EBERHARD KARLS UNIVERSITÄT TÜBINGEN

DSAR
DATA SCIENCE & ANALYTICS RESEARCH

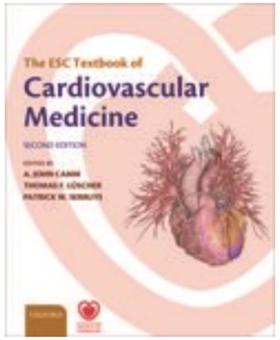
TOWARDS REALISTIC COUNTERFACTUAL EXPLANATIONS FOR TABULAR DATA

GJERGJI KASNECI

AI can be a game changer in medicine

OXFORD ACADEMIC

Subject ▾ Journals Books



The ESC Textbook of Cardiovascular Medicine (2 edn)

A. John Camm (ed.) et al.

Contents

▸ Front Matter

1 The Cardiovascular History and Physical Examination

CHAPTER

32 Sports and Heart Disease

Get access ›

Domenico Corrado, Cristina Basso, Antonio Pelliccia, Gaetano Thiene

https://doi.org/10.1093/med/9780199566990.003.032 Pages 1215–1238

Published: August 2009

Cite Permissions Share ▾

Extract

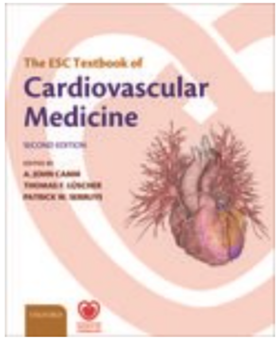
Summary

Sports activity is recommended by the medical community because it improves fitness and reduces cardiovascular morbidity and mortality. However, physical exercise may precipitate acute fatalities in both adults and young competitive athletes with concealed heart diseases.

AI can be a game changer in medicine

OXFORD ACADEMIC

Subject ▾ Journals Books



The ESC Textbook of Cardiovascular Medicine (2 edn)

A. John Camm (ed.) et al.

Contents

- Front Matter
- 1 The Cardiovascular History and Physical Examination

nature medicine

Explore content ⌄ About the journal ⌄ Publish with us ⌄ Subscribe

nature › nature medicine › research highlights › article

RESEARCH HIGHLIGHT | 14 April 2022

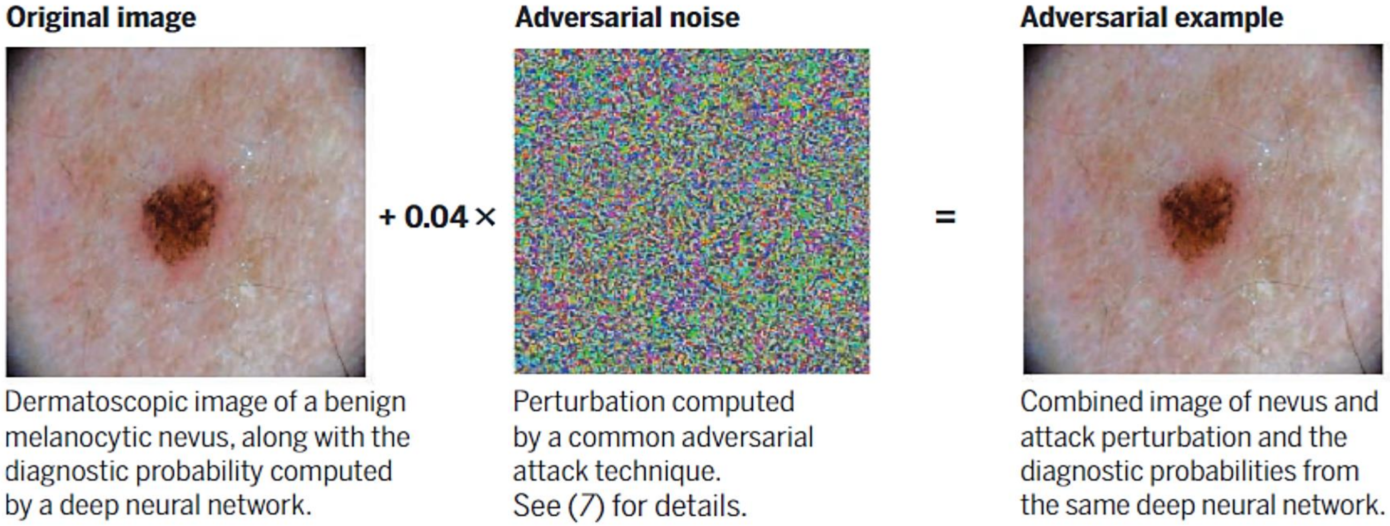
Using AI to predict future cardiac arrest

A deep-learning model predicts the likelihood of, and time to, sudden cardiac death in patients with heart disease — providing an opportunity for clinical intervention.

Karen O'Leary

... but must handle uncertainties and must be explainable

Example: Using AI to recognize cancerous moles [S. G. Finlayson et al., *Adversarial attacks on medical machine learning*, Science 2019]

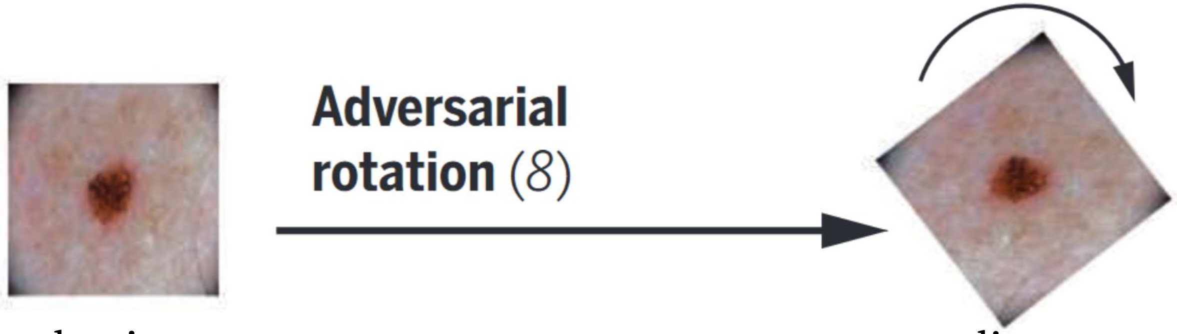


99% confidence benign

100% confidence malignant

“If dermatologists were to get reimbursed only for removing a mole by insurance companies if an AI agreed that it was malignant, there could be an incentive to alter borderline cases to ensure payment for more procedures”
– Samuel G. Finlayson.

Conclusion: It isn’t always clear what factors AI uses for prediction, which can cause problems when the tools are used in the real world.

