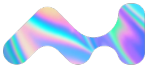


LiCE

Generating Likely Counterfactuals using Sum-Product Networks

Jiří Němeček, Tomáš Pevný, Jakub Mareček

nemecek.jiri@fel.cvut.cz
Artificial Intelligence Center
Faculty of Electrical Engineering
Czech Technical University in Prague

March 25, 2025

Counterfactual Explanations (CEs)

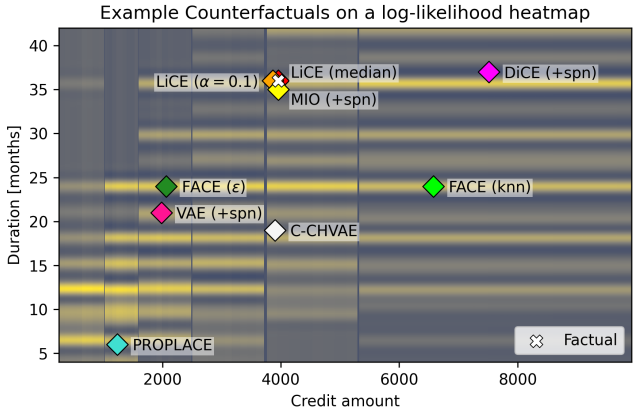
AI CENTER FEE CTU

- Minimal change of input to achieve desired output.
- Example: “You would get the loan if your income was $100,000 instead of $75,000.”
- Many desiderata for CEs [1].
  - **Plausibility:** The counterfactual should have a *high likelihood* w.r.t. the data distribution.
- A trade-off between plausibility and similarity [2].

LiCE

AI CENTER FEE CTU

- Successfuly optimize of plausibility and similarity at once.
  - Under further constraints (e.g., actionability, discrete data).



Plausibility in LiCE

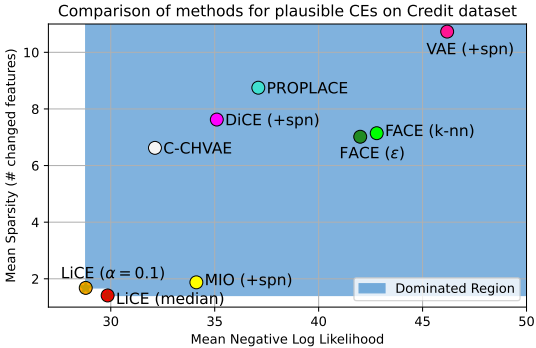
AI CENTER FEE CTU

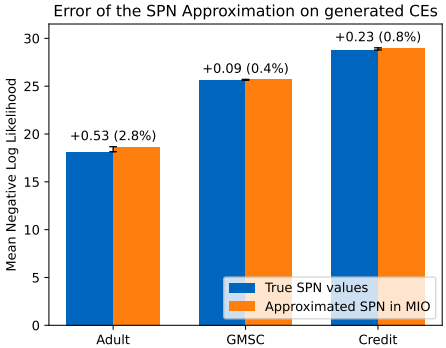
- Utilizing trained **Sum-Product Networks**:
  - DAG, computing sample likelihood.
  - Leaves are (univariate) **distributions**.
  - Inner nodes are
    - weighted **sums** (mixtures)
    - or **products** of children.
  - They model both discrete and real-valued random variables.
- We formulate them in **Mixed-Integer Linear Optimization**.
- And solve with a constraint on counterfactual's likelihood.
  - Or optimize a linear combination of similarity and plausibility.

Conclusion

AI CENTER FEE CTU

- LiCE **outperforms** other methods for plausible CEs on **likelihood and similarity**.
- We tightly approximate SPNs within MIO, opening up various future directions.





Thank you

AI CENTER FEE CTU

- Please contact me with any questions at:

  `contact@nemecekjiri.cz`

- LiCE implementation is available at:

  `github.com/Epanemu/LiCE`

References

Riccardo Guidotti.
Counterfactual explanations and how to find them: literature review and benchmarking.
Data Mining and Knowledge Discovery, April 2022.

Ulrike Kuhl, André Artelt, and Barbara Hammer.
Keep Your Friends Close and Your Counterfactuals Closer: Improved Learning From Closest Rather Than Plausible Counterfactual Explanations in an Abstract Setting.
In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, pages 2125–2137, New York, NY, USA, June 2022. Association for Computing Machinery.