Bayesian Model Comparison for Neural Models of Decision Making

Maryam Meghdadi, Eric Schulz*, Marcel Binz*
Helmholtz Institute for Human-Centered AI, Munich, Germany

HELMHOLTZ MUNICH
HCAI

TL;DR

Goal: We aim to discover theories of human decision-making using neural networks (NNs).

Previous Work: Fit structured NNs on large-scale datasets [1]. Select the best model via Cross-Validation (CV).

The Gap: CV tends to favor more flexible models and can overlook the true underlying mechanism.

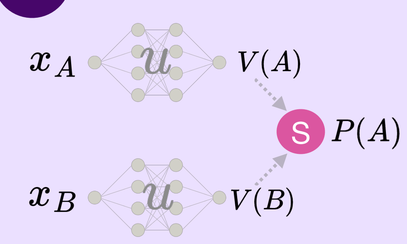
Our Approach: Apply **Bayesian Model Comparison** to select models based on **Marginal Likelihood**: balances goodness-of-fit against an explicit penalty for model flexibility.

1. Task

Human data: **Choices13k** [1]
~14k Participants
~13k Problems

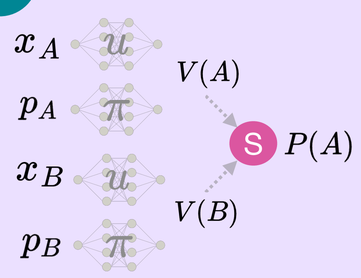
Option A	Option B
16 w. certainty	1 w. prob. 0.6 44 w. Prob. 0.1 48 w. Prob. 0.1 50 w. Prob. 0.2

EU Expected Utility



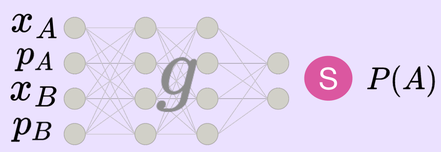
$$V(A) = \sum_{i \in A} u(x_i) p_i$$

PT Prospect Theory



$$V(A) = \sum_{i \in A} u(x_i)\pi(p_i)$$

CD Context-Dependent



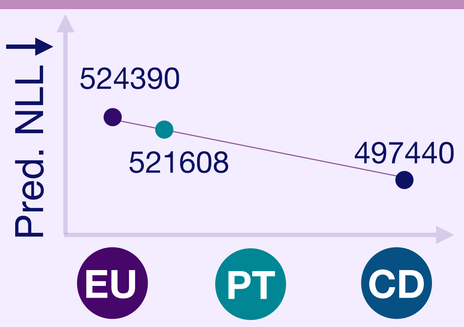
$$P(A) = g(x_A, p_A, x_B, p_B)$$

S Softmax

2. Predictive Model Comparison

- Fit on training set
- Select best model based on test set (CV) performance

Limitation:
Favors more complex models.



3. Bayesian Model Comparison

In a Bayesian model we have:

$$p(\boldsymbol{\theta} \mid \mathcal{D}, \mathcal{M}) = \frac{p(\mathcal{D} \mid \boldsymbol{\theta}, \mathcal{M})\, p(\boldsymbol{\theta} \mid \mathcal{M})}{p(\mathcal{D} \mid \mathcal{M})}$$

Posterior · Likelihood · Prior · Marginal Likelihood

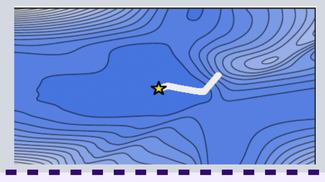
The **Marginal Likelihood** can be used for **Model selection**.

$$p(\mathcal{M} \mid \mathcal{D}) \propto p(\mathcal{D} \mid \mathcal{M})\, p(\mathcal{M})$$

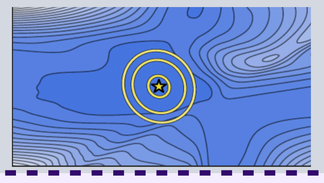
But the Integral is intractable. $p(\mathcal{D} \mid \mathcal{M}) = \int p(\mathcal{D} \mid \boldsymbol{\theta}, \mathcal{M}) p(\boldsymbol{\theta} \mid \mathcal{M}) d\boldsymbol{\theta}$

Laplace Approximation: Estimate the posterior by $\mathcal{N}(\boldsymbol{\theta}^*, A^{-1})$

(a) Find the **MAP** estimate (local maximum) of the posterior: $\boldsymbol{\theta}^*$



(b) Compute the **inverse Hessian** (local curvature) at MAP: $A^{-1}, A = -\nabla^2 p(\mathcal{D}, \boldsymbol{\theta}^* \mid \mathcal{M})$

