Lecture #3: Review of Probability Theory, continued

Objectives
- Build a sound framework for representing uncertainty
- Understand the fundamentals of probability theory
- Prepare to use Bayesian networks / directed graphical models extensively!
- Specifically today: understand Bayes’ Law (aka Bayes’ Theorem)

Bayes’ Theorem

Given \( P(A|B) \) and \( P(A) \) and \( P(B) \), can we calculate \( P(B|A) \)?

\[
P(B | A) = \frac{P(A | B) P(B)}{P(A)}
\]

Bayes, Thomas (1763)
“An essay towards solving a problem in the doctrine of chances.”
Philosophical Transactions of the Royal Society of London 53:370-418

Proof of Bayes’ Theorem

\[
\begin{align*}
\text{Proof.} \\
1. \quad & P(A \cap B) = P(A | B) \cdot P(B) \\
2. \quad & P(A \cap B) = P(B | A) \cdot P(A) \\
3. \quad & P(A) = P(B | A) \cdot P(A) \\
4. \quad & P(A) = \frac{1}{P(A)} \\
\end{align*}
\]

Announcements
- Feedback on Reading Report #1
- Assignment #1
  - Early day: Tuesday
  - Due: Wednesday
  - Submit on Learning Suite
- Start planning your 2015 Summer now:
  - Full-time or Internship
  - Graduate school
  - Aim high!
  - Start here: bridge.byu.edu
The Denominator

We need to show that $A \cap B$ and $A \cap \overline{B}$ are mutually exclusive.

Thus, we can apply the third axiom of probability ...

The Law of Total Probability

Computing $P(A)$ from Partition

More generally,

Bayes’ Theorem (2)

What is the relationship of the numerator to the denominator?
Using Bayes’ Theorem: Diagnosis

- Test T for some rare phenomenon G (1 in 100,000)
- In presence of G, T will have positive indication 95% of time.
- In absence of G, T will have positive indication 0.005% of time.
- Suppose T gives a positive indication.
- What is the probability that G is actually present?

Take out pencil and paper. You solve!

Diagnosis (cont.)

\[
p(T \mid G) = \frac{p(G)p(T \mid G)}{p(T)}
\]

\[
p(G) = \frac{p(G)p(T \mid G)}{p(T)}
\]

\[
p(T) = \frac{p(T \mid G)p(G) + p(T \mid \neg G)p(G)}{p(T)}
\]

Diagnosis (cont.)

\[
p(T \mid G) = 0.95 \quad p(T \mid \neg G) = 0.00005
\]

\[
p(G) = 0.0001 \implies p(\neg G) = 1 - p(G)
\]

\[
p(T) = \frac{p(G)p(T \mid G)}{p(T)} + \frac{p(\neg G)p(T \mid \neg G)}{p(T)}
\]

\[= 0.1597 = 16\%
\]

Using Bayes’ Theorem: Classification

Given A, want most probable vector \(\mathbf{b}\):

\[
\mathbf{b} = \underset{\mathbf{b}_j}{\text{arg max}} \ p(\mathbf{A} \mid \mathbf{b}_j)
\]

\[= \underset{\mathbf{b}_j}{\text{arg max}} \ p(\mathbf{A} \mid \mathbf{b}_j)p(b_j) \]

Note: We can use our distributions \(p(\mathbf{A} \mid \mathbf{B}_j)\) and \(p(\mathbf{B}_j)\) to make classification decisions!
What's Next?

- Remainder of Probability Theory
- Random Variables
- Important Ideas from Information Theory
- Bayesian Networks
- Joint (Generative) Models