Announcements

Group office hours confirmed (but not necessarily this week)
   Fridays, 10:30-11:30 in GS 919

Good times to schedule an appointment confirmed
   Monday, Tuesday, Thursday: 9:30

First preview slide set is available (expect changes!)

Optional homework assignment now posted.

Why does modeling work at all?

Complex example:
   Video feed from a camera watching the world

Summary points:
   There is structure in the world
   Real world high dimensional data is not!
   Brute force representation of a high dimensional distribution is a bad idea for two reasons
      It is completely impractical
      It misses the forest for the trees

More summary

Our models should capture what is important

1) Mechanistically
2) Statistically (simplify the joint density)

Clustering in high dimensions

\[ p(x_1, x_2, \ldots, x_n) = \sum_c p(c) \prod_i p(x_i | c) \]
Generative models

Informally, tells a story about how data comes to be
Illustrated using ancestral sampling

\[ p(a,b,c) = p(c)p(a|c)p(b|c) \]

Bayesian inference

- likelihood function for the parameters
- prior probability
- posterior probability

\[ P(\Theta | x) = \frac{P(x | \Theta) P(\Theta)}{P(x)} \]

normalizer, often is not of interest

Simple example*

- What you know
  - John is coughing
- What do you conclude?
  - John has a cold
  - John has lung cancer
  - John has stomach problems

Why this approach

Separates representation, modeling, and inference
Model is separated into prior and likelihood
Encourages being precise about the relationship between models and observed data
Handles fitting and learning similarly
What is known is always represented as a distribution

*Adopted from Josh Tenenbaum

Read first chapter of K&F now linked from the lecture notes page.