Computational Cognitive Science

Lecture 15: Overhypotheses

Chris Lucas

School of Informatics

University of Edinburgh

November 8, 2022

Reading

Kemp, Perfors, and Tenenbaum, 2007. (link)

Priors

Having good priors is useful – we need them to generalize.

Where do priors come from?

Priors

Implicit, basic:

- Light comes from overhead
- If I got sick, it was probably something I ate (Garcia effect)

Priors

Yesterday's posterior:

- How likely is rain tomorrow?
- How long will it take to find a free spot at the library?
- How reliably do blickets activate my machine?

Priors

We have focused on priors about concrete hypotheses:

- Biases in coins and consumer choices
- How often background causes produce an effect
- What number concepts are people likely to think of

Generalization and abstraction

This is all useful, but what about

- learning to generalize beyond our experience?
- learning abstract concepts?

Generalization and abstraction

E.g., discovering

- people don't always agree: Pineapple pizza
- people aren't always right: whipped garlic, not hummus
- shape is often a marker of category membership
- causality without contact is more common in some domains than others

Generalization and abstractions

Enter **overhypotheses**: Hypotheses about hypotheses.

Overhypotheses

Imagine you have a bag of marbles. You pull out a red marble.

What's the probability that the next marble is red?

What distributions of colors are likely?

(Based on Kemp, Perfors, & Tenenbaum (2007;link))

Overhypotheses

Your answer depends on how homogeneous you think the bag is.

Now imagine you've seen five bags of marbles and pulled two marbles out of each:

- green, green
- blue, blue
- blue, blue
- red, red
- yellow, yellow

Now how likely is it that the next marble is red?

Overhypotheses

What if you'd instead seen:

- green, blue
- blue, blue
- blue, red
- red, yellow
- yellow, green

Now how likely is it that the next marble is red?

Feature variability

We are able form expectations about how variable features are.

Kemp et al. developed a computational model to explain these phenomena.

Kemp et al.'s model

Recall the Dirichlet distribution, and how we can parameterize it in terms of concentration (α) and bias (β):

- $\sum_i \beta_{\mathbf{i}} = 1$
- $\alpha > 0$

Previously, we picked α and β. Here we learn them.

Kemp et al.'s model

- $\alpha \sim \mathrm{exponential}(1)$ (i.e., homogeneous bags are more likely)
- $\beta \sim \mathrm{Dir}(1, 1, ..., 1)$ (uniform)
- $\theta^i \sim \mathrm{Dir}(\alpha\beta_1, \alpha\beta_2, ...)$: The proportions for the i^{th} bag
- $p(\alpha|d) \propto p(d|\alpha)p(\alpha)$
- $p(d|\alpha) = \int d\theta p(d|\theta)p(\theta|\alpha)$

We can infer distributions over α, β, and θ w/Monte Carlo methods.

Kemp et al.'s model

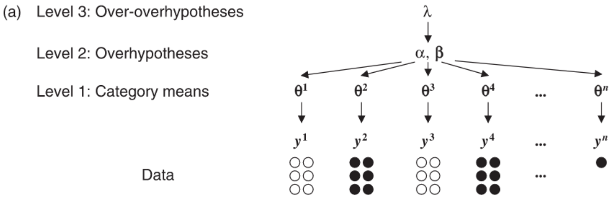


Figure 1a from Kemp et al.

Kemp et al.'s model

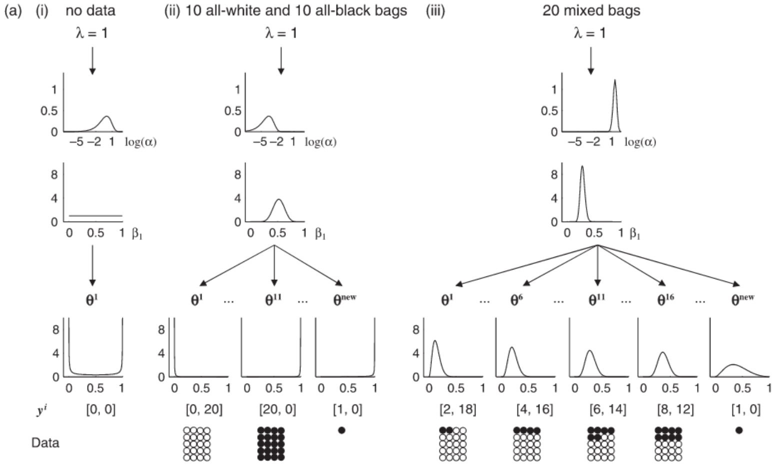


Figure 3 from Kemp et al.: Predictions of their hierarchical model

Kemp et al.'s model

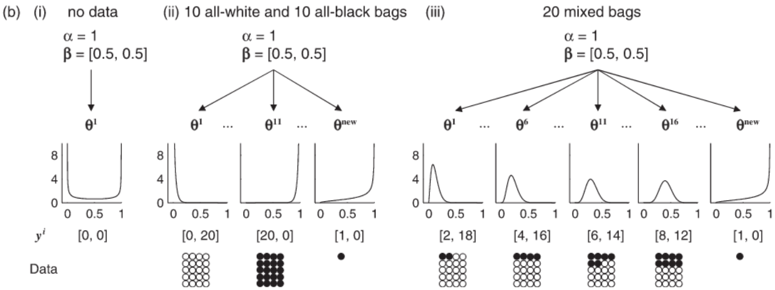
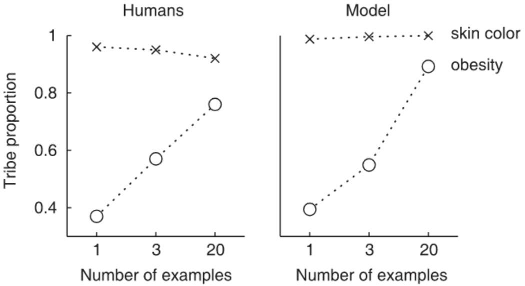