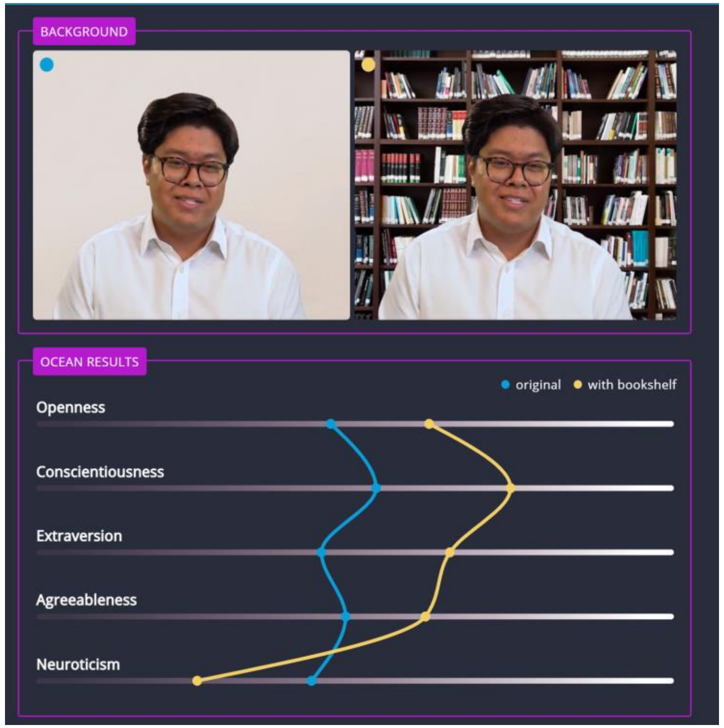


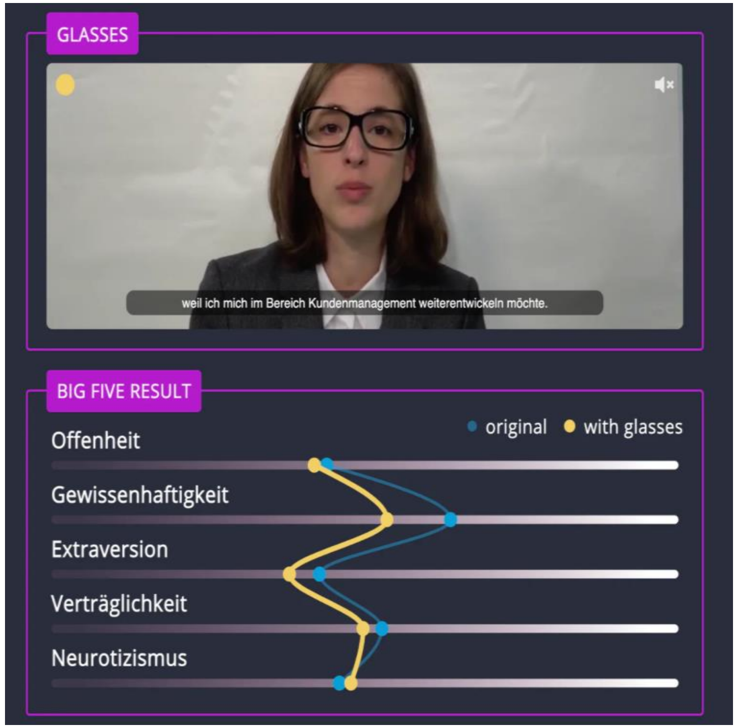
benign

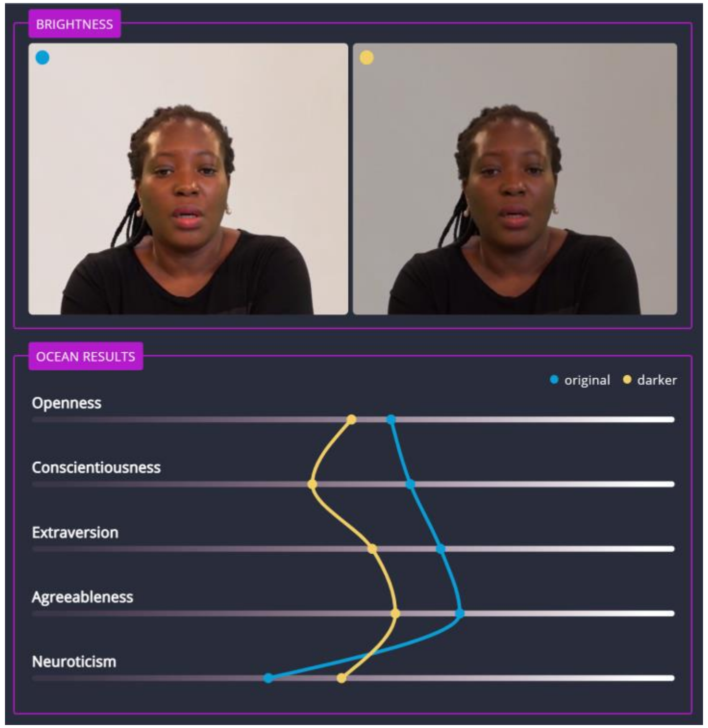
malignant

This could lead to many cases of misclassification in day-to-day practice

AI has influence on personal lives







Source: https://web.br.de/interaktiv/ki-bewerbung/en/?utm_keyword=referral_input

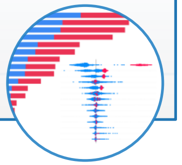
- **BR24:** How can it be that factors not related to facial expressions or gestures play a role in the evaluation of personality traits?
- **Retorio.com**: *As in a normal job interview, such factors are also included in the assessment. All of this is done without being asked, without any pressure of the kind that can arise in an interview situation.*

- **Q1: What are the main factors that led to the result?**
- **Q2: What can I do to achieve the desired outcome?**

Algorithmic recourse and involved explanation strategies

► **Algorithmic Recourse** aims to **explain an automated decision** and **suggest actionable changes to achieve favorable outcomes for the end user** (emphasizing feasibility)

Feature Attribution



- **Central question:** Which **inputs** are **responsible** for the produced output and what is the **order of importance** or what are the **interactions** between those inputs?
- **Popular approaches:** SHAP (SHapley Additive exPlanations) framework by [5], LIME [6], and others [7], [8]).

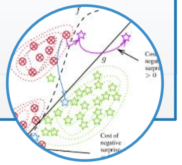
Explanations by example

- **Central question:** What are the **instances that lead to specific results** and how are they distributed?
- Select **particular instances** from the dataset or exploit particular instances **provided by human experts** to explain the behavior of an ML model or explain the underlying distribution [9, 10].

Contrastive Explanations

- **Central question:** Why a model predicted the actual output and not another **close alternative**?
- Grounded in **cognitive psychology** our explanation-seeking behavior is considered as rather **contrastive** [11] (it is argued that such explanations are **easy to process** from a cognitive perspective).

Counterfactual Explanations



- **Central question:** What can the user do to **achieve a desired outcome**?
- These approaches (see [12]) provide model explanations by **highlighting important features** and **suggesting actionable feature changes**, e.g., paying off small loans in time, to achieve favorable outcomes in the future.

[5] Lundberg & Lee. *A unified approach to interpreting model predictions*. NeurIPS 2017.
[6] Ribeiro, Singh & Guestrin. *Why should i trust you?: Explaining the predictions of any classifier*. KDD 2016.
[7] Shrikumar, Greenside & Kundaje. *Learning Important Features Through Propagating Activation Differences*. ICML 2018.
[8] Bach et al. *On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation*. PloS one 10, 7 (2015).
[9] Gade et al. *Explainable AI in industry*. KDD 2019.
[10] Mittelstadt, Russell & Wachter. 2019. *Explaining explanations in AI*. In FAT* 2019.
[11] Peter Lipton. *Inference to the best explanation*. Routledge 2003.
[12] Wachter, Mittelstadt & Russell. *Counterfactual explanations without opening the black box: automated decisions and the GDPR*. Harvard Journal of Law & Technology 31, 2 (2018).

Regulatory challenges for recourse on tabular data

Algorithmic recourse should respect GDPR principles!?

Data protection and minimality	Right for explanation	Right to be forgotten
Use only relevant and as few data points as possible	Need for human-interpretable explanations	Data deletion applies also to the models

What are counterfactual explanations and why do we need them?

Given a classifier $f: \mathbb{R}^n \rightarrow [0,1]$ and fixed threshold $t \in [0,1]$
(1) Explain how the *factual input* $\mathbf{x}^F$ influences $f(\mathbf{x}^F)$
(2) If $f\left(\mathbf{x}^F\right) \leq t$, find *counterfactual input* $\mathbf{x}^{CF} \sim \mathbf{x}^F$ that fulfills certain feasibility constraints and $f\left(\mathbf{x}^{CF}\right) > t$

▶ Practically viable CEs should be

- **Realistic** (i.e., the suggestions are realistically achievable)
- **Robust** (e.g., noise or small changes to the data distribution or to the classifier, e.g., through recalibration, should not invalidate the CE)
- **Simple** (small set of easy-to-implement or easy-to-attain suggestions increases the probability of success)

▶ CEs are a powerful mean towards

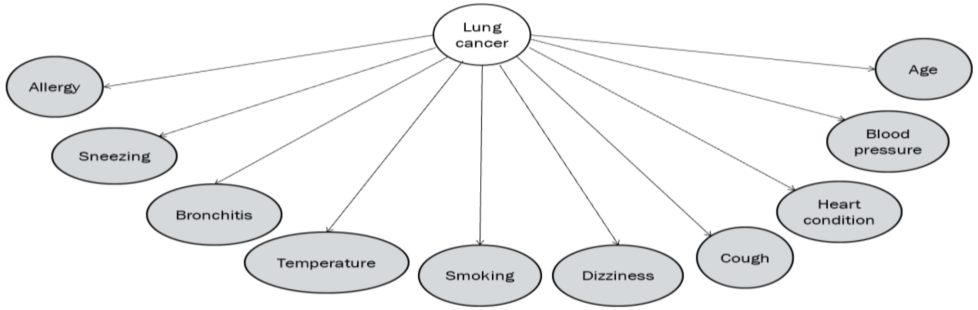
- **Context-wise understanding** through individualized actionable suggestions
- **Empowerment** through education and step-wise personal improvements
- **Perceived fairness** is increased when individual improvements are possible (in constructive interaction with the system)

