Fairness through Feature Acquisition

Hortense Fong, Vineet Kumar, Anay Mehrotra and Nisheeth Vishnoi

Yale University

Algorithmic Decisions

- Where do Algorithms make decisions in business + society?

Algorithmic Decisions

- Where do Algorithms make decisions in business + society?
  - Human Capital: Resume screening, University admissions / enrollment

Algorithmic Decisions

- Where do Algorithms make decisions in business + society?
  - Human Capital: Resume screening, University admissions / enrollment
  - Medical Care: Which patients to monitor more intensively or even escalate care?

Algorithmic Decisions

- Where do Algorithms make decisions in business + society?
  - Human Capital: Resume screening, University admissions / enrollment
  - Medical Care: Which patients to monitor more intensively or even escalate care?
  - Loans: Which customers to approve for a auto / home / personal loan?

Algorithmic Decisions

- Where do Algorithms make decisions in business + society?
  - Human Capital: Resume screening, University admissions / enrollment
  - Medical Care: Which patients to monitor more intensively or even escalate care?
  - Loans: Which customers to approve for a auto / home / personal loan?
  - Criminal Justice: Pre-trial bail

Algorithmic Decisions

- Where do Algorithms make decisions in business + society?
  - Human Capital: Resume screening, University admissions / enrollment
  - Medical Care: Which patients to monitor more intensively or even escalate care?
  - Loans: Which customers to approve for a auto / home / personal loan?
  - Criminal Justice: Pre-trial bail
- Lots of places where it is not obvious that algorithms might be involved.

Algorithmic Decisions

- Where do Algorithms make decisions in business + society?
  - Human Capital: Resume screening, University admissions / enrollment
  - Medical Care: Which patients to monitor more intensively or even escalate care?
  - Loans: Which customers to approve for a auto / home / personal loan?
  - Criminal Justice: Pre-trial bail
- Lots of places where it is not obvious that algorithms might be involved.

Algorithmic Decisions

- Where do Algorithms make decisions in business + society?
  - Human Capital: Resume screening, University admissions / enrollment
  - Medical Care: Which patients to monitor more intensively or even escalate care?
  - Loans: Which customers to approve for a auto / home / personal loan?
  - Criminal Justice: Pre-trial bail
- Lots of places where it is not obvious that algorithms might be involved.

Major Challenge:
Algorithmic Bias across groups (think race, gender, age, income)

Humans may be able to override decisions, but often have limited time / energy / attention

Impact of Bias: Human Capital

RETAIL OCTOBER 10, 2018 / 7:04 PM / UPDATED 4 YEARS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin

8 MIN READ

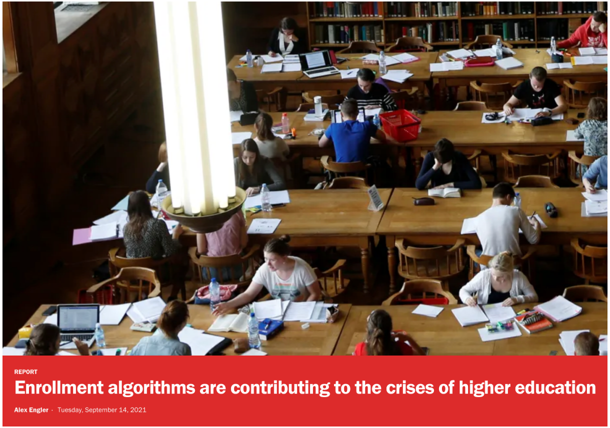
SAN FRANCISCO (Reuters) - Amazon.com Inc's AMZN.O machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

ML Problem:

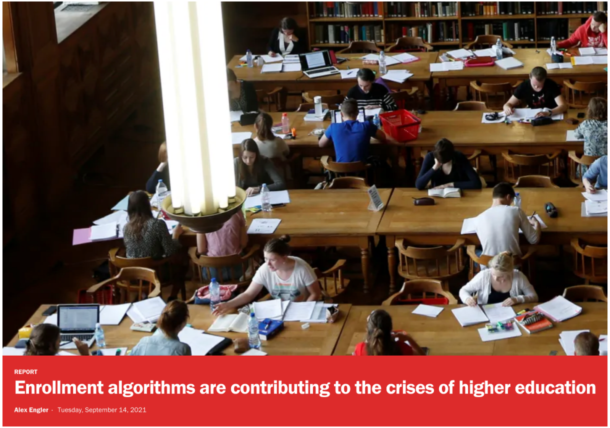
Prediction Problem: Predict quality or fit of applicant (one to five stars)
Input to algorithm: Resume
Decision: Interview or Not
Bias: women-related words decreased stars

Impact of Bias: Human Capital

Impact of Bias: Human Capital



Impact of Bias: Human Capital



Impact of Bias: Human Capital

REPORT

Enrollment algorithms are contributing to the crises of higher education

Alex Engler · Tuesday, September 14, 2021

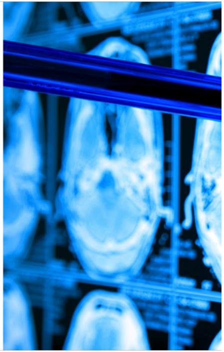
ML Problem:

Prediction problem (Y): Likelihood of accepting
Input to algorithm (X): Student information
Decision: How much financial aid to offer
Bias: More accurate for higher income

Impact of Bias: Medical Care

Algorithms Are Making Decisions About Health Care, Which May Only Worsen Medical Racism

Unclear regulation and a lack of transparency increase the risk that AI and algorithmic tools that exacerbate racial biases will be used in medical settings.

