Announcement

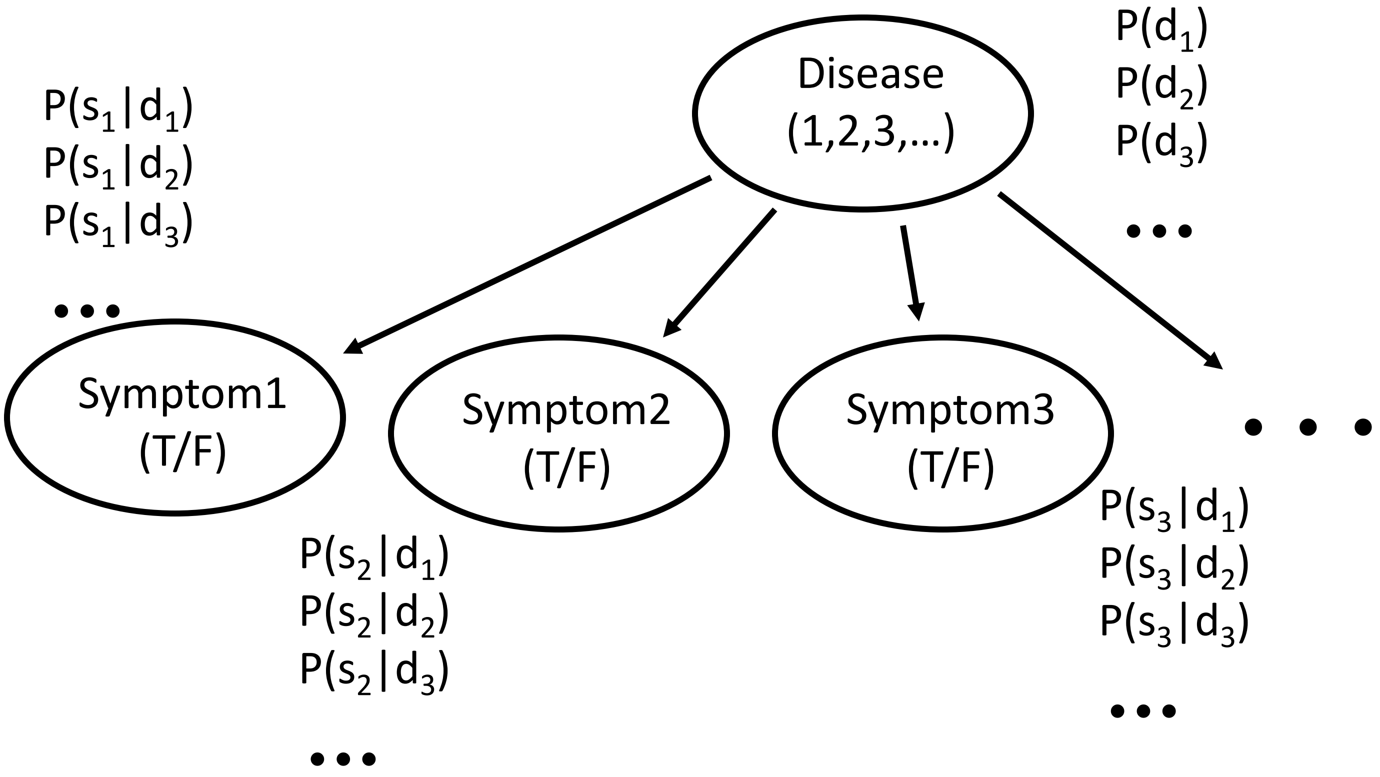
- HW4 on BNs due Tuesday
- Machine Learning Next
- Chapters 18 & 20 in text

Naïve Bayes

- Symptoms (attributes) are conditionally independent of each other given the Disease (classification)
- Stringent assumptions / Impoverished expressiveness
- Works surprisingly (?) well in practice
- Common first choice

NOTE the use of intentionally impoverished model (for tractability - recall coffee cup)

Naïve Bayes



P(s_1|d_1)
P(s_1|d_2)
P(s_1|d_3)
…

Disease
(1,2,3,…)

P(d_1)
P(d_2)
P(d_3)
…

Symptom1
(T/F)

Symptom2
(T/F)

Symptom3
(T/F)

• • •

P(s_2|d_1)
P(s_2|d_2)
P(s_2|d_3)
…

P(s_3|d_1)
P(s_3|d_2)
P(s_3|d_3)
…

Infer likely disease:

$$\boxed{\underset{d_i}{\arg\max}\, P\left(d_i \mid \overline{S}\right)}$$

Naïve Bayes

Diagnose by inference with:

$$\arg\max_{d_i} P(d_i \mid \overline{S})$$

$$\arg\max_{d_i} \frac{P(d_i)\cdot P(\overline{S} \mid d_i)}{P(\overline{S})}$$

$$\arg\max_{d_i} \frac{P(d_i)\cdot \prod_j P(s_j \mid d_i)}{P(\overline{S})}$$

$$\arg\max_{d_i} \left(P(d_i) \prod_j P(s_j \mid d_i) \right)$$

|D| Functions, each assigns probability
over the Boolean hypercube

Which are / are not probability models?

