ICLR 2025

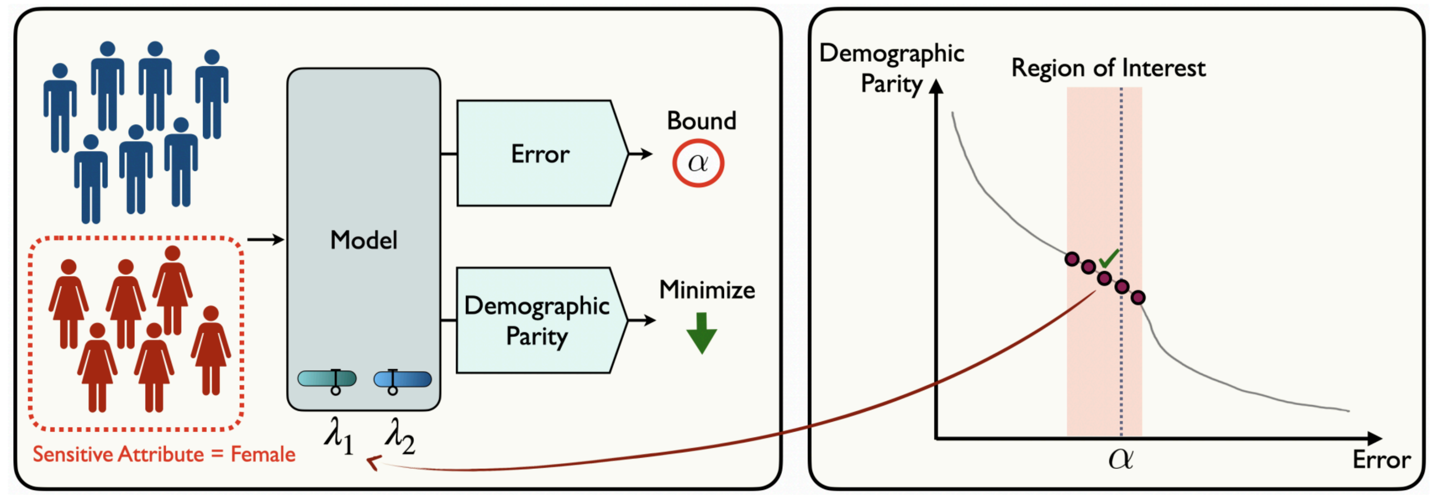
MIT CSAIL

School of Electrical Engineering & Computer Engineering, ECE
The Iby and Aladar Fleischman Faculty of Engineering
Tel Aviv University

# Risk Controlling Model Selection via Guided Bayesian Optimization

Bracha Laufer Goldshtein[1], Adam Fisch[2], Regina Barzilay[3], Tommi Jaakkola[3]

Tel Aviv University[1], Goole Deepmind[2], MIT CSAIL[3]




## Idea

Computationally and statistically efficient method for selecting model configurations that control multiple risks while minimizing an additional free objective function.

## Problem Formulation

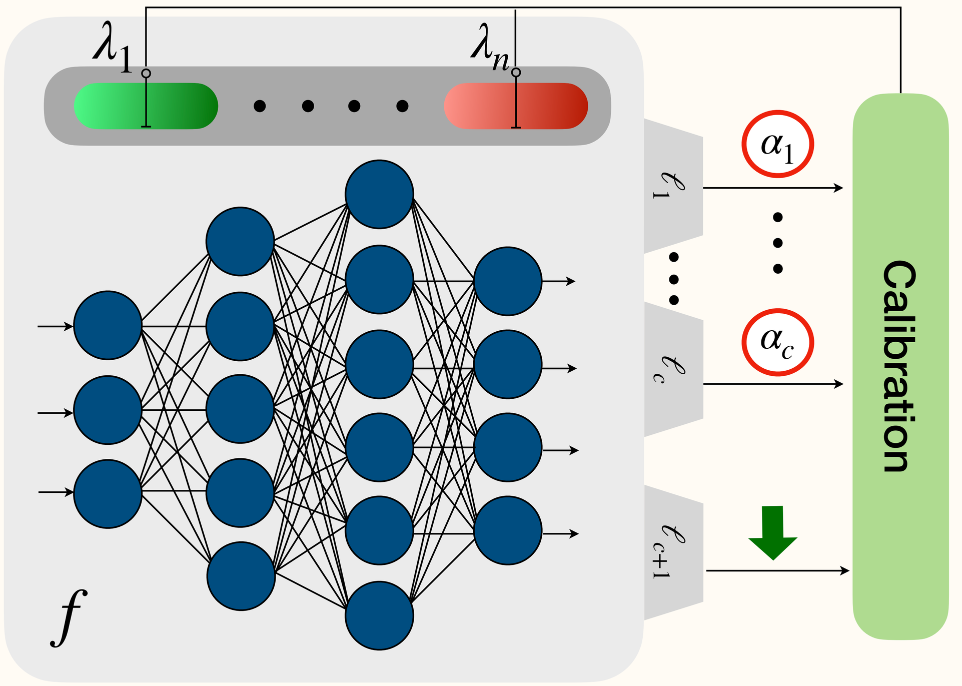
Model $f: X \times \Lambda \to Y$ configured by $n$ hyper-parameters $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n) \in \Lambda$