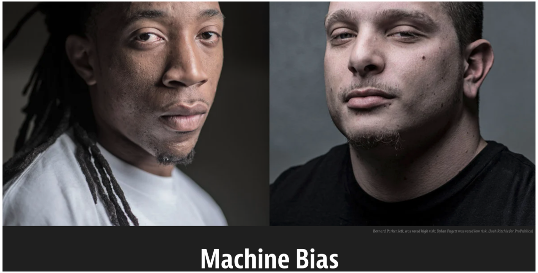


ML Problem:

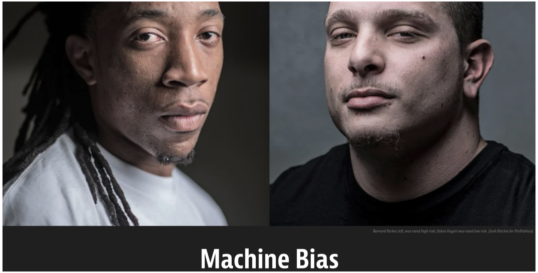
Prediction problem (Y): who is likely to have a serious condition
Input to algorithm (X): insurance claims, diagnosis codes, etc.
Decision: extra medical attention and care
Bias: For same risk assessment, Black patients sicker than White patients

Impact of Bias: Criminal Justice

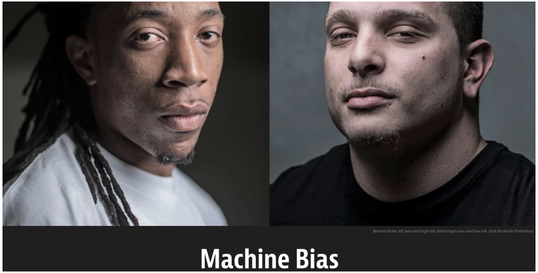
Impact of Bias: Criminal Justice



Impact of Bias: Criminal Justice



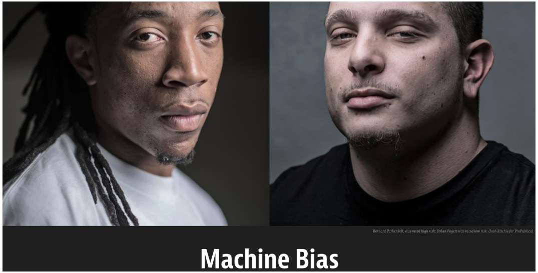
Impact of Bias: Criminal Justice



ML Problem:

Prediction problem (Y): likelihood of re-offending
Input to algorithm (X): 137 question survey
Decision: Offer Bail or Not
Bias: Higher FPR among Blacks

Impact of Bias: Criminal Justice



ML Problem:

Prediction problem (Y): likelihood of re-offending
Input to algorithm (X): 137 question survey
Decision: Offer Bail or Not
Bias: Higher FPR among Blacks

- Accuracy might not be a good performance metric because it cannot distinguish between FP and FN (Type 1 and Type 2)

Research Context

Bank trying to predict loan default $Y_i \in \{0, 1\}$

	Id	Income	Age	Experience	Married/Single	House_Ownership	Car_Ownership	Profession
0	1	1303834	23	3	single	rented	no	Mechanical_engineer
1	2	7574516	40	10	single	rented	no	Software_Developer
2	3	3991815	66	4	married	rented	no	Technical_writer
3	4	6256451	41	2	single	rented	yes	Software_Developer
4	5	5768871	47	11	single	rented	no	Civil_servant

Our approach focuses on acquiring new columns (not rows)

Why Feature Acquisition?

What is feature acquisition?

Our approach focuses on acquiring new columns (not rows)

> *"...most of our interviewees report ... data collection, rather than model development, as the most important place to intervene"*

—Holstein et al. (2019)

—Chen et al. (2018)

Why Feature Acquisition?

What is feature acquisition?

Our approach focuses on acquiring new columns (not rows)

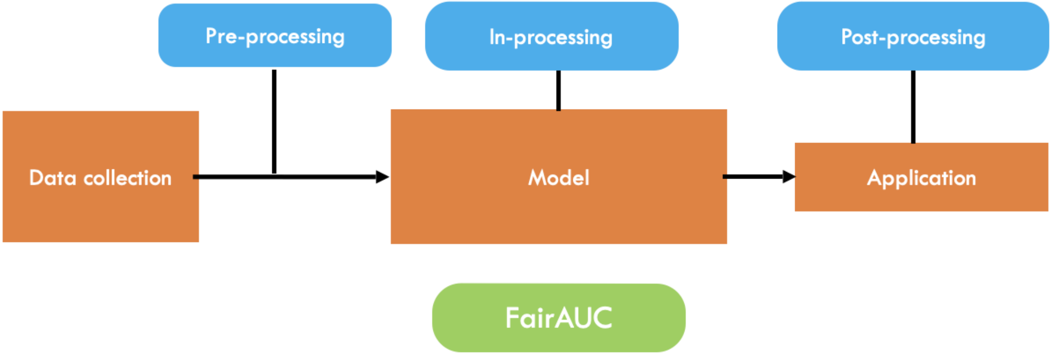
> *"... most of our interviewees report ... data collection, rather than model development, as the most important place to intervene"*
>
> —Holstein et al. (2019)