CEs by model design vs. CEs for black-box models

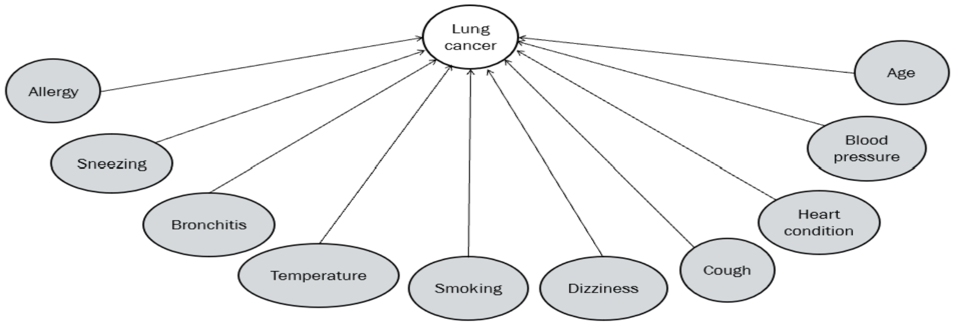
Generative

Discriminative

Simple

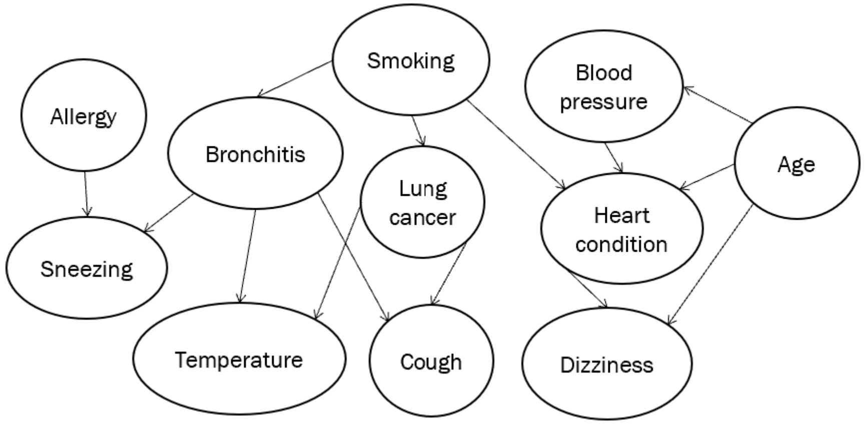


$$p(\mathbf{x}, y) = p(\mathbf{x}|y)p(y) = p(y)\prod_{i=1}^{k} p(x_i|y)$$

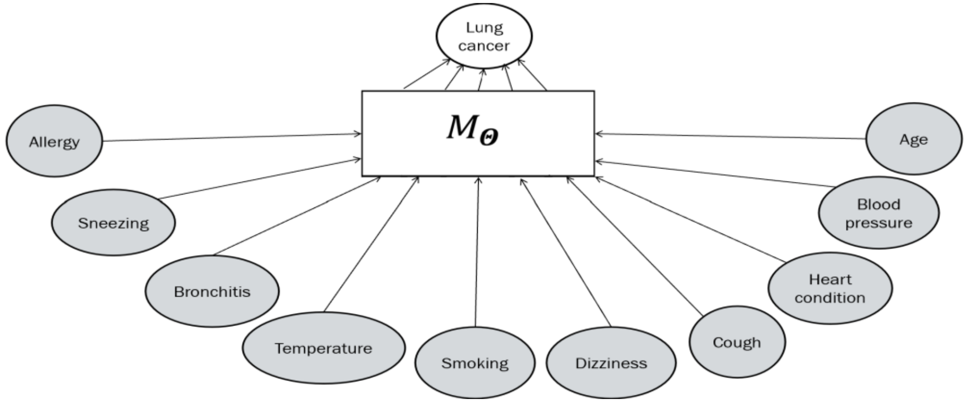


$$p(y|\mathbf{x};\mathbf{w}) \approx y \cdot \sigma\left(w_0 + \sum_{i=1..k} w_i x_i\right) + (1-y)\left(1 - \sigma\left(w_0 + \sum_{i=1..k} w_i x_i\right)\right)$$

Intricate

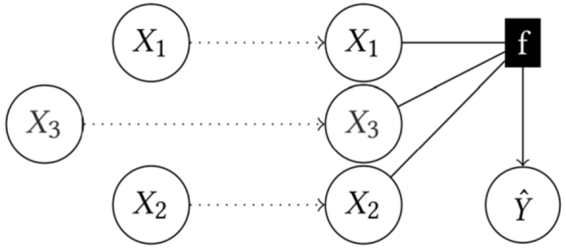


$$p(x_1, x_2, \ldots, x_{k+1}) = p(x_1)p(x_2|x_1)p(x_3|x_2, x_1)p(x_4|x_2)\ \ldots$$

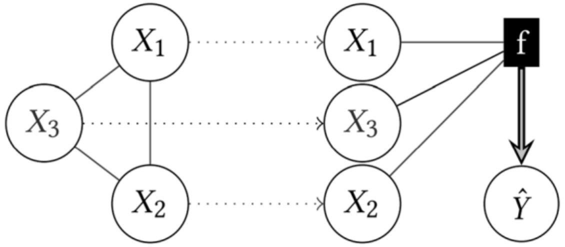


$$p(y|\mathbf{x};\boldsymbol{\Theta}) \approx y \cdot M_{\boldsymbol{\Theta}}(\mathbf{x}) + (1-y)\big(1 - M_{\boldsymbol{\Theta}}(\mathbf{x})\big)$$

CEs for black-box models



X_1 e.g., blood pressure
X_2 e.g., age
X_3 e.g., heart condition



► Independence-based approaches

- Impose **independence assumption** on input features
- **Combinatorial solvers** generate recourse suggestions in the presence of feasibility constraints, e.g. [13,14,15,16]
- **Neighborhood search** [17,18] with decision trees, random search, and SVMs aim to fulfill fairness constraints
- Gradient-based **optimization** to find **low-cost CEs** given multiple constraints, such as feasibility and diversity [10,19,20]
- Shortcoming: neglecting input dependencies leads to **overly optimistic intervention costs in practice**

[13] *Karimi et al. Model-Agnostic CE for Consequential Decisions. AISTATS 2020.*
[14] *Rawal & Lakkaraju. Beyond Individualized Recourse: Interpretable and Interactive Summaries of Actionable Recourses. NeurIPS 2020.*
[15] *Russell. Efficient Search for Diverse Coherent Explanations. FAT* 2019.*
[16] *Ustun, Spangher & Liu. Actionable recourse in linear classification. FAT* 2019.*
[17] *Lash et al. Generalized inverse classification. SIAM 2017.*
[18] *Tolomei et al. Interpretable predictions of tree-based ensembles via actionable feature tweaking. KDD 2017.*
[19] *Dhurandhar et al. Explanations based on the Missing: Towards Contrastive Explanations with Pertinent Negatives. NeurIPS 2018.*
[10] *Mittelstadt, Russell & Wachter. Explaining explanations in AI. FAT* 2019.*
[20] *Mothilal, Sharma & Tan. Explaining ML Classifiers through Diverse CEs. FAT* 2020.*

► Dependency-based approaches

- **Bridge the gap** between the **strong independence assumption** and the **strong causal assumption**, e.g. [21,22,23,24,25]
- Main idea: exploit **factors of variation in lower-dimensional latent space** to capture input dependencies [26,27]
- Feasibility **constraints can be encoded into the CE model** [23] and classification **uncertainty can be modeled**, e.g., CLUE [28]
- Shortcoming: **data handling for CE generation** is a challenge

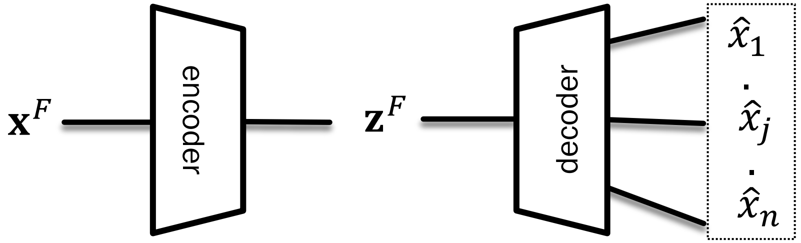
[21] *Downs et al. CRUDS: Counterfactual Recourse Using Disentangled Subspaces. ICML WHI 2020.*
[22] *Joshi et al. Towards Realistic Individual Recourse and Actionable Explanations in Black-Box Decision Making Systems. arXiv preprint arXiv:1907.09615(2019).*
[23] *Mahajan, Tan, & Sharma. Preserving causal constraints in counterfactual explanations for machine learning classifiers. arXiv preprint arXiv:1912.03277(2019).*
[24] *Pawelczyk, Broelemann & Kasneci. Learning Model-Agnostic CEs for Tabular Data. WWW 2020.*
[25] *Pawelczyk, Broelemann & Kasneci. On CEs under Predictive Multiplicity. UAI 2020. PMLR.*
[26] *Kingma & Welling. Auto-encoding variational bayes. ICLR 2013.*
[27] *Nazabal et al. Handling incomplete heterogeneous data using vaes. PR 2020.*
[28] *Antorán, et al. Getting a CLUE: A Method for Explaining Uncertainty Estimates. ICLR 2021.*

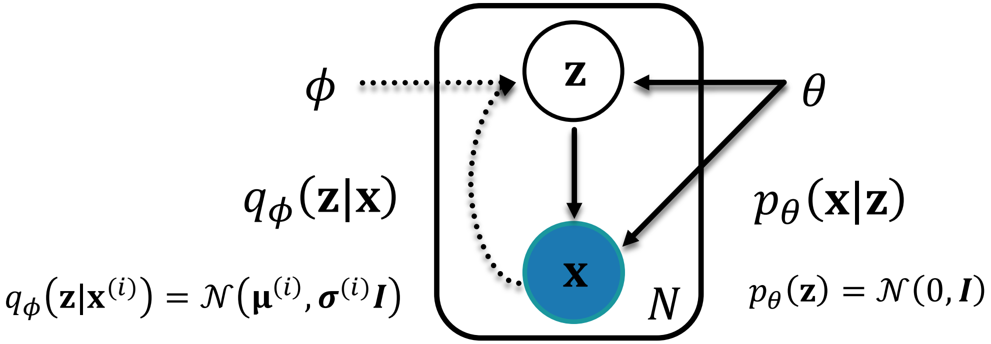
Variational Autoencoder models for learning CEs for tabular data