### User-defined objective functions

$$\ell_i(\boldsymbol{\lambda}) = \mathbb{E}\left[L_i(X, Y; \boldsymbol{\lambda})\right] \quad i \in \{1, \dots, c+1\}$$

**Control**: $\ell_1(\boldsymbol{\lambda}), \dots, \ell_c(\boldsymbol{\lambda})$ bound by $\alpha_1, \dots, \alpha_c$ with high probability

**Minimize**: $\ell_{c+1}(\boldsymbol{\lambda})$




Let $\mathcal{D}_{\text{cal}} = \{(X_i, Y_i)\}_{i=1}^m$ be an i.i.d. calibration set used for selecting a hyper-parameter combination $\hat{\boldsymbol{\lambda}}$.

Chosen combination is $(\alpha, \delta)$-risk controlling if:

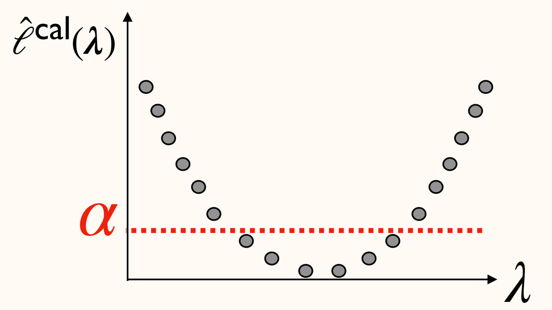
$$\mathbb{P}\left(\ell_i(\hat{\boldsymbol{\lambda}}) < \alpha_i\right) \geq 1 - \delta \quad \forall i \in \{1, \dots, c\}$$

**Goal**: Select $(\alpha, \delta)$-risk controlling configuration, which minimizes $\ell_{c+1}$.

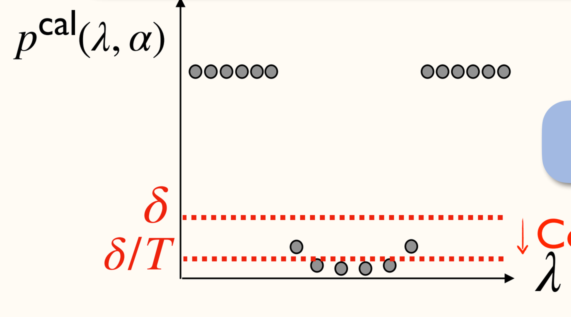
## Learn then Test [Angelopoulos et. al. ,2021]

Configuration selection as Hypothesis Testing - null hypothesis $H_\lambda : \ell(\lambda) > \alpha$

1. Compute empirical risk




2. Compute p-values



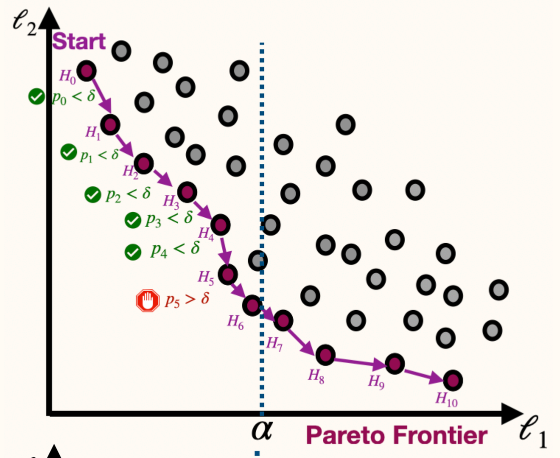

3. Test

Hyper-parameter dimension ↑
Statistical Power ↓
Computation ↑

## Pareto Testing [Laufer-Goldshtein et. al. ,2023]

### Key Steps

- Solve a multi-objective optimization problem
- Recover the Pareto front
- Perform fixed sequence testing over the front