> *"Much research has been devoted to constraining models to satisfy cost-based fairness in prediction... The impact of data collection on discrimination has received comparatively little attention."*
>
> —Chen et al. (2018)

How and Where to Acquire Features?

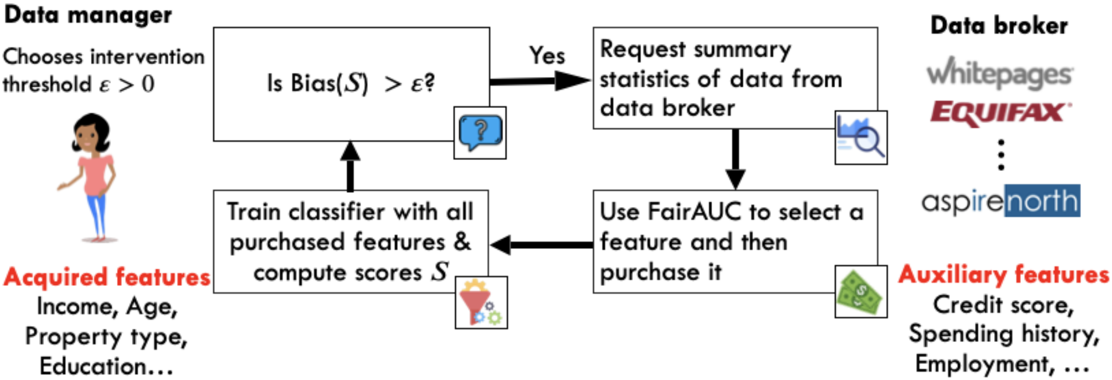
Problem[6]	Prediction Outcome $\hat{Y}$	First-party Data Examples $\hat{\mathbf{X}}$	Auxiliary Features Examples $\hat{\mathbf{Z}}$	Source
Loan provision	Default	Name, address, SSN, credit history[7]	Work history, college major, spending and saving behavior, social network data[8]	Data vendor, social data vendor
Bail decision	Recidivism	Criminal history, questionnaire responses[9]	Spending and saving behavior, credit history, social network data	Data vendor, social data vendor
Hiring	Promotion	Resume, referral, interview	Social network data engagement	Social data vendor
Extra Medical Attention	Hospital Readmission	Biomarker values, comorbidities	Wearables, social network data	Devices[10], social data vendor

Research Context



Big Picture Overview

FairAUC in the context of loan decisions:



Advantages of Method

- Dynamic approach to feature acquisition in order to mitigate bias
  $\implies$ Provides guarantees that bias will reduce with feature acquisition

Advantages of Method

- Dynamic approach to feature acquisition in order to mitigate bias
  $\implies$ Provides guarantees that bias will reduce with feature acquisition
- Can we used with any classification algorithm, with guarantees on lower bound of improvement for GLM models

Advantages of Method

- Dynamic approach to feature acquisition in order to mitigate bias
  $\Longrightarrow$ Provides guarantees that bias will reduce with feature acquisition
- Can we used with any classification algorithm, with guarantees on lower bound of improvement for GLM models
- Works with or without using group membership in classification

Advantages of Method

- Dynamic approach to feature acquisition in order to mitigate bias
  $\implies$ Provides guarantees that bias will reduce with feature acquisition
- Can we used with any classification algorithm, with guarantees on lower bound of improvement for GLM models
- Works with or without using group membership in classification
- Only requires summary statistics (mean, variances, and co-variances) of the data

Advantages of Method

- Dynamic approach to feature acquisition in order to mitigate bias
  $\implies$ Provides guarantees that bias will reduce with feature acquisition
- Can we used with any classification algorithm, with guarantees on lower bound of improvement for GLM models
- Works with or without using group membership in classification
- Only requires summary statistics (mean, variances, and co-variances) of the data
- Robust to issues of *reverse discrimination* because the disadvantaged group can change over time as we acquire features

Preview of Results

- Unconditional variance of features not predictive of bias

Preview of Results

- Unconditional variance of features not predictive of bias
- Class conditional variance is what is important for AUC and fairness

Preview of Results

- Unconditional variance of features not predictive of bias
- Class conditional variance is what is important for AUC and fairness
- Equalizing size of data across groups might not reduce bias

Preview of Results

- Unconditional variance of features not predictive of bias
- Class conditional variance is what is important for AUC and fairness
- Equalizing size of data across groups might not reduce bias
  - Even as number of observations grows