Fourth is not normalized (why does that work?)

Parameters are adjusted to best fit the world samples

What are the parameters?

A kind of machine learning

We will consider the probability functions
the log probability functions
the decision boundaries

Naïve Bayes

$$\boxed{\arg\max_{d_i}\left(P(d_i)\prod_j P(s_j \mid d_i)\right)}$$

- Suppose we always reason from observed symptoms to diseases
- Characterize the boundaries
- Log(x) is monotonically increasing, so:

$$\boxed{\arg\max_{d_i} Log\left(P(d_i)\prod_j P(s_j \mid d_i)\right)}$$

- Log of a product is…

Naïve Bayes

$$\boxed{\arg\max_{d_i} Log\left(P(d_i)\prod_j P(s_j \mid d_i) \right)}$$

- We represent the symptoms S (evidence) is a Boolean vector:

  s_k = 1 if k'th symptom is present
  = 0 if absent

- Becomes argmax $_i$ of $F_i(S)$
  where S is the Boolean vector of evidence

- $F_i(S) = \text{Log } P(d_i) + \sum \text{Log } P(s_k \mid d_i)$

- Form is $F_i(S) = a_i + B_i \cdot S$ B_i is a vector of weights (one for each…?)

  $a_i = \text{Log } P(d_i) + \sum \text{Log } P(s_k=0 \mid d_i)$
  $b_{ik} = \text{Log } P(s_k=1 \mid d_i) - \text{Log } P(s_k=0 \mid d_i)$

- The Log Prob fcns are hyperplanes over the Boolean hypercube

Naïve Bayes

$$\boxed{\arg\max_{d_i} Log\left(P(d_i)\prod_j P(s_j \mid d_i) \right)}$$

- Form $F_i(S) = a_i + B_i \cdot S$

  Each is a linear polynomial in S

- Diagnose using the highest valued function at a point S

- What are the Naïve Bayes decision boundaries?

- (How do hyperplanes interact?)

- How do we determine a's and B's?