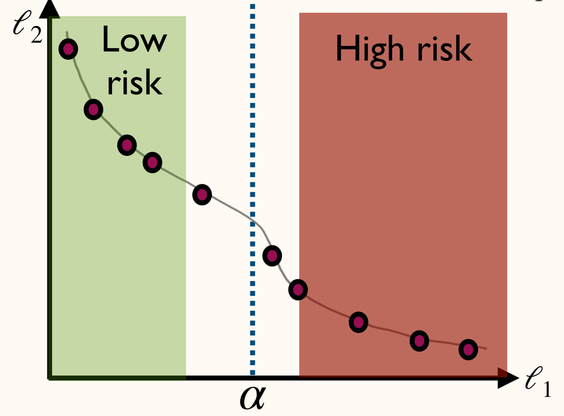



### Drawback

Pareto front includes irrelevant configurations

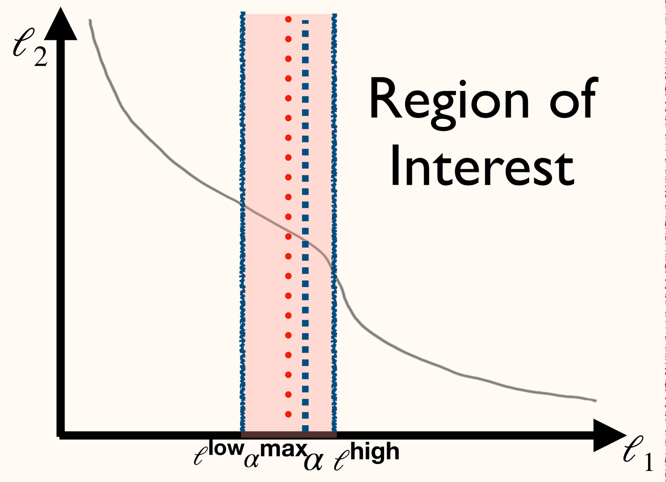
- High risk - unlikely to pass the test
- Low risk - inefficient w.r.t. the free objective

It might be too sparse near the limit




## Region of Interest

- Define a region of interest in the objective space.
- Focus on configurations that are both efficient and valid.
- Include values that are likely to correspond to $\alpha^{\max}$, the maximum value that can pass the test.




## GuideBO

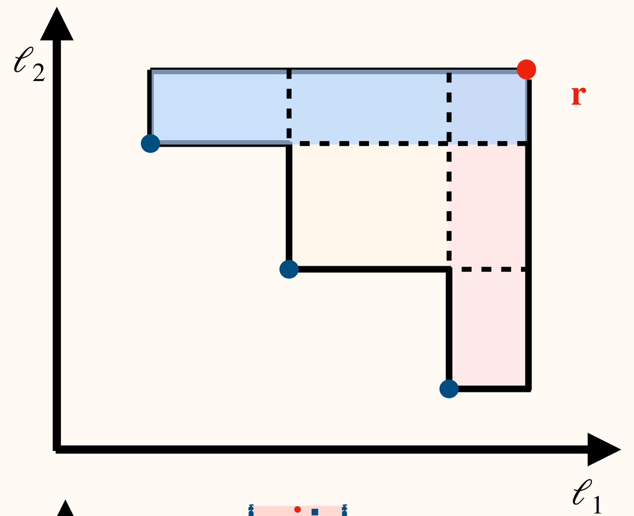
- Given an approximated Pareto front $\hat{\mathcal{P}}$, the hypervolume indicator:

$$HV(\hat{\mathcal{P}};\, \mathbf{r}) = \int_{\mathbb{R}^d} \mathbf{1}_{H(\hat{\mathcal{P}}, \mathbf{r})},$$

$$H(\hat{\mathcal{P}}; \mathbf{r}) = \{\mathbf{z} \in \mathbb{R}^d : \exists \mathbf{p} \in \hat{\mathcal{P}} : \mathbf{p} \prec \mathbf{z} \prec \mathbf{r}\}.$$