Preview of Results

- Unconditional variance of features not predictive of bias
- Class conditional variance is what is important for AUC and fairness
- Equalizing size of data across groups might not reduce bias
  - Even as number of observations grows
- Algorithm that is focused on *greedily* minimizing bias can actually perform worse on both bias and performance than algorithm that focuses only on performance

Preview of Results

- Unconditional variance of features not predictive of bias
- Class conditional variance is what is important for AUC and fairness
- Equalizing size of data across groups might not reduce bias
    - Even as number of observations grows
- Algorithm that is focused on *greedily* minimizing bias can actually perform worse on both bias and performance than algorithm that focuses only on performance
- We use a canonical dataset (COMPAS) and acquire new features for these individuals

Preview of Results

- Unconditional variance of features not predictive of bias
- Class conditional variance is what is important for AUC and fairness
- Equalizing size of data across groups might not reduce bias
  - Even as number of observations grows
- Algorithm that is focused on *greedily* minimizing bias can actually perform worse on both bias and performance than algorithm that focuses only on performance
- We use a canonical dataset (COMPAS) and acquire new features for these individuals
  - Our algorithm focused on the (currently) disadvantaged group works much better in terms of both bias and performance

Problem Setting

Data: There are N samples (N is large)

- *Acquired features:* Each sample i, contains:

Problem Setting

Data: There are N samples (N is large)

- *Acquired features:* Each sample i, contains:
  - a class label $Y_i \in \{-1, 1\}$

Problem Setting

Data: There are N samples (N is large)

- *Acquired features:* Each sample i, contains:
  - a class label $Y_i \in \{-1, 1\}$
  - group $A_i \in \{a, b\}$

Problem Setting

Data: There are N samples (N is large)

- *Acquired features:* Each sample i, contains:
  - a class label $Y_i \in \{-1, 1\}$
  - group $A_i \in \{a, b\}$
  - (d) features $X_i^1, X_i^2, \ldots, X_i^d \in \mathbb{R}$

Problem Setting

Data: There are N samples (N is large)

- *Acquired features:* Each sample i, contains:
  - a class label $Y_i \in \{-1, 1\}$
  - group $A_i \in \{a, b\}$
  - (d) features $X_i^1, X_i^2, \ldots, X_i^d \in \mathbb{R}$
- *Unacquired features:* For each sample i, there are $m - d$ unacquired features $Z_i^1, Z_i^2, \ldots, Z_i^{m-d} \in \mathbb{R}$

Problem Setting

Data: There are N samples (N is large)

- *Acquired features:* Each sample i, contains:
  - a class label $Y_i \in \{-1, 1\}$
  - group $A_i \in \{a, b\}$
  - (d) features $X_i^1, X_i^2, \ldots, X_i^d \in \mathbb{R}$
- *Unacquired features:* For each sample i, there are $m - d$ unacquired features $Z_i^1, Z_i^2, \ldots, Z_i^{m-d} \in \mathbb{R}$

Problem Setting

Data: There are N samples (N is large)

- *Acquired features:* Each sample i, contains:
  - a class label $Y_i \in \{-1, 1\}$
  - group $A_i \in \{a, b\}$
  - (d) features $X_i^1, X_i^2, \ldots, X_i^d \in \mathbb{R}$
- *Unacquired features:* For each sample i, there are $m - d$ unacquired features $Z_i^1, Z_i^2, \ldots, Z_i^{m-d} \in \mathbb{R}$

Manager: The manager has access to the all acquired features and can obtain Unacquired features at a cost

- Can acquire features from **Data Vendors** (e.g., Whitepages and Acxiom)

Problem Setting

Data: There are N samples (N is large)

- *Acquired features:* Each sample i, contains:
  - a class label $Y_i \in \{-1, 1\}$
  - group $A_i \in \{a, b\}$
  - (d) features $X_i^1, X_i^2, \ldots, X_i^d \in \mathbb{R}$
- *Unacquired features:* For each sample i, there are $m - d$ unacquired features $Z_i^1, Z_i^2, \ldots, Z_i^{m-d} \in \mathbb{R}$

Manager: The manager has access to the all acquired features and can obtain Unacquired features at a cost

- Can acquire features from **Data Vendors** (e.g., Whitepages and Acxiom)
- Selects a "hypothesis class" of classifiers to be any Generalized Linear Model (e.g., SVM or Logistic regression)

Problem Setting

Goal: Train a classifier $f : (X, Z) \to Y$.
Classifier takes features (X, Z) as input predicts the class label Y, while ensuring that f has "high performance" on both groups.

Disadvantaged Group:

Group on which the classifier f **currently** has lower performance (AUC)

Benchmarks

Problem Setting

Goal: Train a classifier $f : (X, Z) \to Y$.
Classifier takes features (X, Z) as input predicts the class label Y, while ensuring that f has "high performance" on both groups.

Disadvantaged Group:

Group on which the classifier f **currently** has lower performance (AUC)

Benchmarks

- FairAUC: Our proposed approach

Problem Setting

Goal: Train a classifier $f : (X, Z) \to Y$.
Classifier takes features (X, Z) as input predicts the class label Y, while ensuring that f has "high performance" on both groups.