- Model for data generating process

$$\mathbf{x}^F = g(\mathbf{z}^F),\ z_j^F \perp z_k^F \text{ for } j \neq k$$

- Need not assume that $x_j^F \perp x_k^F$





- Explanation Model

$$\delta_{\mathbf{z}}^* = \arg\min_{\delta_{\mathbf{z}}} \|\mathbf{x}^F - \overbrace{g(\mathbf{z}^F + \delta_{\mathbf{z}})}^{\text{Possible } \mathbf{x}^{CF}}\|$$

such that $f\left(g(\mathbf{z}^F + \delta_{\mathbf{z}})\right) > t$

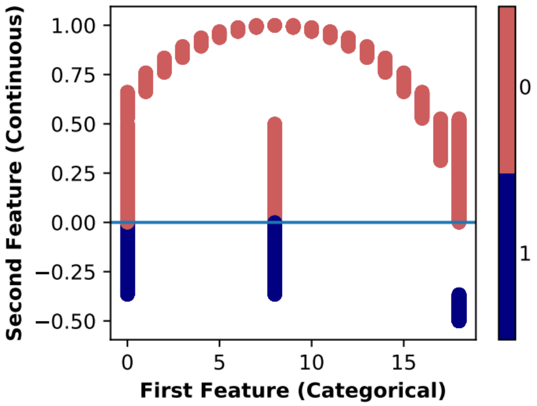
where $f(\mathbf{x}) \in [0,1]$ as before.

- $\|\cdot\|_1$ leads to sparse explanations

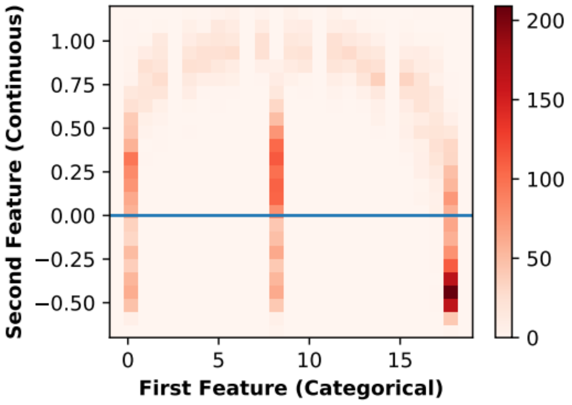
$$\mathcal{L}(\phi, \theta, \mathbf{x}) = D_{KL}\left(q_\phi(\mathbf{z}|\mathbf{x}) \parallel p_\theta(\mathbf{z})\right) \overbrace{- \frac{1}{k} \sum_{j=1..k} \log p_\theta(\mathbf{x}|\mathbf{z}^{(j)})}^{\text{Expected reconstruction error with different losses for different data types}}$$

[24] Pawelczyk, Broelemann & Kasneci. 2020. *Learning Model-Agnostic CEs for Tabular Data*. WWW 2020.

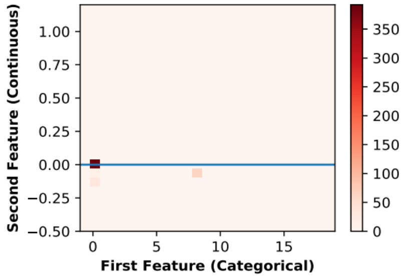
Variational Autoencoder models for learning CEs for tabular data



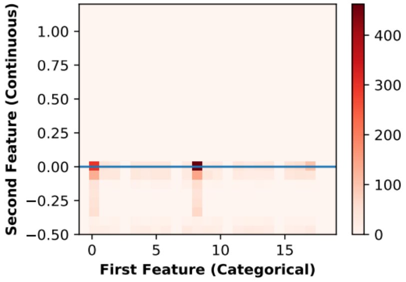
Data generating process



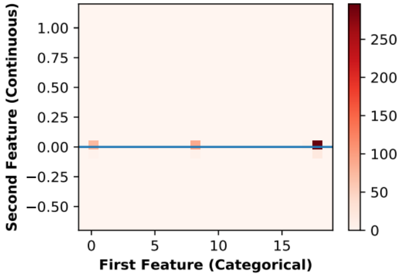
Reconstructed data (by our model)



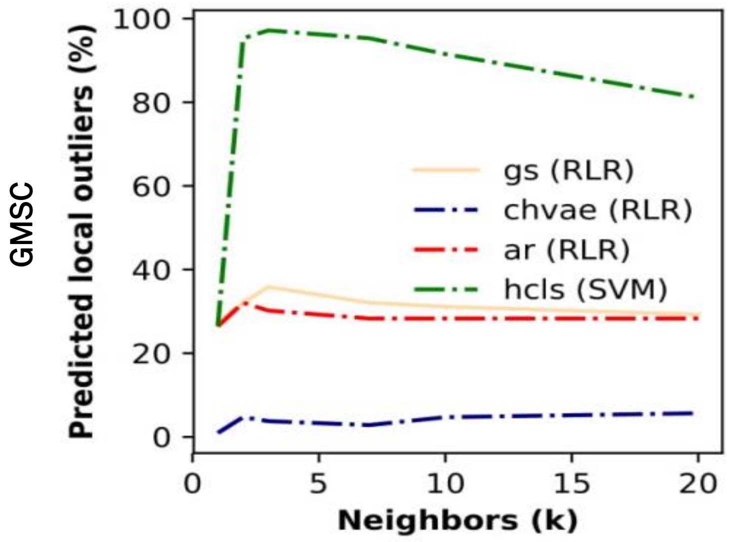
CEs by AR [16]

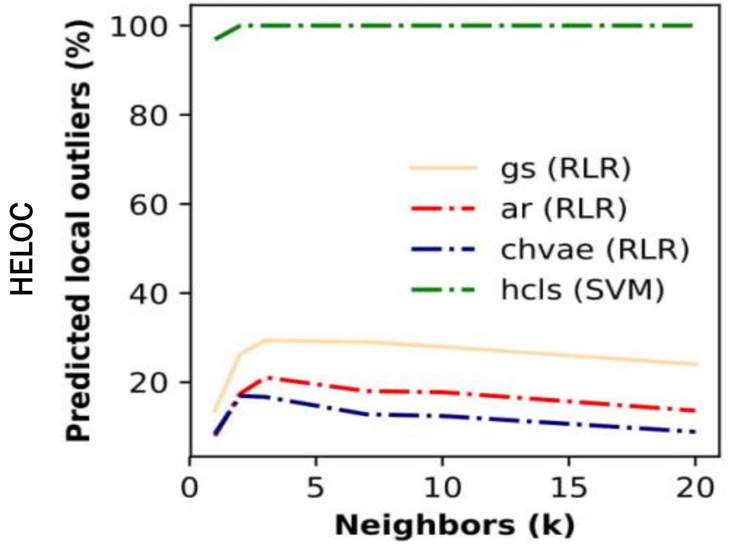


CEs by GS [31]



CEs by CCHVAE [24] (ours)





[24] Pawelczyk, Broelemann & Kasneci. *Learning Model-Agnostic CEs for Tabular Data*. WWW 2020.
[16] Ustun, Spangher & Liu. *Actionable recourse in linear classification*. FAT* 2019.
[31] Laugel et al. *Inverse Classification for Comparison-based Interpretability in Machine Learning*. arXiv preprint arXiv:1712.08443 (2017).
GMSC: https://www.kaggle.com/brycecf/give-me-some-credit-dataset
HELOC: https://community.fico.com/s/explainable-machine-learning-challenge

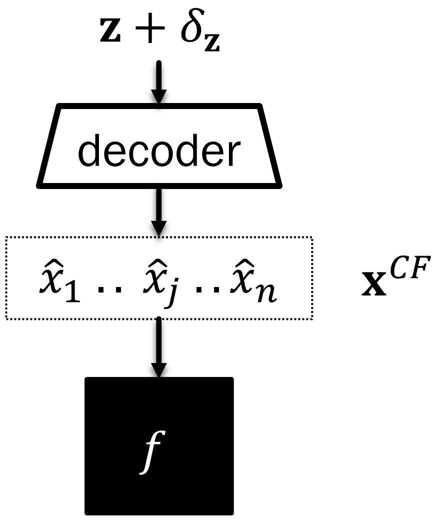
Variational Autoencoder models faithfully reflect recourse costs