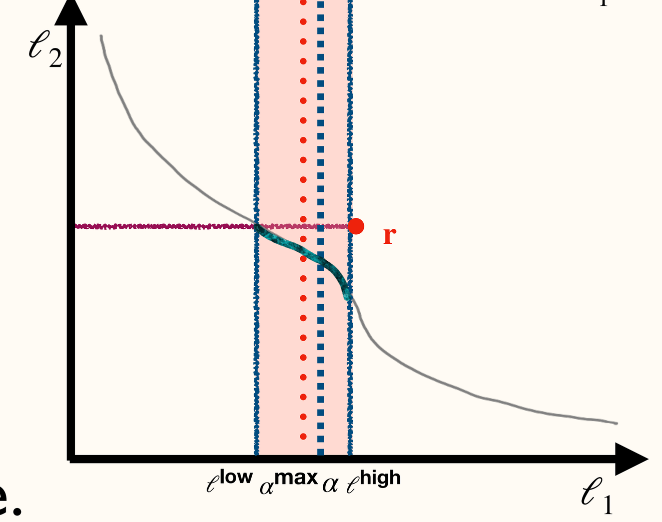
The Hypervolume improvement (HVI):

$$HVI(\boldsymbol{\ell}(\boldsymbol{\lambda}), \hat{\mathcal{P}}; \mathbf{r}) = HV(\boldsymbol{\ell}(\boldsymbol{\lambda}) \cup \hat{\mathcal{P}};\, \mathbf{r}) - HV(\hat{\mathcal{P}};\, \mathbf{r})$$




### Key Idea

- Modify HVI to capture the region of interest.
- Define $\mathbf{r} \in \mathbb{R}^{c+1}$ to enclose the desired region.
- Incorporate into a Bayesian optimization procedure.

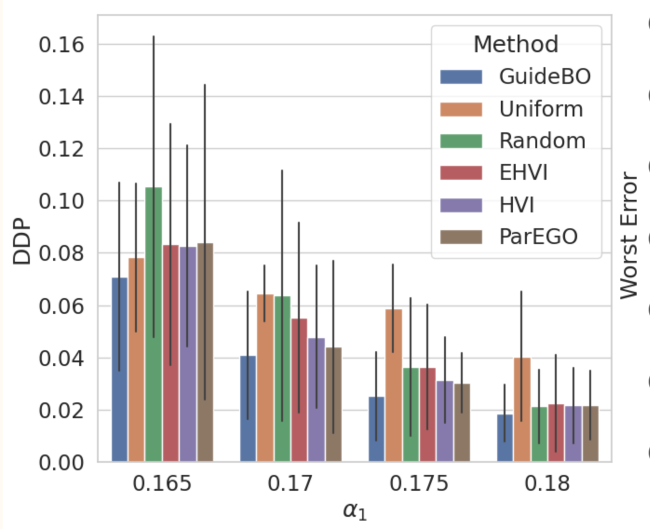



## Tasks

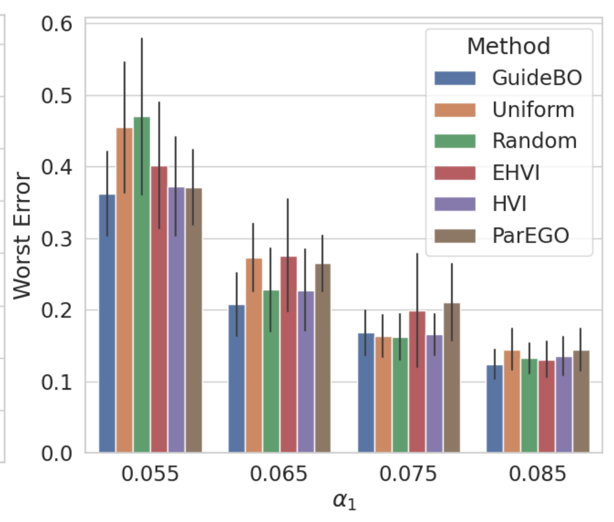
| | Constrain | Minimize | Hyperparameters | Dim |
|---|---|---|---|---|
| Fairness | Avg. error | Demographic Parity | Loss weights | 2 |
| Robustness | Avg. error | Worst error | Control data balance | 4 |
| Selective + Robustness | Avg. error & Miscoverage | Worst error | Control data balance and selection threshold | 5 |
| Early-Time Classification | Acc. Difference | Halt time | Stopping threshold per time | 8-12 |
| Pruning | Acc. Difference | Relative Cost | Degree of Sparsification | 3 |
| Generation | Reconstruction error | Disentanglement | KL regularization weight & latent space dimensionality | 2 |