Disadvantaged Group:

Group on which the classifier f **currently** has lower performance (AUC)

Benchmarks

- FairAUC: Our proposed approach
- minBias: Choose feature that minimizes bias between groups

Problem Setting

Goal: Train a classifier $f : (X, Z) \to Y$.
Classifier takes features (X, Z) as input predicts the class label Y, while ensuring that f has "high performance" on both groups.

Disadvantaged Group:

Group on which the classifier f **currently** has lower performance (AUC)

Benchmarks

- FairAUC: Our proposed approach
- minBias: Choose feature that minimizes bias between groups
- maxAUC: Choose feature to maximize performance (AUC)

Problem Setting

Goal: Train a classifier $f : (X, Z) \rightarrow Y$.
Classifier takes features (X, Z) as input predicts the class label Y, while ensuring that f has "high performance" on both groups.

Disadvantaged Group:

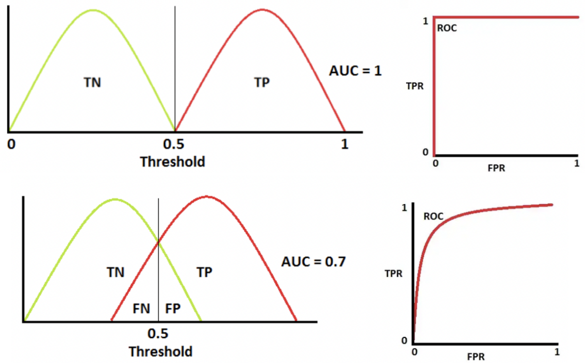
Group on which the classifier f **currently** has lower performance (AUC)

Benchmarks

- FairAUC: Our proposed approach
- minBias: Choose feature that minimizes bias between groups
- maxAUC: Choose feature to maximize performance (AUC)
- Random

Performance: AUC (ROC Curve)

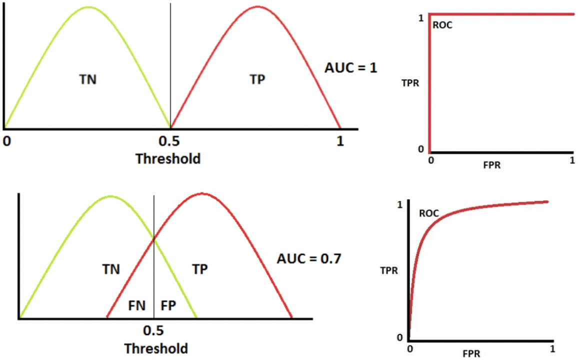
- ROC curve is the plot of TPR versus FPR



Credit: `https://towardsdatascience.com`

Performance: AUC (ROC Curve)

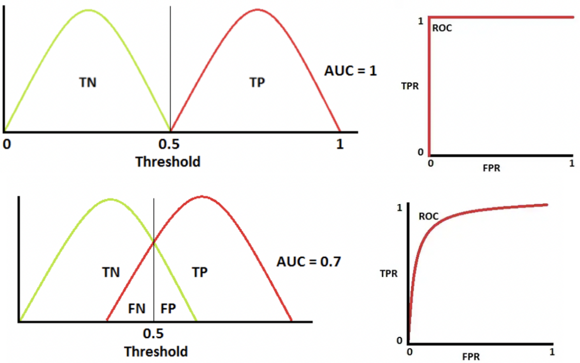
- ROC curve is the plot of TPR versus FPR
- AUC is the "Area Under Curve" and includes both Type 1 and Type 2 errors



Credit: `https://towardsdatascience.com`

Performance: AUC (ROC Curve)

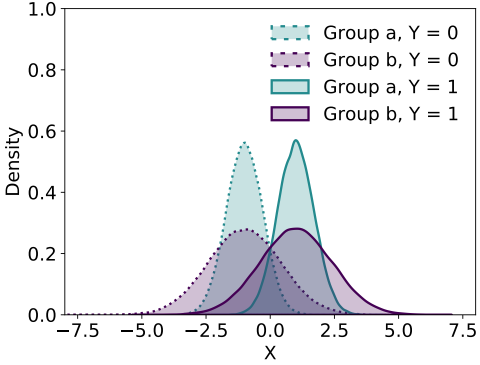
- ROC curve is the plot of TPR versus FPR
- AUC is the "Area Under Curve" and includes both Type 1 and Type 2 errors
- AUC lies between 0 and 1, higher is better $\implies$ **Performance Metric**

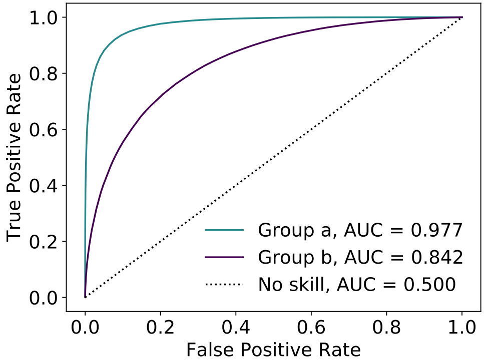


Credit: `https://towardsdatascience.com`

Extending this to multiple groups $g \in \{a, b\}$

$$\text{Bias} := 1 - \frac{\min_g(\text{AUC}_g)}{\max_g(\text{AUC}_g)}.$$





Multivariate Framework

Multivariate Normal Distribution with arbitrary Correlation structure: Conditioned on the class label $Y = y$ and the group $A = a$ (for any $y \in \{-1, 1\}$ and $a \in \{a, b\}$), all m features are drawn from a multi-variate normal distribution:

$$(X, Z) \mid (Y = y, A = a) \sim \mathcal{N}(\mu_{ya}, \Sigma_{ya}).$$

Focus is on easily interpretable summary statistics and theoretical guarantees:

Multivariate Framework

Multivariate Normal Distribution with arbitrary Correlation structure: Conditioned on the class label $Y = y$ and the group $A = a$ (for any $y \in \{-1, 1\}$ and $a \in \{a, b\}$), all m features are drawn from a multi-variate normal distribution:

$$(X, Z) \mid (Y = y, A = a) \sim \mathcal{N}(\mu_{ya}, \Sigma_{ya}).$$

Focus is on easily interpretable summary statistics and theoretical guarantees:

Proposition

For any family of Generalized Linear Models (e.g., Logistic Regression or SVM), the best AUC achievable by a classifier f in this family on group $g \in \{a, b\}$ is

$$\Phi\left(\sqrt{(\mu_{1g} - \mu_{0g})^\top (\Sigma_{0g} + \Sigma_{1g})^{-1} (\mu_{1g} - \mu_{0g})}\right).$$

Where $\Phi : \to [0, 1]$ is the CDF of the standard normal distribution.

Theory: Unconditional distribution does not inform AUC

Exploratory analysis revealed that many works analyze unconditional distributions of features for groups (Corbett-Davies and Goel, 2018) and (Chen et al., 2018)

However, unconditional distribution mixes together base rates, class-conditional means, and class-conditional variances, obscuring the relationship between the data and AUC

Observation

There are distributions D_1, D_2, D_3 *satisfying* $\mathrm{Var}[X|A = a] > \mathrm{Var}[X|A = b]$ *such that*

Theory: Unconditional distribution does not inform AUC

Exploratory analysis revealed that many works analyze unconditional distributions of features for groups (Corbett-Davies and Goel, 2018) and (Chen et al., 2018)

However, unconditional distribution mixes together base rates, class-conditional means, and class-conditional variances, obscuring the relationship between the data and AUC

Observation

There are distributions D_1, D_2, D_3 satisfying $\mathrm{Var}[X|A=a] > \mathrm{Var}[X|A=b]$ such that

- *AUC of group a is greater than AUC of group b with distribution D_1*

Theory: Unconditional distribution does not inform AUC

Exploratory analysis revealed that many works analyze unconditional distributions of features for groups (Corbett-Davies and Goel, 2018) and (Chen et al., 2018)

However, unconditional distribution mixes together base rates, class-conditional means, and class-conditional variances, obscuring the relationship between the data and AUC

Observation

There are distributions D_1, D_2, D_3 *satisfying* $\mathrm{Var}[X|A=a] > \mathrm{Var}[X|A=b]$ *such that*

- *AUC of group a is greater than AUC of group b with distribution* D_1
- *AUC of group a is equal to AUC of group b with distribution* D_2

Theory: Unconditional distribution does not inform AUC

Exploratory analysis revealed that many works analyze unconditional distributions of features for groups (Corbett-Davies and Goel, 2018) and (Chen et al., 2018)

However, unconditional distribution mixes together base rates, class-conditional means, and class-conditional variances, obscuring the relationship between the data and AUC

Observation

There are distributions D_1, D_2, D_3 satisfying $\mathrm{Var}[X|A=a] > \mathrm{Var}[X|A=b]$ such that

- *AUC of group a is greater than AUC of group b with distribution D_1*
- *AUC of group a is equal to AUC of group b with distribution D_2*
- *AUC of group a is less than AUC of group b with distribution D_3*

Theory: Result

FairAUC improves AUC of current disadvantaged group, rather than minimizing bias

Theorem (Performance guarantee of FairAUC)

Under the Binormal Framework and any Generalized Linear Model, at each iteration $t = 1, 2, \ldots,$ FairAUC increases the AUC of the disadvantaged group by at least

$$\max_{\ell} \frac{1}{18} \cdot (\gamma \cdot \beta_\ell \cdot (1 - \delta_\ell))^2,$$

and does not decrease the AUC of the advantaged group

Theory: Result

FairAUC improves AUC of current disadvantaged group, rather than minimizing bias

Theorem (Performance guarantee of FairAUC)

Under the Binormal Framework and any Generalized Linear Model, at each iteration $t = 1, 2, \dots$, FairAUC increases the AUC of the disadvantaged group by at least

$$\max_{\ell} \frac{1}{18} \cdot (\gamma \cdot \beta_\ell \cdot (1 - \delta_\ell))^2,$$

and does not decrease the AUC of the advantaged group

- $1 - \gamma$ is the (current) AUC of the disadvantaged group

Theory: Result

FairAUC improves AUC of current disadvantaged group, rather than minimizing bias

Theorem (Performance guarantee of FairAUC)