Let $\mathbf{x} \in \mathbb{R}^n$ with $g(\mathbf{z}) = \mathbf{x}$ and $f(\mathbf{x}) \leq t$,
and let $\hat{\mathbf{x}} = g(\mathbf{z} + \delta_{\mathbf{z}})$ with $(\hat{\mathbf{x}}) > t$

Let $J_{\mathbf{z}}^{(\mathbf{x},g)} := \frac{\partial g(\mathbf{z})}{\partial \mathbf{z}}\Big|_{\mathbf{z}}$ (controls to what extent the dimensions of $\mathbf{x}$ are affected by latent actions $\delta_{\mathbf{z}}$)

Theorem: $\|\delta_{\mathbf{x}}\|^2 = \|\mathbf{x} - \hat{\mathbf{x}}\|^2 = \|g(\mathbf{z}) - g(\mathbf{z} + \delta_{\mathbf{z}})\|^2$ (recourse cost)

$$\approx \left\| g(\mathbf{z}) - \left(g(\mathbf{z}) + J_{\mathbf{z}}^{(\mathbf{x},g)} \delta_{\mathbf{z}} \right) \right\|^2 = \delta_{\mathbf{z}}^T \left({J_{\mathbf{z}}^{(\mathbf{x},g)}}^T J_{\mathbf{z}}^{(\mathbf{x},g)} \right) \delta_{\mathbf{z}}$$



Variational Autoencoder models avoid low-density recourse

Corollary: For $g(\mathbf{z}) = \boldsymbol{\mu}(\mathbf{z}) + \boldsymbol{\sigma}(\mathbf{z}) \odot \epsilon$ where $\epsilon \sim \mathcal{N}(0, \boldsymbol{I})$ we have

$$\mathbb{E}\left(\|\delta_{\mathbf{x}}\|^2 | \delta_{\mathbf{z}}\right) \approx \delta_{\mathbf{z}}^T \left({J_{\mathbf{z}}^{(\mathbf{x},\boldsymbol{\mu})}}^T J_{\mathbf{z}}^{(\mathbf{x},\boldsymbol{\mu})} + {J_{\mathbf{z}}^{(\mathbf{x},\boldsymbol{\sigma})}}^T J_{\mathbf{z}}^{(\mathbf{x},\boldsymbol{\sigma})} \right) \delta_{\mathbf{z}}$$

The induced expected cost of recourse will be large, if the generator (i.e., decoder) is uncertain in regions of the latent space

Main advantages of this model

(1) Reconstructions adhere to input correlations
(2) Reconstructions happen in dense regions (with data support)

[24] Pawelczyk, Broelemann & Kasneci. *Learning Model-Agnostic CEs for Tabular Data*. WWW 2020.
[29] Pawelczyk et al. *Algorithmic Recourse for Correlated Inputs with Independent Mechanisms*. Under review

Learning CEs for tabular data by independent mechanisms

- Model for data generating process

$$\mathbf{x}^F = g(\mathbf{v}^F, \mathbf{x}_J^F), \ \mathbf{v}^F \perp \mathbf{x}_J^F \ \text{ for } J \subset \{1, \dots, n\}$$

- Do recourse intervention on $\mathbf{x}_J^F$

- Explanation Model

$$\delta_{\mathbf{z}}^* = \arg\min_{\delta_{\mathbf{x}_J}} \left\| \mathbf{x}^F - g\left(\mathbf{v}^F, \mathbf{x}_J^F + \delta_{\mathbf{x}_J}\right) \right\|$$

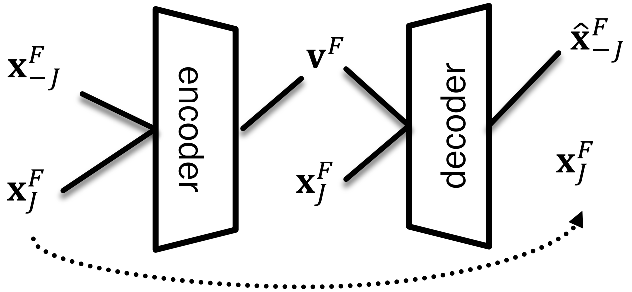
$$\text{such that } f\left(g\left(\mathbf{v}^F, \mathbf{x}_J^F + \delta_{\mathbf{x}_J}\right)\right) > t$$

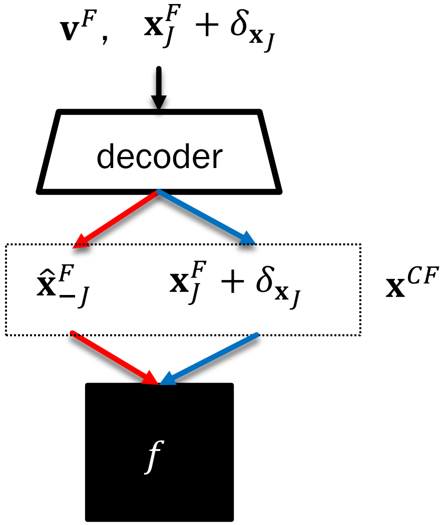
where $f(\mathbf{x}) \in [0,1]$ and t as before.

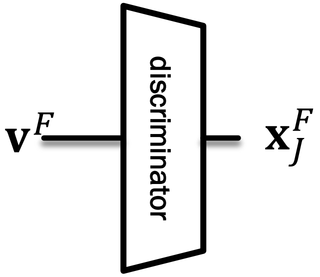
- $\|\cdot\|_1$ for rather sparse explanations

- $\|\delta_{\mathbf{x}}\|^2 = \|\mathbf{x} - \hat{\mathbf{x}}\|^2$

$$= \underbrace{\delta_{\mathbf{x}_J}^T \left(J_{\mathbf{z}}^{(\mathbf{x}_{-J})^T} J_{\mathbf{z}}^{(\mathbf{x}_{-J})} \right) \delta_{\mathbf{x}_J}}_{\text{indirect costs}} + \underbrace{\delta_{\mathbf{x}_J}^T \delta_{\mathbf{x}_J}}_{\text{direct costs}}$$





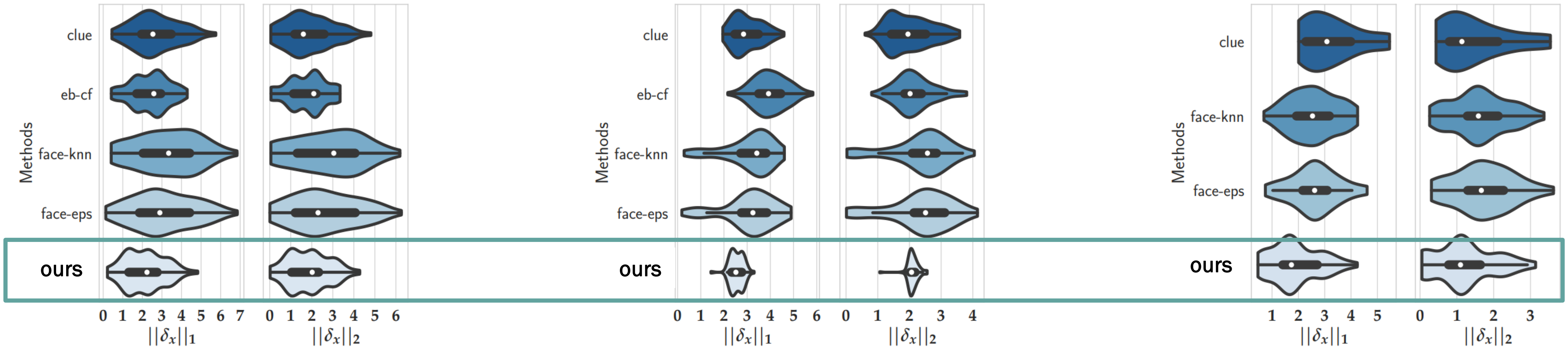


End-to-end Autoencoder training with adv. Regularizer:

$$\mathcal{L}(\Theta, \mathbf{x}) = \left\| dec\left(\mathbf{v}^F, \mathbf{x}_J^F\right) - \mathbf{x}_{-J}^F \right\| - \lambda \left\| disc\left(\mathbf{v}^F\right) - \mathbf{x}_J^F \right\|$$

[29] Pawelczyk et al. *Algorithmic Recourse for Correlated Inputs with Independent Mechanisms*. Under review

Learning CEs for tabular data by independent mechanisms



(a) ANN – Adult (b) ANN – Give Me Some Credit (c) ANN – COMPAS

Competitors

- EB-CF [23]
- CLUE [28]
- FACE [32]

Summary

- Median individual costs are lower
- Less cost variability
- Worst-case CE less costly

[23] Mahajan, Tan, & Sharma. *Preserving causal constraints in counterfactual explanations for machine learning classifiers*. arXiv preprint arXiv:1912.03277(2019).
[28] Antorán, et al. *Getting a CLUE: A Method for Explaining Uncertainty Estimates*. ICLR 2021.
[20] Pawelczyk et al. *Algorithmic Recourse for Correlated Inputs with Independent Mechanisms*.
[32] Poyiadzi et al. *Face: Feasible and actionable counterfactual explanations*. AAAI-AIES 2020.