- These are *generative models* for the diseases

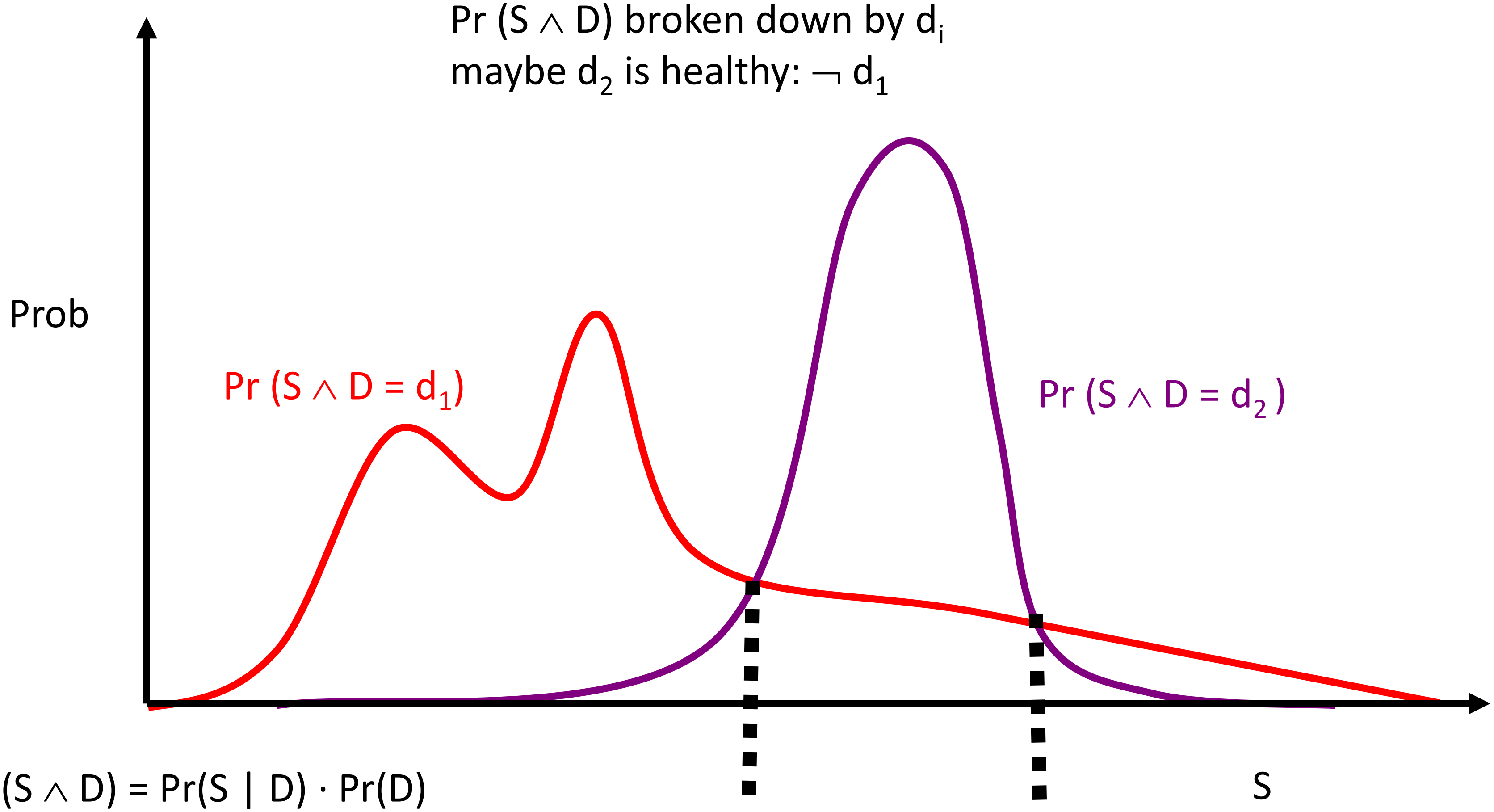
Generative vs. Discriminative Models

- Generative
  - Model “generation” of data (if perfect, don’t need the world to sample)
  - Represent joint probability approximately* in some form
    - Pr($D=d_i$, $S=s_k$) or Log Pr($D=d_i$, $S=s_k$) or …
    - Usually a compact representation (e.g., Bayes net)
  - Infer d_i query probabilities from the generative model
    - From observed evidence S
    - Compute Pr($D=d_i \mid S=s_k$) for each i,
    - Choose highest
    - Note: Can infer other things, like Pr($S=s_k$)
- Discriminative
  - Model the boundaries between d_i values of D
  - Represent the query probabilities directly
    - Perhaps model Pr($D=d_i \mid S=s_k$)
    - Or more directly F: $S \rightarrow D$
    - Note: Can no longer infer Pr($S=s_k$)

* sometimes VERY approximately

Generative vs. Discriminative Models

Pr (S $\wedge$ D) broken down by d_i
maybe d_2 is healthy: $\neg$ d_1



Discriminative
Boundaries

Pr(S $\wedge$ D) = Pr(S | D) · Pr(D)

Note boundaries as Pr(D) changes

Generative vs. Discriminative Models

Encountered before in

Reinforcement Learning

Which is which?

Machine Learning

using fitted generative model

- Adopt a parametric model
- LEARNING: Train model on data / observations
  - A little training for a simple model
  - A LOT of training for a complex model
- Estimate parameter values from data

- INFERENCE: Apply the model
  - Compute probabilities conditioned on evidence
  - Answer queries
  - Classify new data inputs

Markov Random Fields

- Undirected graphical models
- No ordering of nodes for construction
- Denote “direct” causes
- Markov blanket = neighboring nodes only
- Simple conditional independence given evidence
- BN → MRF, moralize the graph
  (connect unconnected parents)
- Potential functions (unnormalized) over maximal cliques
- Z = sum over assignments of product of potentials is the partition function (from statistical physics)
- Boltzmann Gibbs distribution
- Learning and inference are generally harder (iterative)

Simple Word Probability Models for English

- Train from some text corpus NYT, Shakespeare, alt.politics.paranoid,…
- Pr (w_i) (1st order Markov)

  REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME CAN DIFFERENT NATURAL HERE HE THE A IN CAME THE TO OF TO EXPERT GRAY COME TO FURNISHES THE LINE MESSAGE HAD BE THESE

- Pr ($w_{i+1} \mid w_i$) (2nd order Markov)

  THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHOEVER TOLD THE PROBLEM FOR AN UNEXPECTED

Train Three Model Instances

- alt.politics.paranoid
- alt.politics.republican
- alt.politics.democrat

- Get new texts
  - newsgroup postings
  - Speeches
  - NYT articles
- Compute fit

  (probability of each model given the text)

The Infamous Dr. Bayes

(or is logical inference really so bad?)

- Dr. Bayes has a statistics degree
  (*not* an MD)
- He makes diagnoses using his rule and other notions from statistics
- A plague has descended; there are two treatments: A and B
- He has tried them both and seen the results on his own patients
- He does not see patients but holds phone consultations with other (real) doctors

Find a data sample that justifies the following interchange with Dr. Bayes

Is the patient male or female?
Male
then administer treatment A

Is the patient male or female?
Female
then administer treatment A

Is the patient male or female?
Unknown
then administer treatment B

Is this POSSIBLE?

- How to tell a statistician from a normal individual

How to proceed?

- Build an empirical model (well, hybrid – in fact mostly analytic)
- Dr. Bayes interactions are constraints on model parameters
- Then, either
  - Choose parameters to satisfy constraints

    and

    make up data to yield these parameters
  - OR Convince ourselves of their inconsistency