Secrecy, Criminal Justice, Variable Importance, and Decision Trees

Cynthia Rudin

Professor of Computer Science, Electrical and Computer Engineering, Statistical Science, and Mathematics

Duke University

How **important** is a variable?

Hold on... what does that mean?

How important is a variable ?

$f_1 \longleftarrow$ depends heavily on v

$f_2 \longleftarrow$ doesn't depend on v

Knowing how important a variable is to one model does not tell you how important it is in general.

ow important is a variable for an

ll I get the same accuracy with and without the variable?

algorithm get model f_1 ⟵ doesn’t depend on v

nove v,
algorithm get model f_2 ⟵ doesn’t depend on v

s out there exists f_3 ⟵ depends heavily on v

Knowing how important a variable is to two models does not tell you how important it is in general.

This analysis would show we don’t *need* v in order to perform well.

again:

How important is a variable to ?

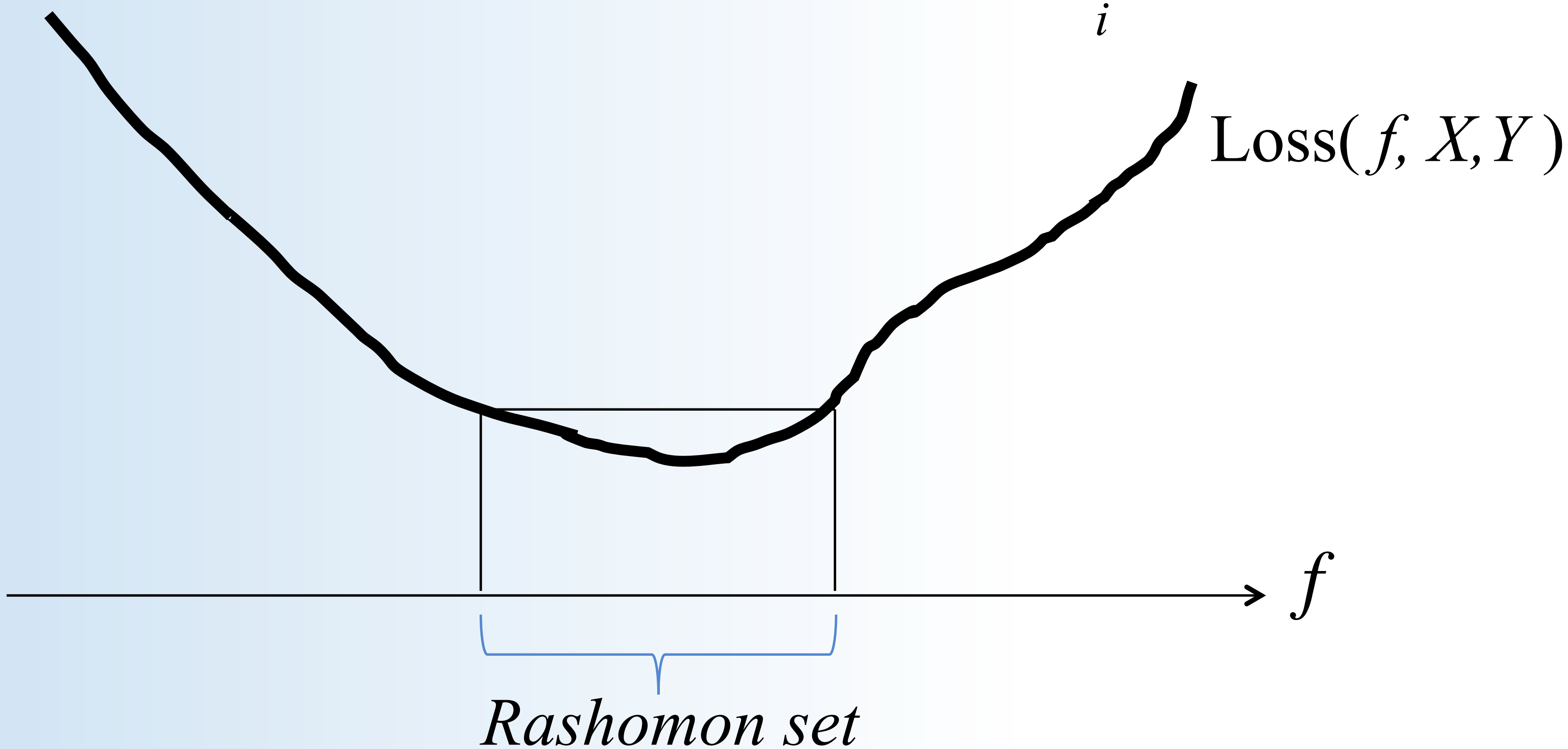
That’s more like it!