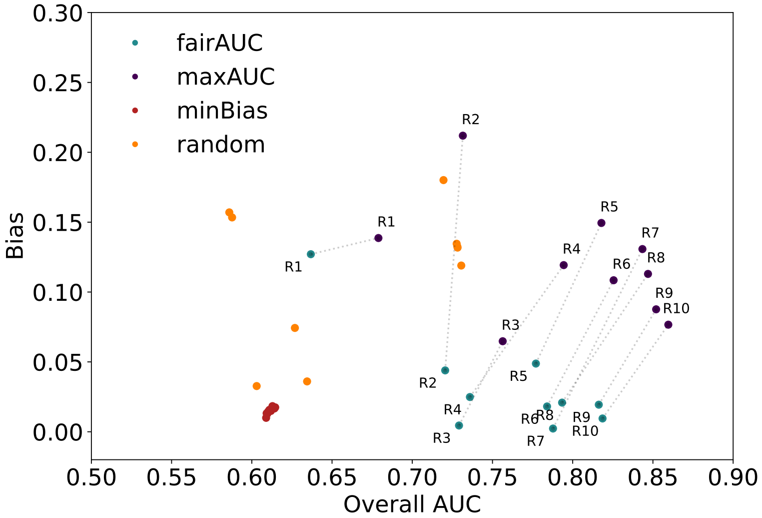
Under the Binormal Framework and any Generalized Linear Model, at each iteration $t = 1, 2, \ldots,$ FairAUC increases the AUC of the disadvantaged group by at least

$$\max_{\ell} \frac{1}{18} \cdot (\gamma \cdot \beta_\ell \cdot (1 - \delta_\ell))^2,$$

and does not decrease the AUC of the advantaged group

- $1 - \gamma$ is the (current) AUC of the disadvantaged group
- β_ℓ and δ_ℓ are data-dependent parameters that can be estimated by the manager

Synthetic Data: Result - Accuracy-Fairness Tradeoff



Feature Acquisition for Criminal Justice

Feature Acquisition for Criminal Justice

Feature Acquisition for Criminal Justice

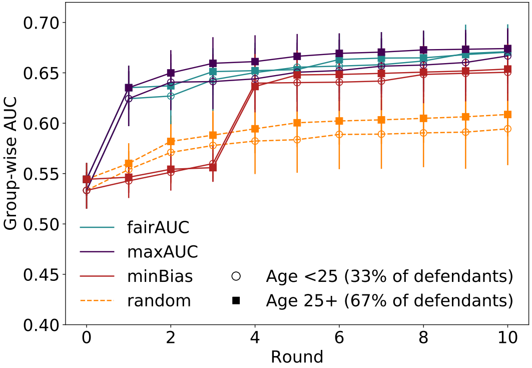
ML Problem:

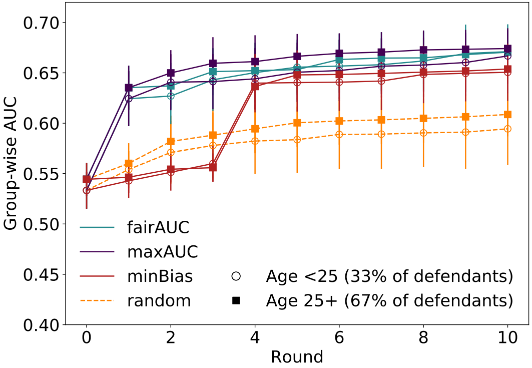
Prediction problem (Y): re-offending
Input to algorithm (X): 137 question survey
Decision: Offer Bail or Not

Feature Acquisition for Criminal Justice

ML Problem:

Prediction problem (Y): re-offending
Input to algorithm (X): 137 question survey
Decision: Offer Bail or Not





ML Problem:

Prediction problem (Y): re-offending
Input to algorithm (X): 137 question survey
Decision: Offer Bail or Not

- Our method fairAUC allows minimal impact on accuracy while addressing bias across groups.

Feature Selection for Healthcare

Feature Selection for Healthcare

Feature Selection for Healthcare

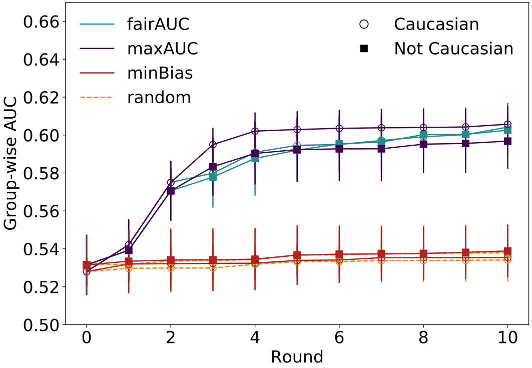
ML Problem:

Prediction problem (Y): Diabetes
Input to algorithm (X): Health variables
Decision: Offer monitoring or nutrition recommendations

Feature Selection for Healthcare

ML Problem:

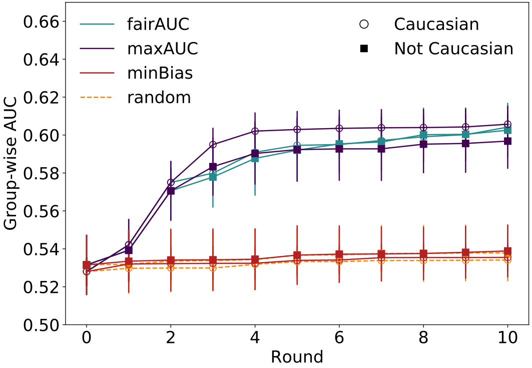
Prediction problem (Y): Diabetes
Input to algorithm (X): Health variables
Decision: Offer monitoring or nutrition recommendations



Feature Selection for Healthcare

ML Problem:

Prediction problem (Y): Diabetes
Input to algorithm (X): Health variables
Decision: Offer monitoring or nutrition recommendations



- Our method fairAUC allows minimal impact on accuracy while addressing bias across groups.

Conclusion

- Bias is increasingly an issue with algorithmic decision making

Conclusion

- Bias is increasingly an issue with algorithmic decision making
- Developing “fair” algorithms is helpful, but impact may be limited by the data features

Conclusion

- Bias is increasingly an issue with algorithmic decision making
- Developing "fair" algorithms is helpful, but impact may be limited by the data features
  - Might result in highly complex and non-explainable models

Conclusion

- Bias is increasingly an issue with algorithmic decision making
- Developing "fair" algorithms is helpful, but impact may be limited by the data features
    - Might result in highly complex and non-explainable models
- Propose an approach for a decision maker to acquire features with fairness in mind with relatively simple models

Conclusion

- Bias is increasingly an issue with algorithmic decision making
- Developing “fair” algorithms is helpful, but impact may be limited by the data features
    - Might result in highly complex and non-explainable models
- Propose an approach for a decision maker to acquire features with fairness in mind with relatively simple models
    - with provable guarantees and fairness to **ALL** groups

Conclusion

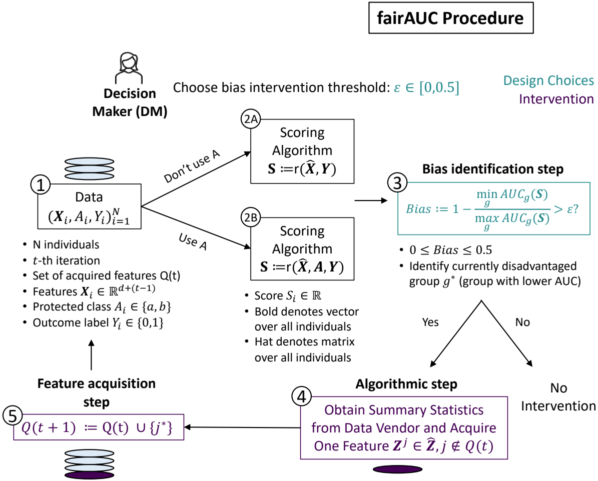
- Bias is increasingly an issue with algorithmic decision making
- Developing "fair" algorithms is helpful, but impact may be limited by the data features
  - Might result in highly complex and non-explainable models
- Propose an approach for a decision maker to acquire features with fairness in mind with relatively simple models
  - with provable guarantees and fairness to **ALL** groups
- Works well in practice in canonical applications

Conclusion

- Bias is increasingly an issue with algorithmic decision making
- Developing "fair" algorithms is helpful, but impact may be limited by the data features
  - Might result in highly complex and non-explainable models
- Propose an approach for a decision maker to acquire features with fairness in mind with relatively simple models
  - with provable guarantees and fairness to **ALL** groups
- Works well in practice in canonical applications
- More broadly, research on fairness is needed across all stages of the ML process

Thank you

Schematic – with Math



- Auxiliary feature $Z_i^j \in \mathbb{R}$, collection of auxiliary features $\widehat{\boldsymbol{Z}}$
- DM provides data vendor $\boldsymbol{S}$ and $\boldsymbol{Y}$ and data vendor provides summary statistics for each auxiliary feature
- [1] fairAUC – For g^*, calculate the AUC increase each $\boldsymbol{Z}^j$

Notation:

ε	Bias intervention threshold
N	Number of individuals
t	Iteration number
Q(t)	Set of acquired features
$\widehat{\boldsymbol{X}}$	Input features to classifier
$\boldsymbol{A}$	Protected class
$\boldsymbol{Y}$	Outcome label
$\boldsymbol{S}$	Score
r	Scoring algorithm
g^*	Disadvantaged group
$\widehat{\boldsymbol{Z}}$	Auxiliary features

[1] For other algorithms:

- maxAUC – calculate the overall increase in AUC weighted by group size each $\boldsymbol{Z}^j$ adds using Fisher's linear discriminant and acquire $\boldsymbol{Z}^j$ that generates the greatest increase, $\boldsymbol{Z}^{j^*}$
- minBias – calculate the bias each $\boldsymbol{Z}^j$ generates and acquire $\boldsymbol{Z}^j$ that minimizes the bias, $\boldsymbol{Z}^{j^*}$
- random – select a $\boldsymbol{Z}^j$ at random

Synthetic Data: Result - AUC by Group over Feature Acquisition Rounds