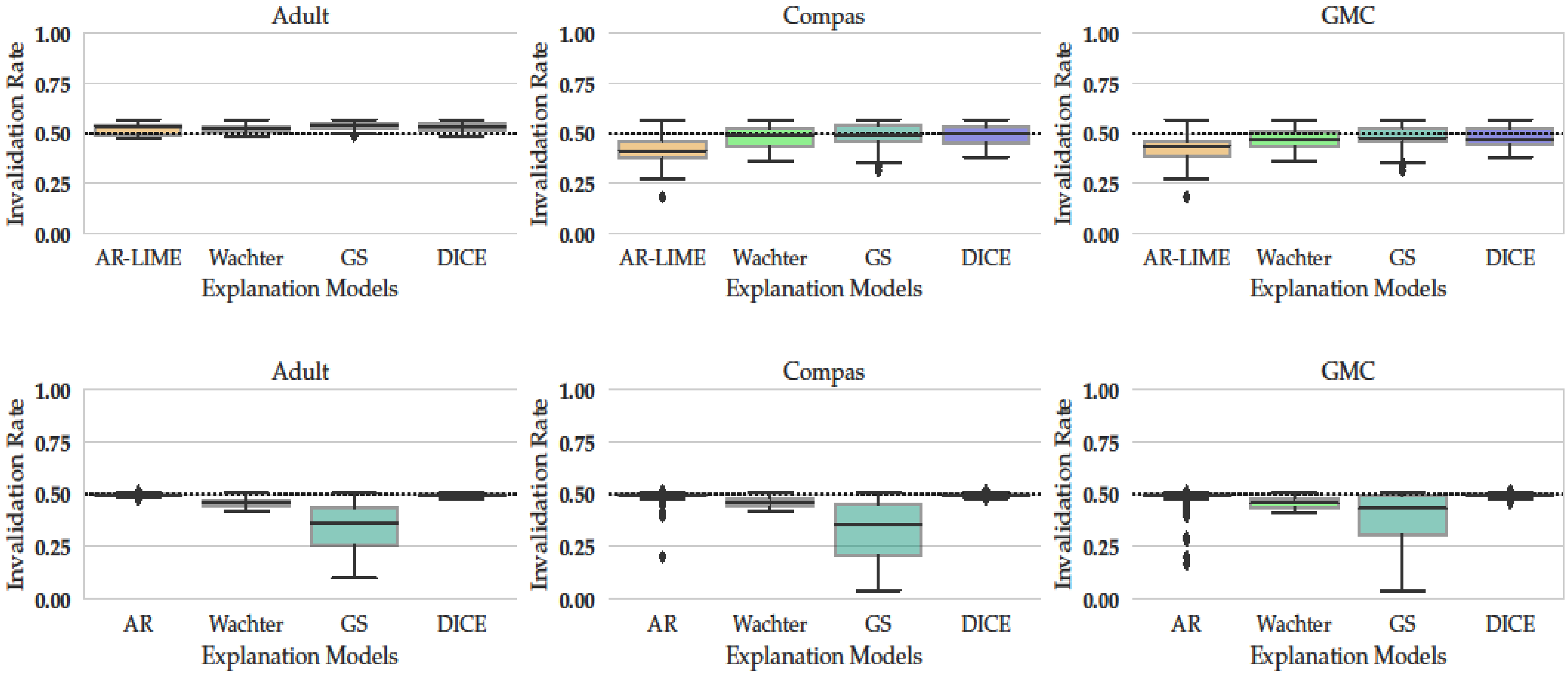
Invalidation rates for SOTA methods

Study robustness to perturbations of prescribed recourses (under review)

Perturb recourses by adding Gaussian RV with mean 0 and variance 0.01

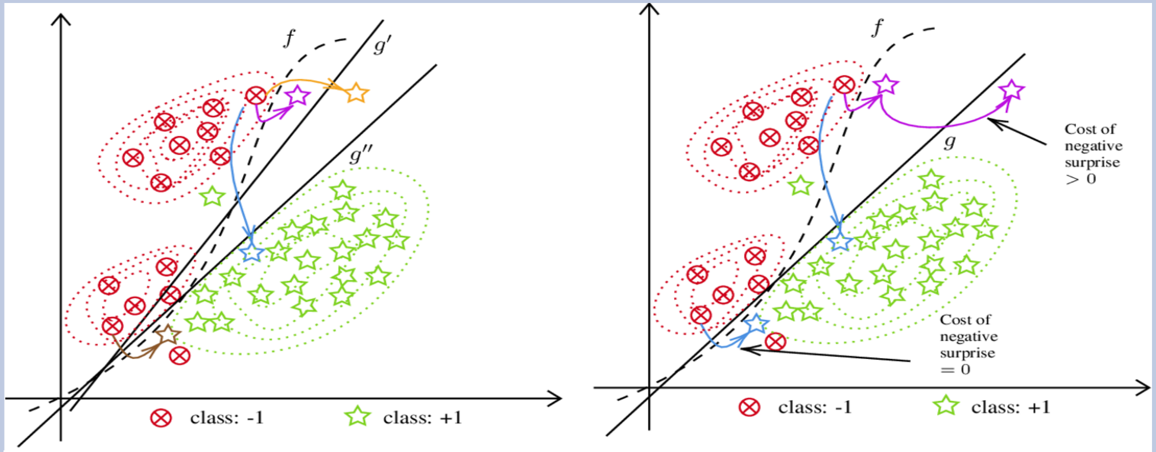
Logistic Regression model

2-layer neural network model

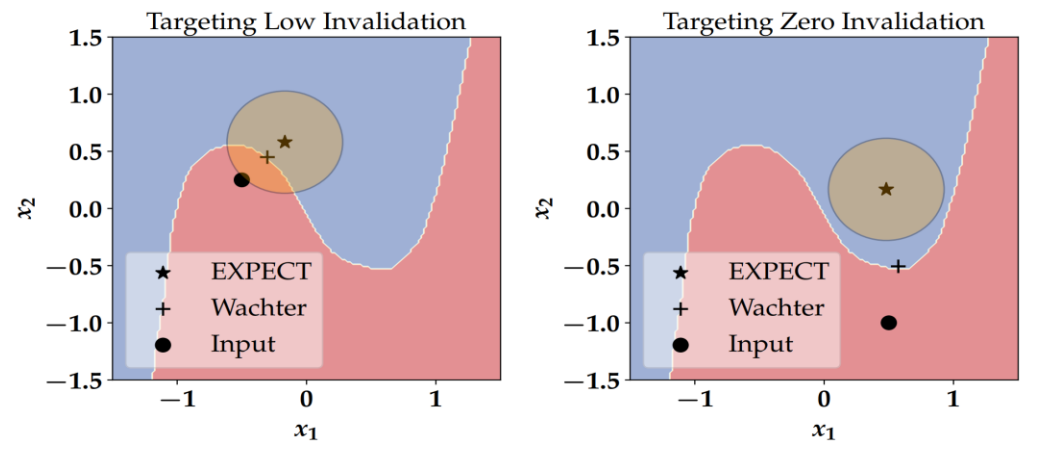


Median recourse invalidation rates are ~50%. Thus, if the recourse responses were noisy, then recourse success would often be **equivalent to a random coin flip.**

Further research on practical CEs



Invalidation under recalibration or model update,
UAI PMLR 2020



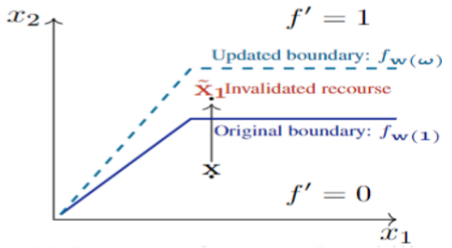
Invalidation under noisy recourse implementation by end users
(joint work with Harvard Business School), under review

Recourse Robustness & Data Deletion Requests

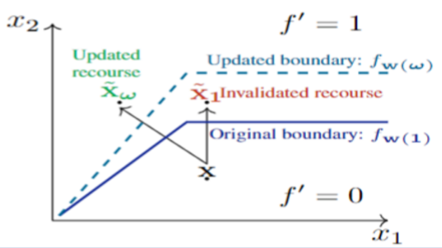
Goal
Study robustness in the presence of deletion requests

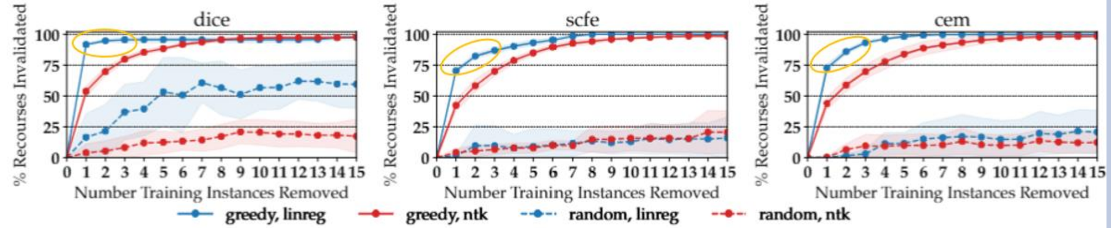
Idea
- Assign every training point a *data weight* $\omega \in \{0,1\}$
- Study changes to recourse $\tilde{x}$ (action stability) or $f(\tilde{x})$ (outcome stability) when $\boldsymbol{\omega}$ changes