In practice, we’ll restrict to a flexible but restricted function class so we can compute and not overfit.

efine the *Rashomon set* as the set of good models within F :

$$\{f : f \in \mathcal{F} \text{ such that } \mathrm{Loss}(f, X, Y) \le \epsilon\}$$

$$\text{perhaps } \mathrm{Loss}(f, X, Y) = \sum_i (f(\mathbf{x}_i) - y_i)^2$$



ermuting a variable:

Reorde
r me!

$$X_{\text{scramble}} \quad X = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix}$$

efine *model reliance* of f on v:

$$\text{Model Reliance}(f, v) = \frac{\text{Loss}(f, X_{\text{scramble}}, Y)}{\text{Loss}(f, X, Y)}$$

If Model Reliance$(f,v)=2$, then Loss doubles if we permute v

If Model Reliance$(f,v)=1$, then f does not depend on v.

efine the *Rashomon set* as the set of good models within F :

$$\{f: f \in F \text{ such that } \mathrm{Loss}(f, X, Y) \leq \epsilon\}.$$

efine *model reliance* of f on v:

$$\text{Model Reliance}(f, v) = \frac{\mathrm{Loss}(f, X_{\mathrm{scramble}}, Y)}{\mathrm{Loss}(f, X, Y)}$$

How important is a variable to <u>*any*</u> good model?

$\approx$

What is the model reliance of functions in the Rashomon set?

Secrecy, Criminal Justice, Variable Importance, and Decision Trees

Cynthia Rudin

Professor of Computer Science, Electrical and Computer Engineering, Statistical Science, and Mathematics

Duke University

arXiv.org > stat > arXiv:1503.07810

Search or Article ID

(Help | Advanced search)

Statistics > Machine Learning

Interpretable Classification Models for Recidivism Prediction

Jiaming Zeng, Berk Ustun, Cynthia Rudin

(Submitted on 26 Mar 2015 (v1), last revised 8 Jul 2016 (this version, v6))

We investigate a long debated question, which is how to create predictive models of recidivism that are sufficiently accurate, transparent, and interpretable to use for decision-making. This question is complicated as these models are used to support different decisions, from sentencing, to determining release on probation, to allocating preventative social services. Each use case might have an objective other than classification accuracy, such as a desired true positive rate (TPR) or false positive rate (FPR). Each (TPR, FPR) pair is a point on the receiver operator characteristic (ROC) curve. We use popular machine learning methods to create models along the full ROC curve on a wide range of recidivism prediction problems. We show that many methods (SVM, Ridge Regression) produce equally accurate models along the full ROC curve. However, methods that designed for interpretability (CART, C5.0) cannot be tuned to produce models that are accurate and/or interpretable. To handle this shortcoming, we use a new method known as SLIM (Super
models along the full ROC curve. T
they are just as accurate as the mo
highly interpretable.

Comments: 45 pages, 17 figures
Subjects: Machine Learning (stat.ML
Cite as: arXiv:1503.07810 [stat.ML
(or arXiv:1503.07810v6 [s

Submission history

From: Jiaming Zeng [view email]
[v1] Thu, 26 Mar 2015 18:21:29 GMT (4
[v2] Fri, 27 Mar 2015 04:32:31 GMT (43
[v3] Fri, 13 Nov 2015 01:09:31 GMT (34
[v4] Fri, 6 May 2016 14:50:11 GMT (791
[v5] Fri, 10 Jun 2016 02:05:32 GMT (79
[v6] Fri, 8 Jul 2016 01:22:05 GMT (382k

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Journal of the Royal Statistical Society
Statistics in Society
Series A

Original Article

Interpretable classification models for recidivism predic

Jiaming Zeng, Berk Ustun, Cynthia Rudin

First published: 05 September 2016 | https://doi.org/10.1111/rssa.12227 |

Read the full text >

PDF TOC

arXiv.org > stat > arXiv:1503.07810

Search or Article ID

(Help | Advanced search)