## Results

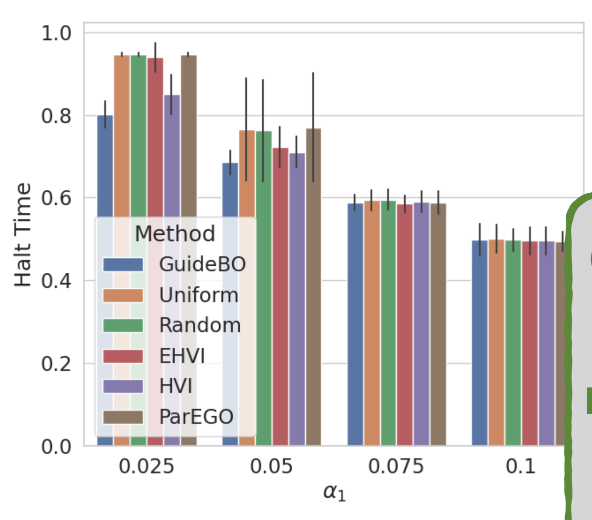
Fairness



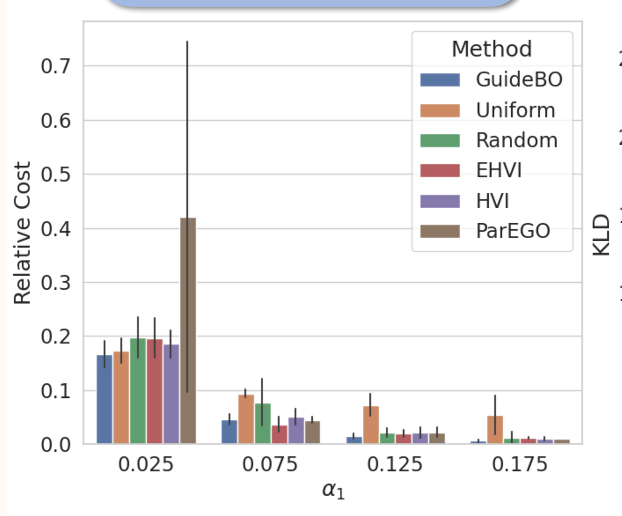

Robustness



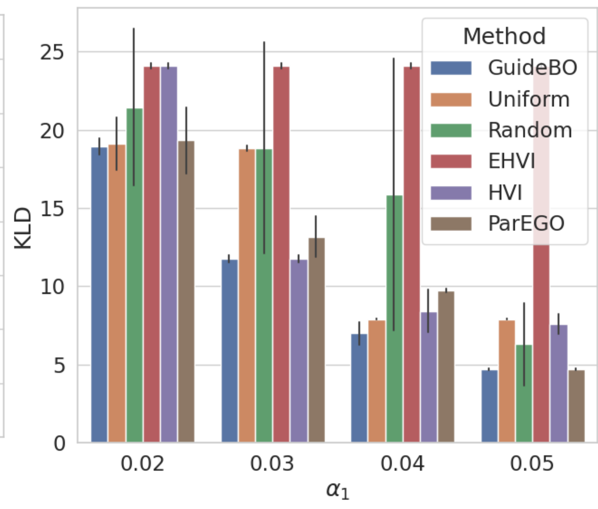

Early-Time Classification



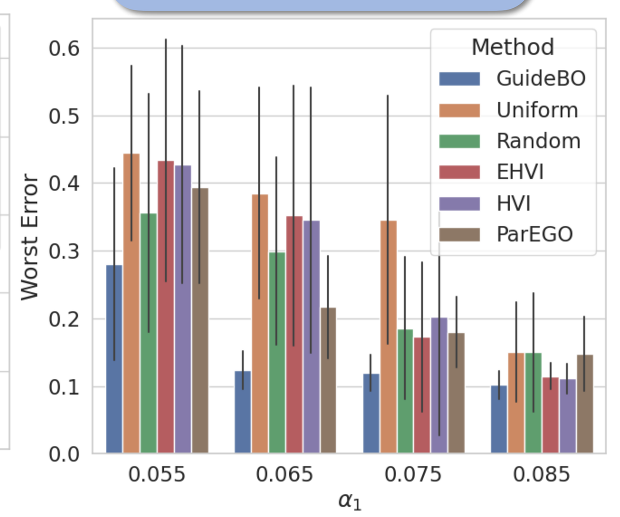

Pruning




Generation




Selective + Robustness




GuideBO better minimizes the free objective function