Pushing the Limits of Learning from Limited Data

Maya Malaviya[1]*, **Ilia Sucholutsky**[2]*, **Thomas L. Griffiths**[2,3]

[1]Department of Computer Science, Stevens Institute of Technology
[2]Department of Computer Science, Princeton University
[2]Department of Psychology, Princeton University
mmalaviy@stevens.edu, is2961@princeton.edu, tomg@princeton.edu

Abstract

What is the mechanism behind people's remarkable ability to learn from very little data, and what are its limits? Preliminary evidence suggests people can infer categories from extremely sparse data, even when they have fewer labeled examples than categories. However, the mechanisms behind this learning process are unclear. In our experiment, people learned 8 categories defined over a 2D manifold from just 4 labeled examples. Our results suggest that people are forming rich representations of the underlying categories despite this limited information. These results push the limits of how little information people need to build strong and systematic category representations.

Introduction

Categorization is a basic problem faced by any organism that hopes to capture the abstract structure underlying perceptual experience and is a classic topic of research in cognitive psychology (e.g., Bruner and Austin 1956; Medin and Schaffer 1978; Nosofsky 1986). Humans can learn categories from limited data, inferring novel concepts from just a few examples (e.g., Xu and Tenenbaum 2007). In recent work, we showed people can infer categories even if given fewer examples (M) than the number of categories (N) (Malaviya et al. 2022). Although there is theoretical evidence that machines can categorize in this few-shot regime (Sucholutsky and Schonlau 2021), more work is needed to see how we can generalize this data-efficient classification to more complex domains. Our previous human experiments only considered scenarios where categories formed intervals along a 1D manifold, and participants were given $M = N - 1$ labeled examples. So, it is unclear whether people are actually learning sophisticated representations of high-dimensional feature spaces in order to infer the space of the Mth class. In this paper, we probe the limits of the human ability to form category representations with sparse data by generating stimuli on a 2D manifold and showing $M = 4$ examples while soliciting categorization judgments for $N = 8$ categories. Our results provide evidence that people can indeed represent more categories than the number of examples they are shown. Furthermore, people's judgments are similar to predictions from a weighted nearest-neighbor model, providing a way to understand how people, and perhaps machines, form generalizations from very sparse data.

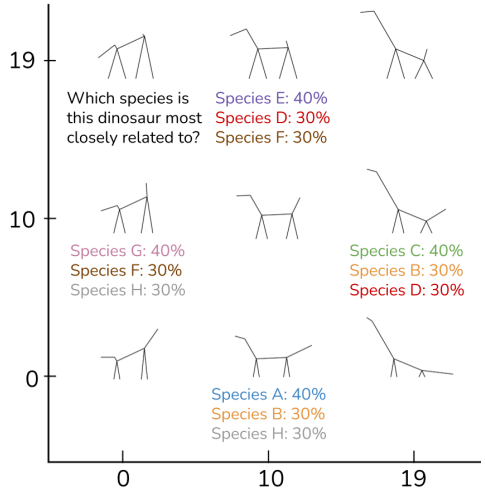

Figure 1: 9 of the 400 stimuli along the 2D manifold, 4 of which are the soft-labeled examples. The stimulus at (0,19) is annotated with the question participants saw for each trial.

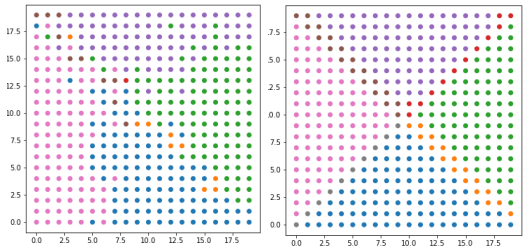

Figure 2: Left: Participant classification majority vote for each stimulus in the 20×20 grid. Right: Simulated classification majority vote using a weighted nearest neighbor model.

Methods

Stimuli The stimuli were images of stick figures representing quadrupeds, adapted from Malaviya et al. 2022; San-

*These authors contributed equally.