Figure 3b from Kemp et al.: Predictions of a non-hierarchical model, for contrast

Kemp et al.'s model

This model can explain other phenomena, e.g., that people are quicker to generalize from some features than others:



(Fig. 4 from Kemp et al., based on data from Nisbett et al. 1983)

Shape bias

This model can also help explain **shape bias**:

- Children tend to classify rigid objects based on their shape, rather than their color, material, or size
- Linda Smith and colleagues (2002) provided evidence that this bias is learned from experience (or at least can be), using a **training study**

Shape bias

In categorization, our hypotheses dictate what gets grouped with what. We can treat category labels as just another feature.

Shape bias

But how do we decide our basis for grouping things together?

What features are informative, and how?

Overhypotheses can help us decide what features are diagnostic of category membership

- e.g., shape, material, size, solidity

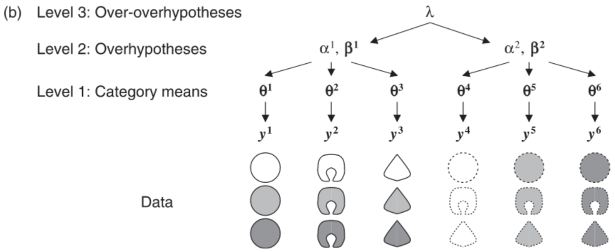


Figure 1b from Kemp et al.

Shape bias

x_i is a (categorical) feature; x_1 is shape, x_2 is material.

For a given general kind of thing k we're dealing with,

$$\alpha_i^k \sim \mathrm{exponential}(1)$$

If we see rigid objects ($k = 1$) with a similar shape go together but can have different materials:

- low values of α_1^1 are likely (homogeneous shape)
- high values of α_2^1 are likely (heterogeneous material)

If we see that piles of the same material ($k = 2$) go together regardless of their shape, the converse is true for α_i^2.

Shape bias

Where does k come from?

How can we learn what different kinds of things there are?

Intuition behind the expanded model:

- Each category is a member of an “ontological kind”
- We don’t know how many ontological kinds there are
- Each ontological kind has an associated set of parameters that govern the behavior of its members

Have we encountered something like this elsewhere in the course?

Shape bias

Recall model of individual differences in Navarro et al.:

- Each person is a member of a cluster
- We don't know how many clusters there are in our data set
- Each cluster has an associated set of parameters that govern the behavior of its members

We can use the same "Chinese restaurant process" prior over partitions.

Shape bias

The full generative model (from the Appendix of Kemp et al.):

$$
\begin{aligned}
\mathbf{z} &\sim \mathrm{CRP}(\gamma) \\
\alpha^k &\sim \mathrm{exponential}(\lambda) \\
\beta^k &\sim \mathrm{Dir}(\mathbf{1}) \\
\theta^i &\sim \mathrm{Dir}(\alpha^{z_i}\beta^{z_i}) \\
\mathbf{y}^i | n^i &\sim \mathrm{multinomial}(\theta^i)
\end{aligned}
$$

Armed with this generative model, we can use MCMC methods like Stan does to make predictions.

Discovering ontological kinds

Trained on category assignments and feature labels. . .

(a) Training

Category	1 1 2 2 3 3 4 4
Shape	1 1 2 2 3 4 5 6
Material	1 2 3 4 5 5 6 6
Size	1 2 1 2 1 2 1 2
Solidity	1 1 1 1 2 2 2 2

(b) Second-order generalization

S

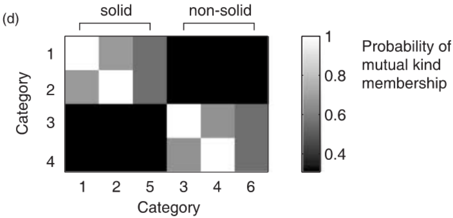
5	? ?
7	7 8
7	8 7
1	1 1
1	1 1

N

6	? ?
8	8 9
8	9 8
1	1 1
2	2 2

(Figure 7 from Kemp et al.)

Discovering ontological kinds



The model inferred two "ontological kinds" and was able to classify items as being in new categories but known kinds.

Questions

- Human cognition is very flexible – how can models capture this flexibility?
  - Hand-picked parametric overhypotheses may not suffice
- People tend to make judgments quickly – how can we capture this efficiency, and the trade-offs that entails?

References

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, 10(3), 307-321.

Nisbett, R.E., Krantz, D.H., Jepson, C., & Kunda, Z. (1983).The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*, 90(4), 339–363.

Smith, L.B., Jones, S.S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science*, 13(1), 13–19.