GENERALIZED INVERSE CLASSIFICATION

SDM '17

Michael T. Lash¹, Qihang Lin², W. Nick Street², Jennifer G. Robinson³, and Jeffrey Ohlmann²

¹Department of Computer Science, ²Department of Management Sciences, ³Department of Epidemiology

www.michaeltlash.com
What is *inverse classification*?

- The process of making **meaningful perturbations** to a test instance such that the probability of a desirable outcome is maximized.
What is *inverse classification*?

- The process of making meaningfull perturbations to a test instance such that the probability of a desirable outcome is maximized.
What is *inverse classification*?

- The process of making *meaningful perturbations* to a test instance such that the probability of a desirable outcome is maximized.

![Diagram](image)
What is *inverse classification*?

- The process of making *meaningful perturbations* to a test instance such that the probability of a desirable outcome is maximized.

A test instance $x$ (e.g., student)
What is *inverse classification*?

- The process of making **meaningful perturbations** to a test instance such that the probability of a desirable outcome is maximized.
What is *inverse classification*?

- The process of making *meaningful perturbations* to a test instance such that the probability of a desirable outcome is maximized.
What is *inverse classification*?

- The process of making *meaningful perturbations* to a test instance such that the probability of a desirable outcome is maximized.
What is *inverse classification*?

- The process of making *meaningful perturbations* to a test instance such that the probability of a desirable outcome is maximized.

What about the *meaningful* part of the definition?
Well...let's visit some past work!

Meaningful Perturbations

\[
\min_x f(x)
\]

Begin with a basic formulation.
\[
\min_x f(x)
\]

[For now] A differentiable classification function e.g., logistic regression, SVM, ANN, etc.
Meaningful Perturbations

\[ \min_{x} f(x) \]
Meaningful Perturbations

Segment features.
Meaningful Perturbations

\[
\min_x f(x)
\]

Some regressor

\[
H(\bar{x}_U, x_D)
\]

Some regressor

\[
H : \mathbb{R}^{|U|+|D|} \rightarrow \mathbb{R}^{|I|}
\]

Estimate indirectly changeable.
\[
\min_{\bar{x}_D} \ f(\bar{x}_U, H(\bar{x}_U, x_D), x_D) \\
\text{s.t. } \phi(x_D - \bar{x}_D) \leq B \\
\quad i; \leq x_i \leq u_i \text{ for } i \in D
\]

Update objective function. Add constraints.
Meaningful Perturbations

\[
\min_{\mathbf{x}_D} f(\bar{\mathbf{x}}_U, H(\bar{\mathbf{x}}_U, \mathbf{x}_D), \mathbf{x}_D)
\]

s.t. \( \phi(\mathbf{x}_D - \bar{\mathbf{x}}_D) \leq B \)

\( l_i \leq x_i \leq u_i \) for \( i \in D \)

\[
\phi(\mathbf{z}) = \sum_{i \in D} c_i^+(z_i)_+ + c_i^-(z_i)_-
\]

Cost-change function
Meaningful Perturbations

\[
\min_{x_D} f(\bar{x}_U, H(\bar{x}_U, x_D), x_D) \\
\text{s.t. } \phi(x_D - \bar{x}_D) \leq B \text{ Budget} \\
l_i \leq x_i \leq u_i \text{ for } i \in D
\]

\[
\phi(z) = \sum_{i \in D} c_i^+(z_i)_+ + c_i^-(z_i)_-
\]
Meaningful Perturbations

$$\min_{x_D} f(\bar{x}_U, H(\bar{x}_U, x_D), x_D)$$

s.t. $\phi(x_D - \bar{x}_D) \leq B$

$$l_i \leq x_i \leq u_i \text{ for } i \in D$$

Bounds
1. Relax assumptions about $f(\cdot)$.

$$\min_{x_D} f(\bar{x}_U, H(\bar{x}_U, x_D), x_D)$$

s.t. $\phi(x_D - \bar{x}_D) \leq B$

$$l_i \leq x_i \leq u_i \text{ for } i \in D$$

Assume:

- $f(z)$ differentiable
- $\|\nabla f(z) - \nabla f(z')\| \leq L \|\nabla f(z) - \nabla f(z')\|$ for $z, z' \in \mathbb{R}^{|D|}$
Main Contributions

1. Relax assumptions about $f(\cdot)$.

$$\min_{x_D} f(\bar{x}_U, H(\bar{x}_U, x_D), x_D)$$

s.t. $\phi(x_D - \bar{x}_D) \leq B$

$$l_i \leq x_i \leq u_i \text{ for } i \in D$$

Assume:
- $f(z)$ differentiable
- $\|\nabla f(z) - \nabla f(z')\| \leq L \|\nabla f(z) - \nabla f(z')\| z, z' \in \mathbb{R}^{|D|}$
1. Relax assumptions about $f(\cdot)$. 

