Improvement of the Bayesian generalization model in order to handle negative examples and discontinuous hypotheses

May 17, 2022

abstract-text

1 Background

The problem our project is focusing on is Shepard's ideal generalization problem. The generalization problem focuses on how humans build hypothesis spaces for a given consequence after observing stimuli. In the paper "Generalization, similarity, and bayesian inference", they discuss how using a model of bayesian inference, we can predict the probability of given stimuli being included within the consequential region [TG02]. The model uses the equation $p(y \in C \mid x) = \sum_{h:y\in h} p(h|x)$ where h is a hypothesis from the hypothesis space $\mathcal{H}$ and $p(h|x)$ is the posterior probability of the hypothesis after observing x. We plan to extend this model to investigate how including negative examples within the x vector(x is the observed stimuli) affect how the model limits hypotheses. We also plan to explore how different distributions and models compare to the original model for generalization.

2 Question

To generalize the model by Tenenbaum et al, we want to find a way, how it can be improved, so that it can handle negative examples and discontinuous hypotheses.

3 Negative examples in the baseline model

The baseline model (Figure 1), which was introduced by Tenenbaum et al. is by itself not capable of handling negative examples.

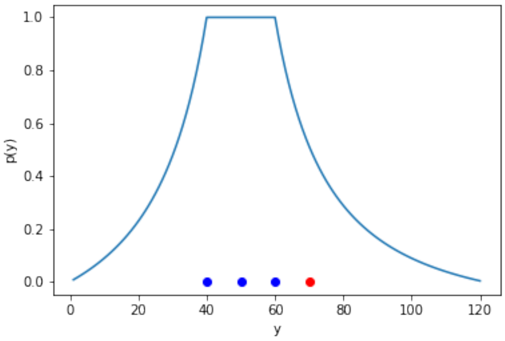

Figure 1: Baseline model fitted on the positive examples [40, 50, 60] (blue) does not incorporate the negative example [70](red)

A naive approach to incorporate negative examples, is to fit a model the same way it was fitted for the positive examples and then calculate the difference. Since the model calculated probabilities, it needs to be ensured, that the sum does not go below zero. If the example from before is used, we see that the negative example is predicted as having probability 0 (Figure 3), which is expected.

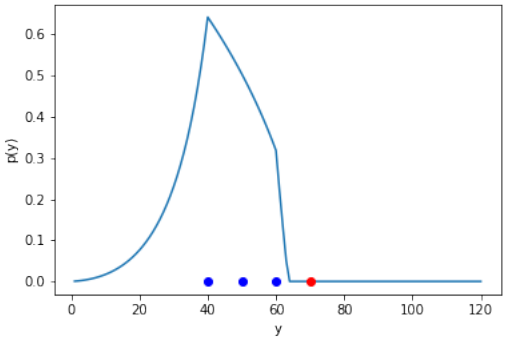

Figure 2: addition model fitted on the positive examples [40, 50, 60] (blue) and negative exaample [70] (red)

The model can also be applied to the case of discontinuous clusters. We can see this in the figure

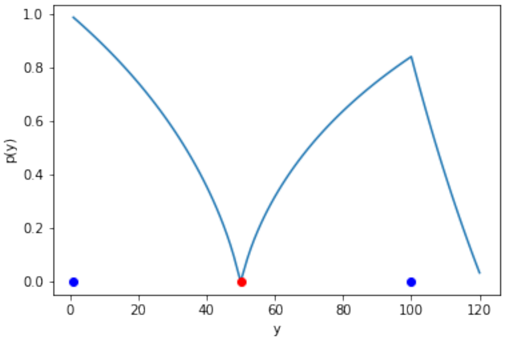

Figure 3: addition model fitted on the positive examples [1, 100] (blue) and negative exaample [70] (red)

Evaluate how negative examples affect the predictions of the original model and use it as a baseline (2) Evaluate the predictions made by the original model for multi cluster sample data

4 Limiting positive regions using negative examples

A typical use case for negative samples under the assumption of a single positive region is limiting that region. We would generally refer to the underlying thought process as one of elimination, which is why we shall call the corresponding model the elimination model. Since our baseline model takes all possible hypotheses into consideration, the process of integrating counterexamples is relatively straightforward: We remove all those hypotheses from consideration which are contradicted by our negative examples. To this end, we modify the likelihood calculation to return zero for all hypotheses which would incorporate a known negative example, therefore removing those hypotheses from consideration. This results in the posterior probabilities for all values on the far side of a negtaive sample (from the perspective of a positive sample) being zeroed out, while retaining the exponential falloff behavior of the base model.

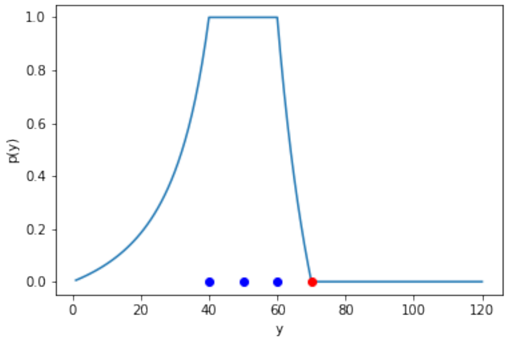

Figure 4: elimination model fitted on the positive examples [40, 50, 60] (blue) and negative exaample [70] (red)

As can be seen in the plot of our example data (Figure 4), the predictions generated by ths approach correctly match the

0% probability we would expect to see for known negative samples. Otherwise, the general shape of the basline curve is retained, leaving its original assumptions intact.

5 Separating positive regions by negative examples

6 Multiple Discontinuous Regions Model

To explore how people predict the consequential region when using a model with multiple discontinuous regions, we will use the following model:

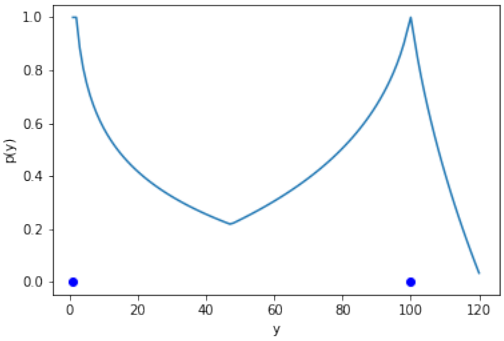

Figure 5: Multiple discontinuous regions model fitted on examples [1, 100] with 2 disjoint regions.

This model is a modification of the original model, which allows for multiple discontinuous regions. It works by splitting the stimuli into n-regions and then calculating the posterior probability for each region. This model is fitted to only positive examples, and works by taking $max(p_i(observation))$

References

[JLK17] Alan Jern, Christopher G. Lucas, and Charles Kemp. "People learn other people's preferences through inverse decision-making". In: *Cognition* 168 (Nov. 2017), pp. 46–64. DOI: 10.1016/j.cognition.2017.

06.017. URL: https://doi.org/10.1016%2Fj.cognition.2017.06.017.

[TG02] Joshua B. Tenenbaum and Thomas L. Griffiths. "Generalization, similarity, and bayesian inference". In: *Behavioral and Brain Sciences* 24 (4 Aug. 2002), pp. 629–640. DOI: 10.1017/s0140525x01000061. URL: https://doi.org/10.1017/s0140525x01000061.