Statistics > Machine Learning

Interpretable Classification Models for Recidivism Prediction

Jiaming Zeng, Berk Ustun, Cynthia Rudin

(Submitted on 26 Mar 2015 (v1), last revised 8 Jul 2016 (this version, v6))

We investigate a long-debated question, which is how to create predictive models of recidivism that are sufficiently accurate, transparent, and interpretable to use for decision-making. This question is complicated as these models are used to support different decisions, from sentencing, to determining release on probation, to allocating preventative social services. Each use case might have an objective other than classification accuracy, such as a desired true positive rate (TPR) or false positive rate (FPR). Each (TPR, FPR) pair is a point on the receiver operator characteristic (ROC) curve. We use popular machine learning methods to create models along the full ROC curve on a wide range of recidivism prediction problems. We show that many methods (SVM, Ridge Regression) produce equally accurate models along the full ROC curve. However, methods that designed for interpretability (CART, C5.0) cannot be tuned to produce models that are accurate and/or interpretable. To handle this shortcoming, we use a new method known as SLIM (Supersparse Linear Integer Models) to produce accurate, transparent, and interpretable models along the full ROC curve. T
they are just as accurate as the mo
highly interpretable.

Comments: 45 pages, 17 figures
Subjects: **Machine Learning (stat.MI**
Cite as: arXiv:1503.07810 [stat.MI
(or arXiv:1503.07810v6 [s

Submission history

From: Jiaming Zeng [view email]
[v1] Thu, 26 Mar 2015 18:21:29 GMT (4
[v2] Fri, 27 Mar 2015 04:32:31 GMT (431kb,D)
[v3] Fri, 13 Nov 2015 01:09:31 GMT (348kb,D)
[v4] Fri, 6 May 2016 14:50:11 GMT (791kb,D)
[v5] Fri, 10 Jun 2016 02:05:32 GMT (791kb,D)
[v6] Fri, 8 Jul 2016 01:22:05 GMT (382kb,D)

- Most ML methods have similar performance across problems, including interpretable modeling methods.
- Race is not useful for predicting recidivism, but correlated with criminal history.

PRO PUBLICA

Donate

Bernard Parker, left, was rated high risk; Dylan Fugett was rated low risk. (Josh Ritchie for ProPublica)

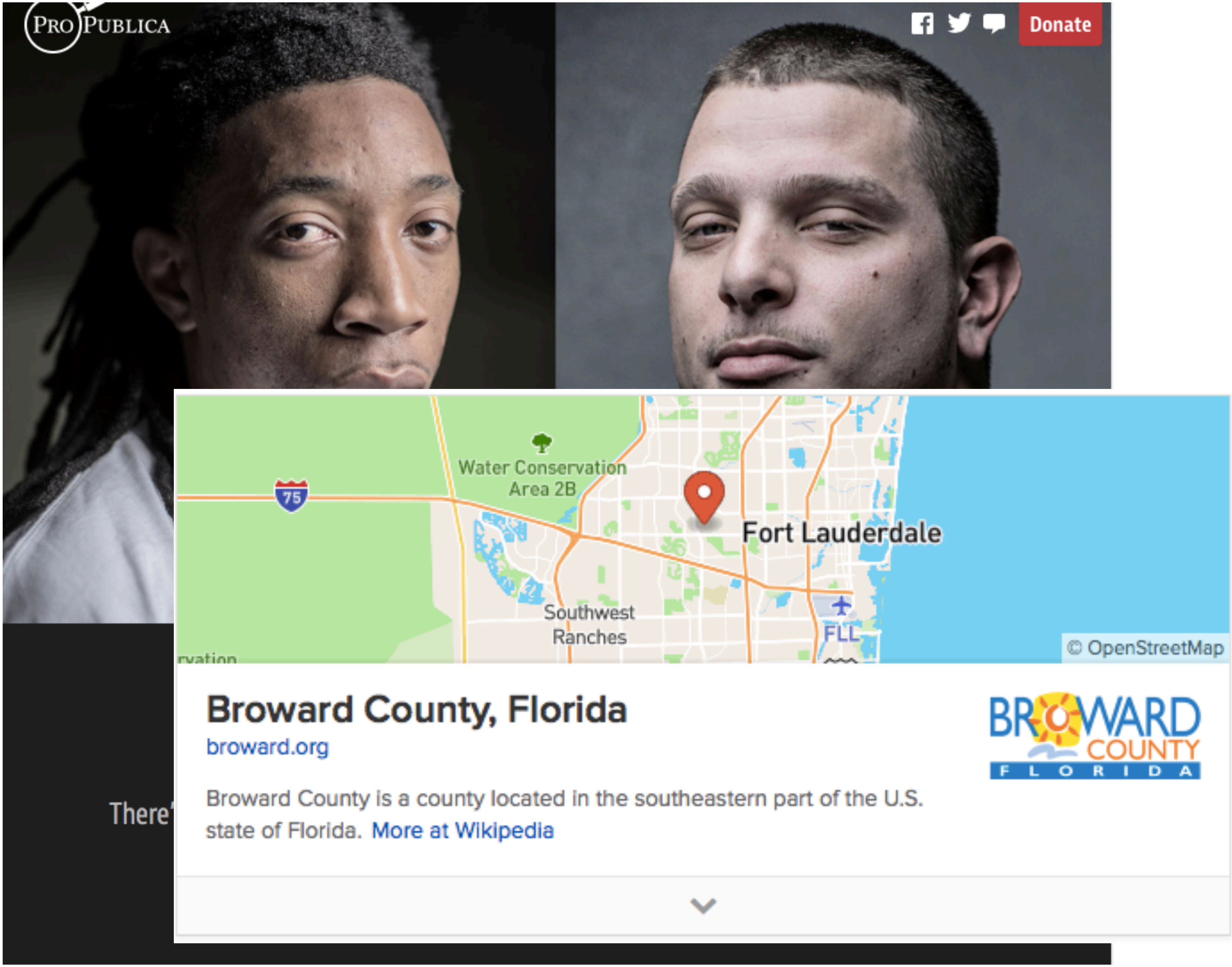
Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica

May 23, 2016

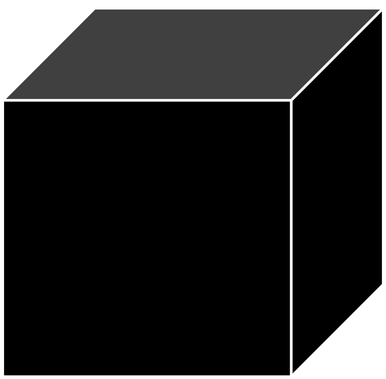
World	vs.	Zeng et al. (JRSS 2016)
Government/COMPAS: Black box is necessary.		Interpretable models are just as good
Propublica: COMPAS depends on race (after conditioning on age and criminal history).		There's no need to use race after conditioning on age and criminal history, so any reasonable model wouldn't use it.
Has $$, has *data*		Has $0



PRO PUBLICA
Donate
75
Water Conservation
Area 2B
Fort Lauderdale
Southwest
Ranches
FLL
© OpenStreetMap
Broward County, Florida
broward.org
Broward County is a county located in the southeastern part of the U.S. state of Florida. More at Wikipedia
BROWARD COUNTY
FLORIDA
There'

COMPAS vs. CORELS

OMPAS: (Correctional Offender
anagement Profiling for Alternative
nctions)