Recourse outcome stability

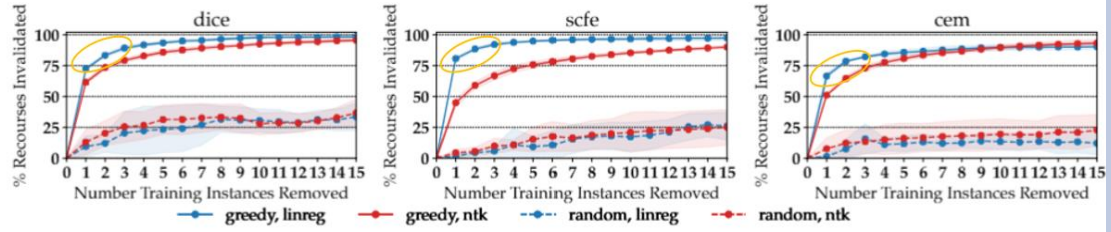


Recourse action stability





(a) Admission ($n = 17301$)

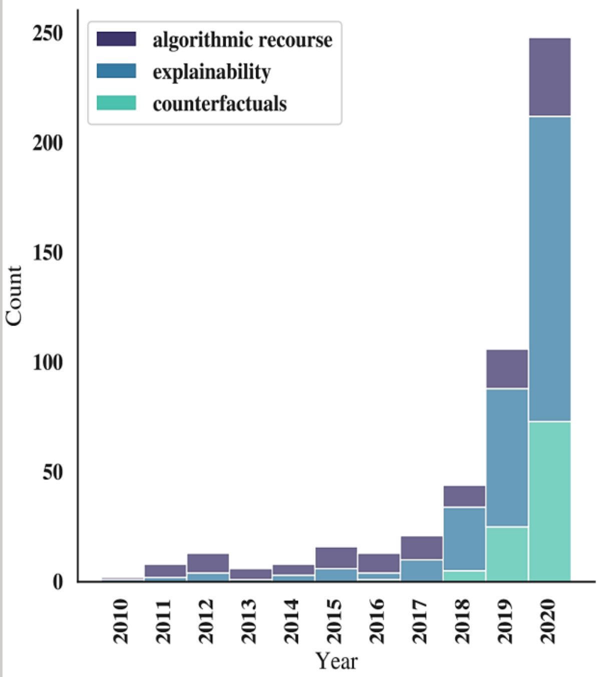


(b) Heloc ($n = 7896$)

CARLA: Counterfactual And Recourse LibrAry

Classifiers × Data sets × Measures × Methods

$$\underbrace{\begin{Bmatrix} \text{Logistic Regression} \\ \text{Neural Networks} \\ \text{RF} \\ \text{Xgboost} \end{Bmatrix}}_{4} \times \underbrace{\begin{Bmatrix} \text{Adult} \\ \text{COMPAS} \\ \text{GiveMeCredit} \\ \text{Heloc} \end{Bmatrix}}_{4} \times \underbrace{\begin{Bmatrix} \ell_0 \\ \ell_1 \\ \vdots \end{Bmatrix}}_{7} \times \underbrace{\begin{Bmatrix} \text{AR} \\ \text{C-CHVAE} \\ \vdots \end{Bmatrix}}_{13}$$

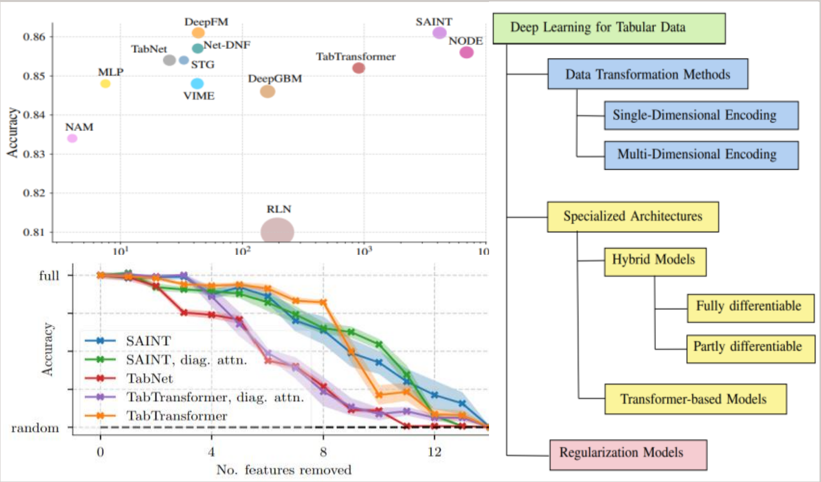


Approach	Method	Model Type	Algorithm	Immutable	Handle categorical	Other
Independent (**I**)	AR	Linear	Integer Prog.	Yes	Binary only	Direction of change
	AR-LIME	Agnostic	Integer Prog.	Yes	Binary only	Direction of change
	AS	Gradient based	Brute Force	Yes	Binary only	Action sets
	CEM	Gradient based	Gradient based	No	No	None
	DICE	Gradient Based	Gradient based	Yes	Binary Only	Generative model
	GS	Agnostic	Random search	Yes	Binary Only	None
Dependent (**D**)	CEM-VAE	Gradient based	Gradient based	No	No	Gen. Model regularizer
	CLUE	Gradient based	Gradient based	No	No	Generative model
	EB-CF	Gradient based	Gradient based	Yes	All	User Constraints
	FACE-EPS	Agnostic	Graph search	Binary Only	Binary Only	CE is from data set
	FACE-KNN	Agnositc	Graph search	Binary Only	Binary Only	CE is from data set

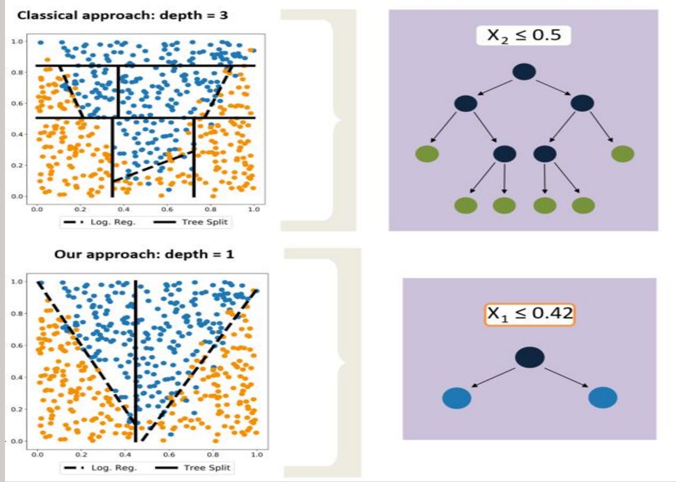
Table 1: Explanation method summary.

CARLA (Counterfactual And Recourse LibrAry) – Python library for standardized benchmarking of CE methods,
NeurIPS Datasets and Benchmarks 2021

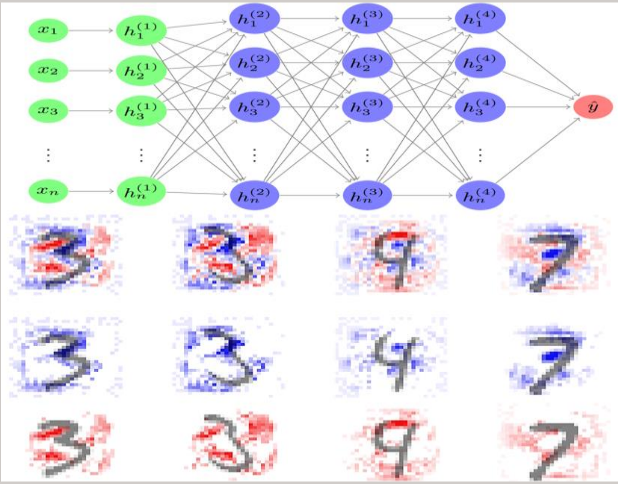
Further research in explainable and reliable AI



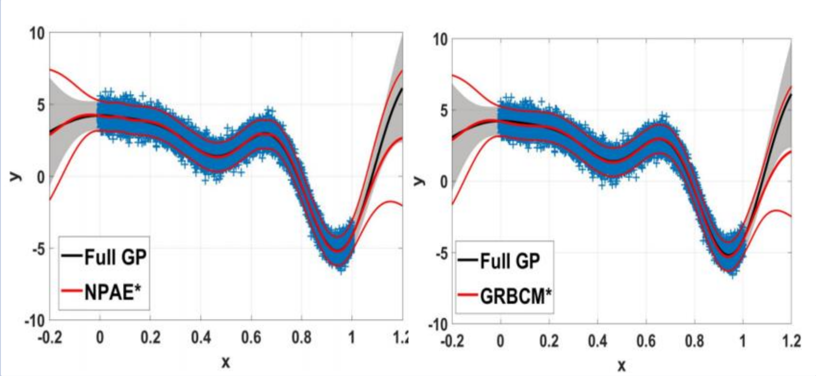
Deep Learning & Tabular Data (arXiv'21, ICML'22, IJDS'22)