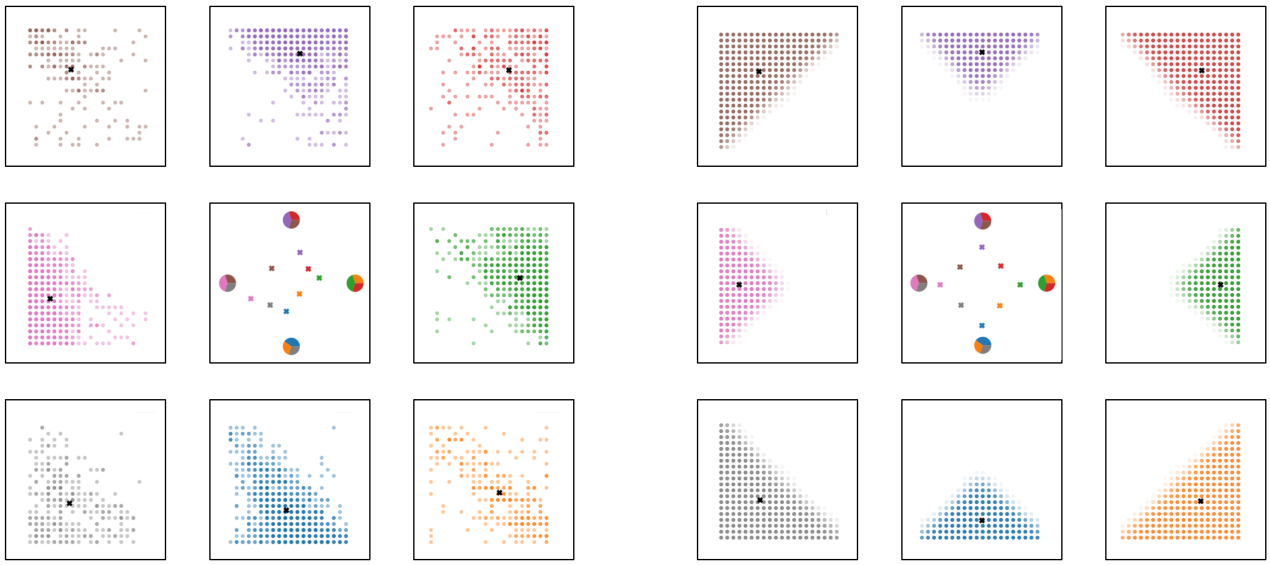

Figure 3: Left: Participant classifications of stimuli along a 2D manifold into 8 classes based on 4 soft-labeled examples. Each outer scatter plot marks points on the manifold that were assigned to each respective class (distinguished by color), with the black 'X' marking the class centroid. The inner scatter plot has a colored 'X' corresponding to each class centroid, and pie charts denoting the soft-labeled examples at their locations on the manifold. Right: Simulated classifications using an exponential distance-weighted 4-nearest neighbor model.

born and Griffiths 2008. They have 9 distinct, continuous features, so they are points in 9D feature space. The stimuli were generated by selecting feature values for three stick figures (X, Y, Z), then taking linear combinations of these feature values to produce more stick figures. Each figure can be represented as $D = i(X - Y) + j(X - Z)$ where $i, j \in \{1 + n * 0.1\}_{n=0}^{10}$ are scaling factors and X, Y, Z are vectors containing the manually selected feature values. This yields 400 stimuli organized in a 20×20 grid, sampling a 2D manifold in a 9D feature space.

Four of the stimuli were labeled (or "soft-labeled") with probabilities of belonging to eight different classes, enumerated A-H, e.g., one stick figure in Figure 1 at location (10,0) on the grid was labeled 40% A, 30% B, 30% H, and 0% C-G.

Experimental Procedure The four labeled stimuli were shown to participants ($n = 41$, from Prolific), who were told they represented models that paleontologists use to summarize dinosaur fossil structures. Soft labels were referred to as "genetic information" that described the labeled dinosaurs' relation to different "species" A-H (classes). Each participant was asked to categorize 100 new stimuli into the most likely class. The underlying dimensions used to construct the stimuli were not explicitly revealed to participants.

Results

Figure 2 (left) shows the majority species vote for each stimulus. We see four classes dominate, likely because they were the four for which a stimulus was labeled as being 40% similar (in contrast to the others which only had 30%). However, some minority classes appear in the boundaries between these classes, indicating that people may form some consensus regarding where these classes lie on the manifold. Figure 3 (left) shows the empirical distribution of classifications for each category, and we see that participants are also categorizing into the minority classes. Furthermore, people systematically disentangle the manifold into the 8 classes, and the centroid of each is located near the examples whose soft labels contain non-zero values for that category.

Figures 2 (right) and 3 (right) show classification results from a 4-nearest neighbor model with exponential inverse-distance weighting (i.e., an exemplar model consistent with the universal law of generalization; Shepard 1987). Participant behavior has some similarity to the model, but there still appear to be systematic differences. For instance, why do some minority categories span less of the manifold than predicted? Perhaps participants show a bias towards majority classes, even if they are also making inferences about minority classes, because we only ask for the most likely class, rather than a distribution over classes.

Conclusion

Our findings contribute to the growing discussion regarding limits of data-efficiency in the fields of artificial intelligence and cognitive science. People can learn from more sparse data than previously thought possible; we present evidence of systematic judgments, even when categorizing into 8 classes from just 4 soft-labeled stimuli. Models that capture human judgments can guide directions for few-shot machine classification. Though this abstract highlights similari-

ties to exponentially-weighted generalization algorithms, future work could probe the mechanisms of these judgments by comparing them to predictions from additional models.

Acknowledgments

This work was funded by a grant from The NOMIS Foundation awarded to TLG and supported by an NSERC fellowship (567554-2022) to IS.

References

Bruner, J. S.; and Austin, G. A. 1956. *A study of thinking*. John Wiley and Sons.

Malaviya, M.; Sucholutsky, I.; Oktar, K.; and Griffiths, T. 2022. Can Humans Do Less-Than-One-Shot Learning? In *Proceedings of the 44th Annual Meeting of the Cognitive Science Society*.

Medin, D. L.; and Schaffer, M. M. 1978. Context theory of classification learning. *Psychological Review*, 85(3): 207–238.

Nosofsky, R. M. 1986. Attention, similarity, and the identification–categorization relationship. *Journal of experimental psychology: General*, 115(1): 39.

Sanborn, A.; and Griffiths, T. L. 2008. Markov chain Monte Carlo with people. *Advances in Neural Information Processing Systems*, 1265–1272.

Shepard, R. N. 1987. Toward a Universal Law of Generalization for Psychological Science. *Science*, 237(4820): 1317–1323.

Sucholutsky, I.; and Schonlau, M. 2021. 'Less than one'-Shot Learning: Learning N classes from M $<$ N samples. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 9739–9746.

Xu, F.; and Tenenbaum, J. B. 2007. Word learning as Bayesian inference. *Psychological Review*, 114(2): 245–272.