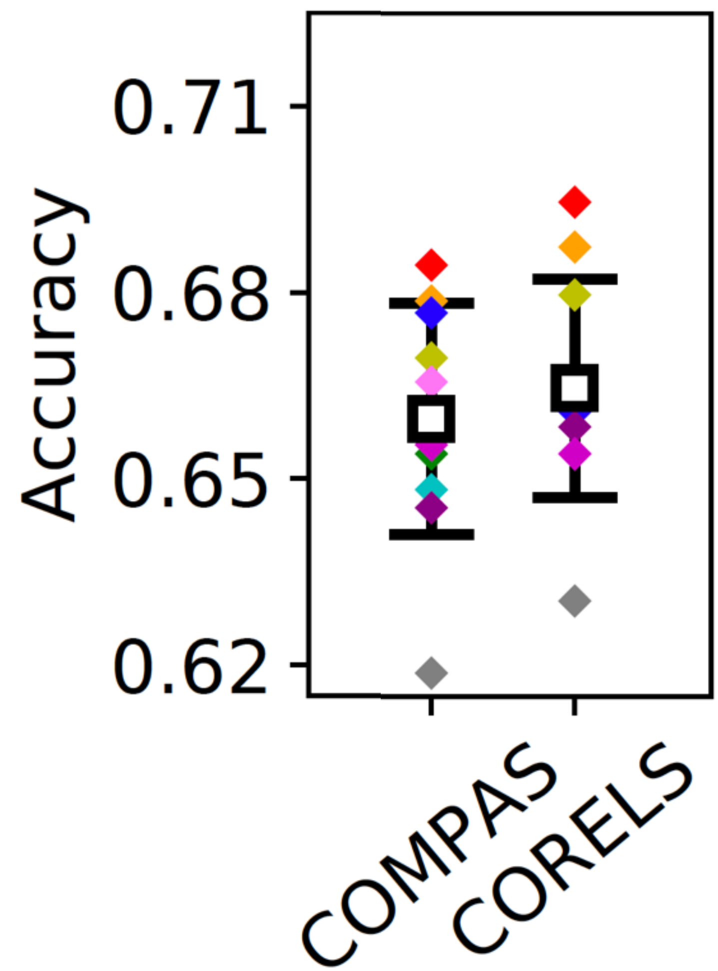


CORELS: (Certifiably Optimal RulE ListS, with Elaine Angelino, Nicholas Larus-Stone, Daniel Alabi, and Margo Seltzer, JMLR 2018)

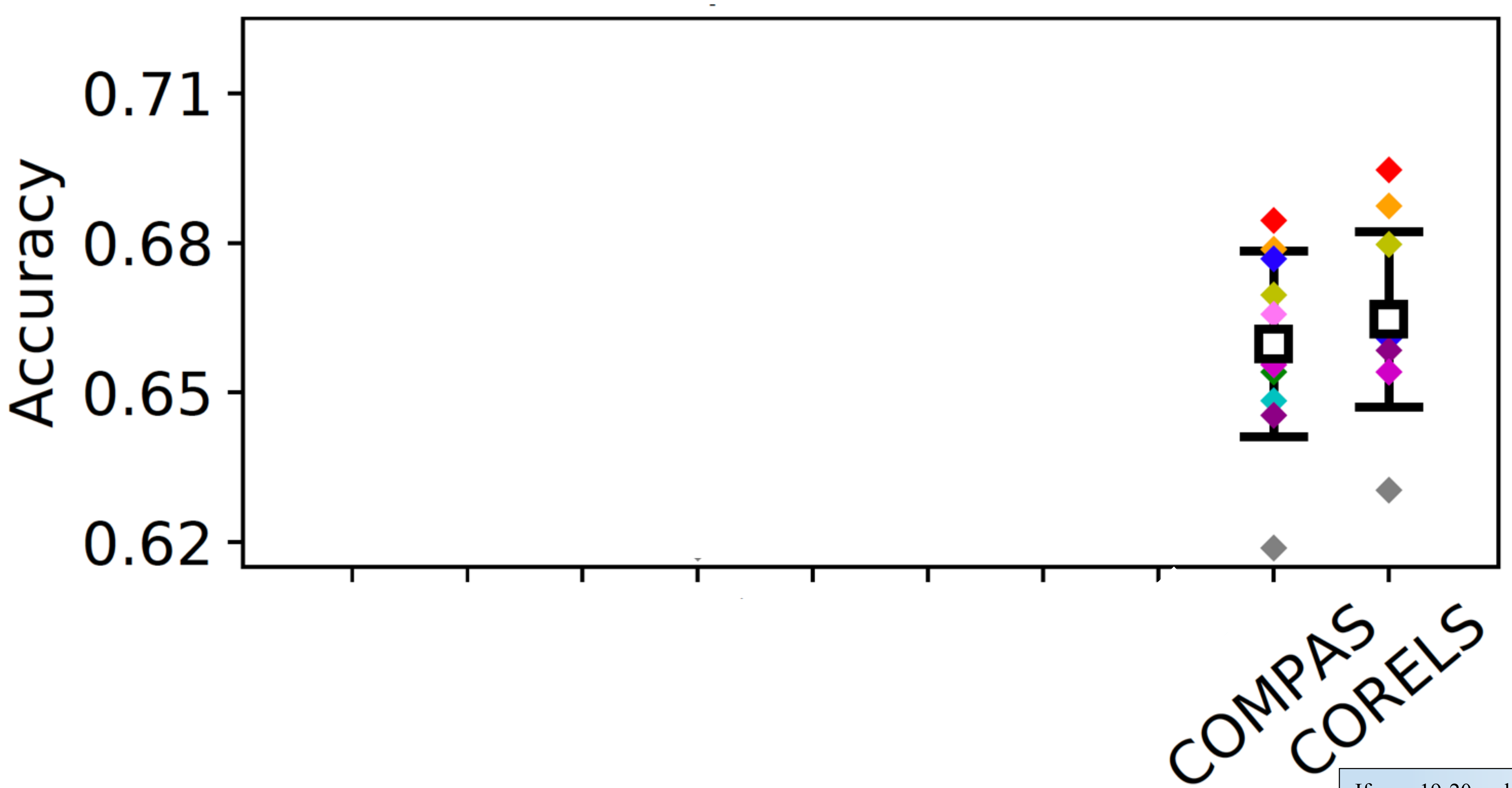
Here is the machine learning model:

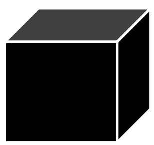
If age=19-20 and sex=male, then predict arrest
else if age=21-22 and priors=2-3 then predict arrest
else if priors >3 then predict arrest
else predict no arrest

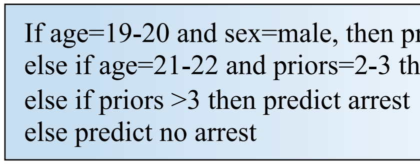
Prediction of arrest within 2 years



Prediction of arrest within 2 years







rld

vernment/COMPAS: Black
is necessary.

publica: COMPAS depends on
e (after conditioning on age
criminal history).

vs.

Zeng et al., JRSS, 2016

Interpretable models
are just as good

There's no need to use
race, so any decent
model wouldn't use it.

opublica created a *linear* model to approximate COMPAS.
efficients for age, criminal history, *and race* were all positive.

es that mean race is an important variable for COMPAS?

f_1 ⟵ depends heavily on v

f_2 ⟵ doesn't depend on v

No way! But what does COMPAS actually do?

MPAS - Correctional Offender Management Profiling for Alternative tions. By Northpointe, Inc.

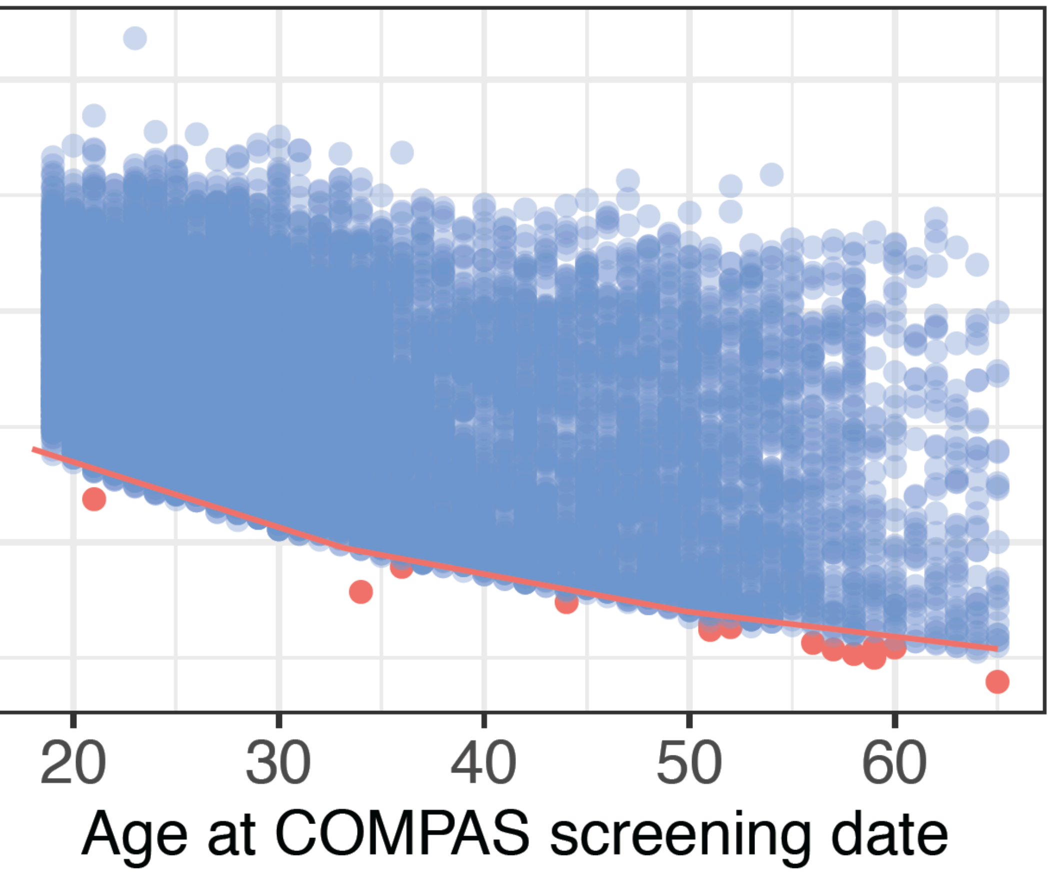
Conjecture: *The COMPAS general recidivism model is a nonlinear additive model. Its dependence on age in Broward County is approximately a linear spline, defined as follows:*