$$\min_{x_D} f(\bar{x}_U, H(\bar{x}_U, x_D), x_D)$$

s.t. $\phi(x_D - \bar{x}_D) \leq B$

$$l_i \leq x_i \leq u_i \text{ for } i \in D$$

Assume:
- $f: \mathbb{R}^p \rightarrow \mathbb{R}$

*Generalized* inverse classification
Main Contributions

1. Relax assumptions about $f(\cdot)$.
2. **Quadratic cost-change function.**

$$\min_{x_D} f(\bar{x}_U, H(\bar{x}_U, x_D), x_D)$$

s.t. $\phi(x_D - \bar{x}_D) \leq B$

$$l_i \leq x_i \leq u_i \text{ for } i \in D$$
Main Contributions

1. Relax assumptions about $f(\cdot)$.
2. **Quadratic cost-change function.**

\[
\begin{align*}
\min_{x_D} & \quad f(\bar{x}_U, H(\bar{x}_U, x_D), x_D) \\
\text{s.t.} & \quad \phi(x_D - \bar{x}_D) \leq B \\
& \quad l_i \leq x_i \leq u_i \text{ for } i \in D
\end{align*}
\]

\[
\phi(z) = \sum_{i \in D} c_i^+(z_i)_+ + c_i^-(z_i)_-
\]
Main Contributions

1. Relax assumptions about $f(\cdot)$.

2. Quadratic cost-change function.

$$\min_{x_D} f(\bar{x}_U, H(\bar{x}_U, x_D), x_D)$$

s.t. $\phi(x_D - \bar{x}_D) \leq B$

$$l_i \leq x_i \leq u_i \text{ for } i \in D$$

$$\phi(z) = \sum_{i \in D} c_i^+(z_i)^2 + c_i^- z(i)^2$$
Main Contributions

1. Relax assumptions about $f(\cdot)$.
2. Quadratic cost-change function.
3. Three real-valued heuristic optimization methods and two sensitivity analysis-based optimization methods.
   * Projection operator to maintain feasibility.
Optimization Methodology

**Heuristic**

- Hill Climbing + Local Search (HC+LS)
- Genetic Algorithm (GA)
- Genetic Algorithm + Local Search (GA+LS)

**Sensitivity Analysis**

- Local Variable Perturbation – First Improvement (LVP-FI)
- Local Variable Perturbation – Best Improvement (LVP-BI)
Experiment Decisions and Data

- $f(\cdot)$: Random forest
- $H(\cdot)$: Kernel regression
- Dataset 1: Student Performance (UCI Machine Learning Repository).
- Dataset 2: ARIC
- One $f$ for optimization, separate $f$ for heldout evaluation.
Results: Student Performance

![Graph showing student performance results over budget. The graph includes lines for various algorithms such as GA, HC+LS, GA+LS, LVP-BI, LVP-FI, and Std 57. The y-axis represents probability, and the x-axis represents budget. Different lines represent different algorithms, with GA shown in red, HC+LS in cyan, GA+LS in blue, LVP-BI in black dots, LVP-FI in purple, and Std 57 in green. The graph illustrates how each algorithm performs across different budget levels.]
Results: Student Performance

- Time out with friends
- Weekday alcohol consumption
- Time spent studying

[Graph showing the relationship between budget and feature change]
Results: ARIC

Need sparsity constraints
Conclusions

- **Generalized Inverse Classification**: can use virtually any learned $f$ (as shown by experiments w/ Random Forest classifier).

- Our proposed methods were successful, although this varied by dataset.
GENERALIZED INVERSE CLASSIFICATION

SDM '17

Michael T. Lash\textsuperscript{1}, Qihang Lin\textsuperscript{2}, W. Nick Street\textsuperscript{2}, Jennifer G. Robinson\textsuperscript{3}, and Jeffrey Ohlmann\textsuperscript{2}

\textsuperscript{1}Department of Computer Science, \textsuperscript{2}Department of Management Sciences, \textsuperscript{3}Department of Epidemiology

www.michaeltlash.com
Causality and Inverse Classification

What we’re doing:

1. Imposing our own causal structure (DAG).
2. We’re not taking the usual counterfactual approach.

Future work will focus on incorporating causal methodology...
Causality and Inverse Classification

- Yes! ....
Causality and Inverse Classification

- Yes! ....
- What we’re doing:
  1. Imposing our own causal structure (DAG).
  2. We’re not taking the usual counterfactual approach.
Yes! ....

What we’re doing:

1. Imposing our own causal structure (DAG).
2. We’re not taking the usual counterfactual approach.

Future work will focus on incorporating causal methodology...