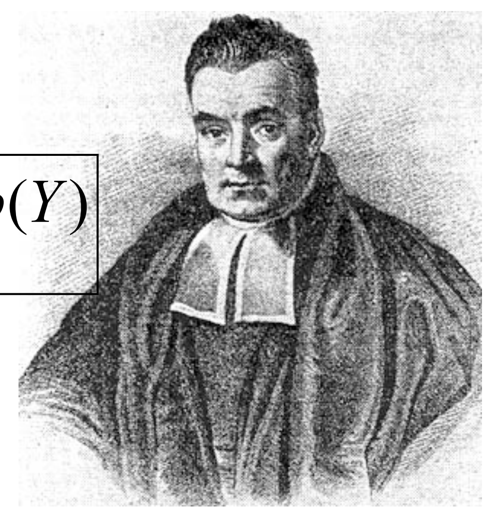
$$p(X,Y) = \frac{n_{ij}}{N} = p(Y \mid X)p(X) = \frac{n_{ij}}{c_i} \times \frac{c_i}{N}$$

- Combining we get Bayes Rule

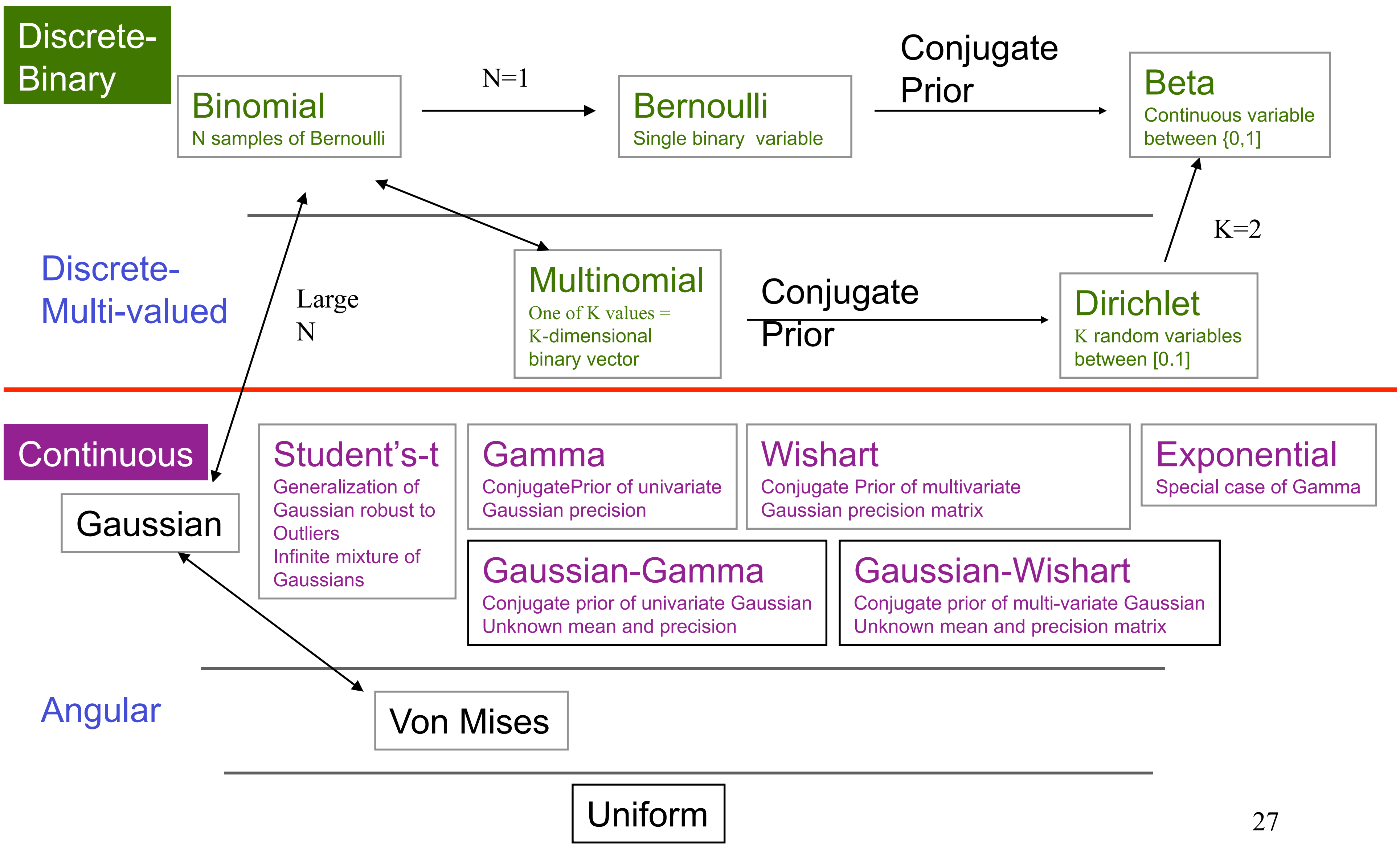
$$p(Y \mid X) = \frac{p(X \mid Y)p(Y)}{p(X)}$$ where $$p(X) = \sum_{Y} p(X \mid Y)p(Y)$$

Viewed as

Posterior α likelihood x prior



Probability Distributions: Relationships



Fully Bayesian Approach

- Feasible with increased computational power
- Intractable posterior distribution handled using either
  - variational Bayes or
  - stochastic sampling
    - e.g., Markov Chain Monte Carlo, Gibbs

Summary

- ML is a systematic approach instead of ad-hockery in designing information processing systems
- Useful for solving problems of classification, regression, clustering and modeling
- The fully Bayesian approach together with Monte Carlo sampling is leading to high performing systems
- Great practical value in many applications
  - Computer science
    - search engines
    - text processing
    - document recognition
  - Engineering