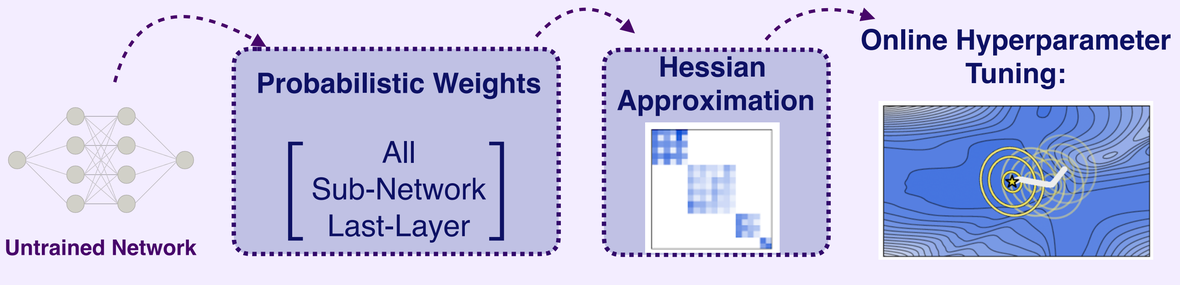


(b) is infeasible: scalable approximations of Hessian exist [2]

Advantage:

- No need for test data.
- **Occam's Razor**: Penalize complex models unless the data truly require them.
- Ground-truth model is recoverable IF: The true model is among the candidates and is identifiable, and the data is infinite.

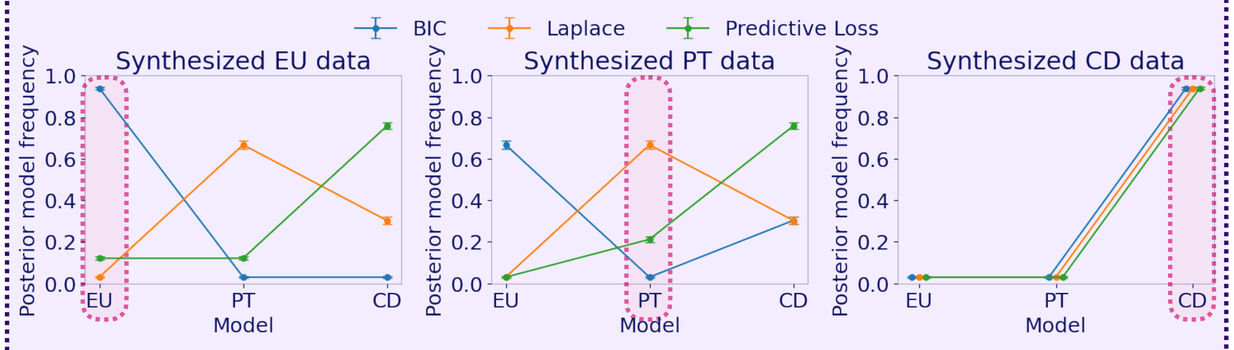
Laplace Package Pipeline [3]



4. Experiments

1. Model Recovery

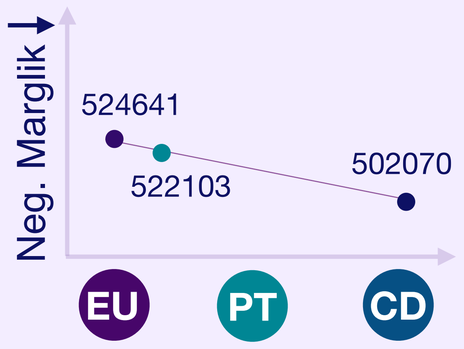
- Fit neural models of EU, PT, and CD on Choices13k.
- Generate synthetic decision choices from each model.
- Run different model selection procedures for 10 seeds.



2. Reproduce previous results with the Laplace approach

- Train on Choices13k
- We reproduce:
  - the *same order*
  - the *same accuracy*

as the predictive approach.



3. Dataset Generalization

- Train Dataset: Choices13k
- Test Dataset: CPC18
- Laplace generalizes better due to Occam's Razor.

CD
Predictive NLL = **42263**
>
Laplace NLL = **40875**

5. Discussion

- Bayesian Model Comparison can scale to Neural Networks.
- With this approach, we can discover theories of human decision-making more reliably.
- We can generalize to more domains and sequential models.
- *Open Question*: How can we resolve remaining mismatch in model recovery?

References

[1] Peterson, Joshua C., et al. "Using large-scale experiments and machine learning to discover theories of human decision-making." Science 372.6547 (2021): 1209-1214.
[2] Immer, Alexander, et al. "Scalable marginal likelihood estimation for model selection in deep learning." International Conference on Machine Learning. PMLR, 2021.
[3] Daxberger, Erik, et al. "Laplace redux-effortless bayesian deep learning." Advances in neural information processing systems 34 (2021): 20089-20103.