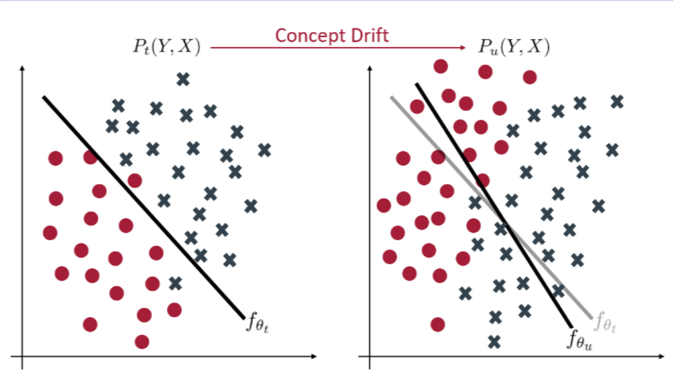
Explainability by design (IJCAI'19, ICDE'22)



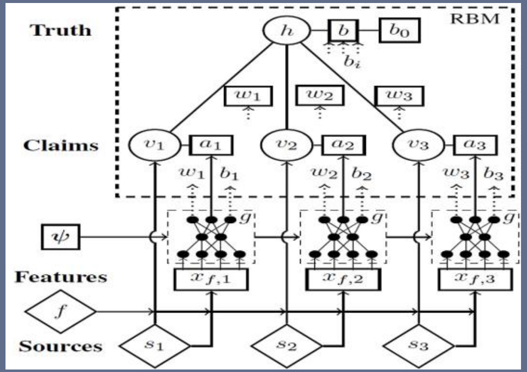
Explainability and feature attribution for neural networks (CIKM'17, ICANN'19)



Scalable Gaussian Processes to understand uncertainty (ICPR'20, IEEE BigData'21, BigComp'22)



Variable importance and concept drift in streaming data (KDD'20, ICPR'20, CIKM'22)



Truth Discovery: Data Quality and Reliability (ICDE'17, CHB'17, ...)

BACK-UP

Automated decision support may have influence on future prosperity

Probability of default: >10%

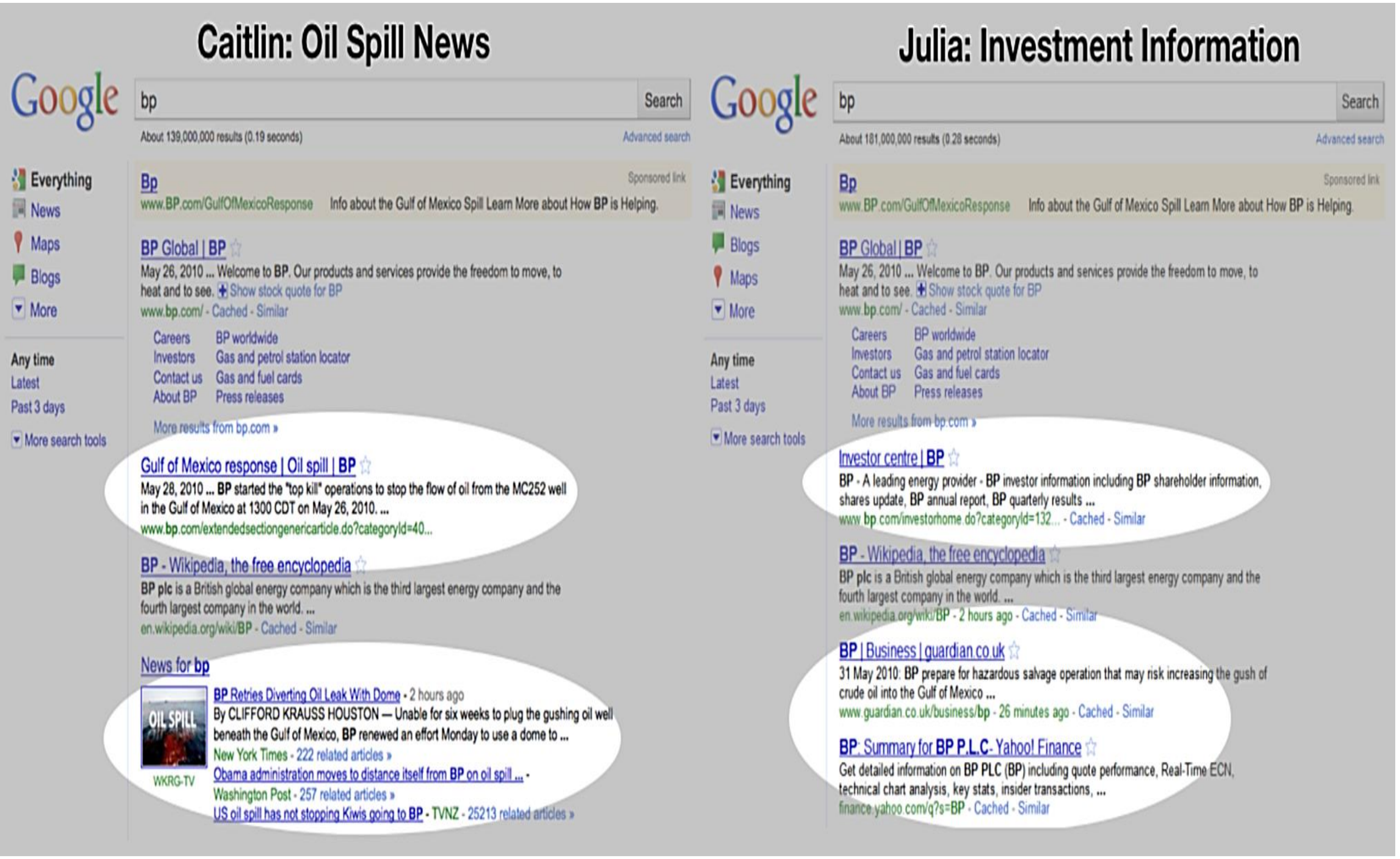
According to the Equal Credit Opportunity Act (ECOA), the Fair Credit Reporting Act (FCRA) in the US, and the General Data Protection Regulation (GDPR) in Europe, creditors are required to offer their customers minimum explanations for why a specific decision was made → **adverse action notice**

Feature Subset	Current Value
# delays elsewhere / year	5
current income	\$1000
tenure w/ current job	4 months
credit file	NaN

Q2: What can I do to achieve the desired outcome?

Feature Subset	Current Value		Required
# delays elsewhere / year	5	→	0
current income	\$1000	→	\$2500
tenure w/ current job	4 months	→	12 months
credit file	NaN	→	True

AI has influence on personal lives



Source: https://medium.com/the-open-book/filter-bubbles-3eca0f892366

Source: Yongfeng Zhang; Xu Chen, Explainable Recommendation: A Survey and New Perspectives, now, 2020.

Handling different data types

Nazabal, Alfredo, et al. "Handling incomplete heterogeneous data using vaes." Pattern Recognition 107 (2020): 107501

- Continuous/real-valued data $$p(x_{nd}|\boldsymbol{\gamma}_{nd}) = \mathcal{N}\left(x_{nd}|\mu_d(\mathbf{z}_n), \sigma_d^2(\mathbf{z}_n)\right)$$
- Positive real-valued data $$p(x_{nd}|\boldsymbol{\gamma}_{nd}) = \log\mathcal{N}\left(x_{nd}|\mu_d(\mathbf{z}_n), \sigma_d^2(\mathbf{z}_n)\right)$$
- Counting data $$p(x_{nd}|\boldsymbol{\gamma}_{nd}) = \mathrm{Poiss}\left(x_{nd}|\lambda_d(\mathbf{z}_n)\right) = \frac{(\lambda_d(\mathbf{z}_n))^{x_{nd}}\exp(-\lambda_d(\mathbf{z}_n))}{x_{nd}!}$$
- Categorical data $$p(x_{nd} = r|\boldsymbol{\gamma}_{nd}) = \frac{\exp(-h_{dr}(\mathbf{z}_n))}{\sum_{q=1}^{R}\exp(-h_{dq}(\mathbf{z}_n))}$$
- Ordinal data $$p(x_{nd} = r|\boldsymbol{\gamma}_{nd}) = p(x_{nd} \leq r|\boldsymbol{\gamma}_{nd}) - p(x_{nd} \leq r-1|\boldsymbol{\gamma}_{nd})$$ $$p(x_{nd} \leq r|\mathbf{z}_n) = \frac{1}{1+\exp(-(\theta_r(\mathbf{z}_n) - h_d(\mathbf{z}_n)))}$$