$$\begin{aligned}
\text{for ages } \leq 33.26,\ f_{\text{age}}(\text{age}) &= -0.056 \times \text{age} - 0.179 \\
\text{for ages between 33.26 and 50.02},\ f_{\text{age}}(\text{age}) &= -0.032 \times \text{age} - 0.963 \\
\text{for ages} \geq 50.02,\ f_{\text{age}}(\text{age}) &= -0.021 \times \text{age} - 1.541.
\end{aligned}$$

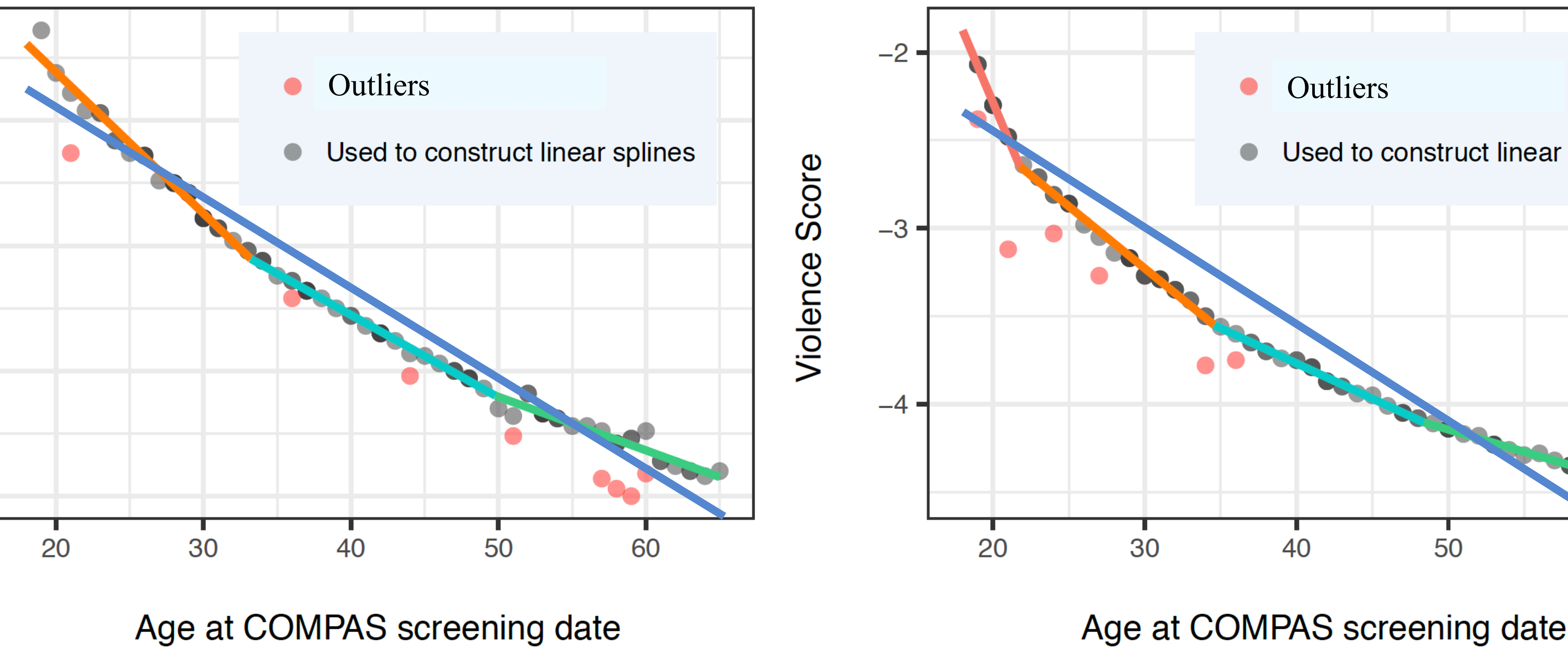
Similarly, the COMPAS violence recidivism model is a nonlinear additive model, with a dependence on age that is approximately a linear spline, defined by:

$$\begin{aligned}
\text{for ages } \leq 21.77,\ f_{\text{viol age}}(\text{age}) &= -0.205 \times \text{age} + 1.815 \\
\text{for ages between 21.77 and 34.58},\ f_{\text{viol age}}(\text{age}) &= -0.070 \times \text{age} - 1.113 \\
\text{for ages between 34.58 and 48.36},\ f_{\text{viol age}}(\text{age}) &= -0.040 \times \text{age} - 2.166 \\
\text{for ages} \geq 48.36,\ f_{\text{viol age}}(\text{age}) &= -0.025 \times \text{age} - 2.882.
\end{aligned}$$

ang, and Coker. The *Age* of Secrecy and Unfairness in Recidivism Prediction. Harvard Data Science Review (



atter plot of COMPAS scores vs age for all individuals in Broward County F



/ang, and Coker. The *Age* of Secrecy and Unfairness in Recidivism Prediction. Harvard Data Science Review (

ProPublica's analysis isn't right. But COMPAS' manual seems wrong. So does COMPAS depend on race other than through age and criminal history?

ry 1:

Subtract off (what we think is) the contribution of age to COMPAS.

Then, run machine learning methods *with and without race* to see if they need race to predict COMPAS well.

	Linear Model	Random Forest	Boosting	SVM
Without Race	0.573	0.532	0.517	0.526
With Race	0.562	0.525	0.506	0.519

able 4: RMSE of machine learning methods for predicting COMPAS general recidivism raw core after subtracting f_{age} with and without race as a feature. There is little difference with nd without race. The differences between algorithms are due to differences in model forms ge at COMPAS screening date and age at first offense are included as a features.

	Linear Model	Random Forest	Boosting	SVM
Without Race	0.498	0.493	0.468	0.475
With Race	0.489	0.482	0.456	0.474

5: RMSE of machine learning methods for predicting COMPAS violence recidivism raw after subtracting $f_{\text{viol age}}$ with and without race as a feature. Age at COMPAS screening and age at first offense are included as a features.

ProPublica's analysis isn't right. But COMPAS' manual seems wrong.
So does COMPAS depend on race other than through age and crimina
history?

ry 1:

Subtract off (what we think is) the contribution of age to COMPAS.

Then, run machine learning methods *with and without race* to see if they explicitly need race to predict COMPAS well.

Knowing how important a variable is to two models does not tell you how important it is in general.

ProPublica's analysis isn't right. But COMPAS' manual seems wrong.
So does COMPAS depend on race other than through age and crimina
history?

ry 2:

Choose a flexible model class. Find the range of Model Reliance of
functions in the Rashomon set.

efine the *Rashomon set* as the set of good models within F :

$$\{f: f \in F \text{ such that } \mathrm{Loss}(f, X, Y) \leq \epsilon\}.$$

fine *model reliance* of f on v:

$$\text{Model Reliance}(f, v) = \frac{\mathrm{Loss}(f, X_{\text{scramble}}, Y)}{\mathrm{Loss}(f, X, Y)}$$

If Model Reliance(f, v)=1, then f does not depend on v.

How important is a variable to <u>*any*</u> good model?

≈

What is the model reliance of functions in the Rashomon set?

efine the *Rashomon set* as the set of good models within F :

$$\{f : f \in F \text{ such that } \text{Loss}(f, \text{X}, \text{Y}) \le \epsilon\}.$$

fine *model reliance* of f on v:

$$\text{Model Reliance}(f, v) = \frac{\text{Loss}(f, \text{X}_{\text{scramble}}, \text{Y})}{\text{Loss}(f, \text{X}, \text{Y})}$$

fine *model class reliance* of F on v:

$$\text{odel Class Reliance}_{+}(F, v) = \max_{f \in \text{Rashomon set}(F, \epsilon)} \text{Model Reliance}(f,$$

$$\text{odel Class Reliance}_{-}(F, v) = \min_{f \in \text{Rashomon set}(F, \epsilon)} \text{Model Reliance}(f,$$

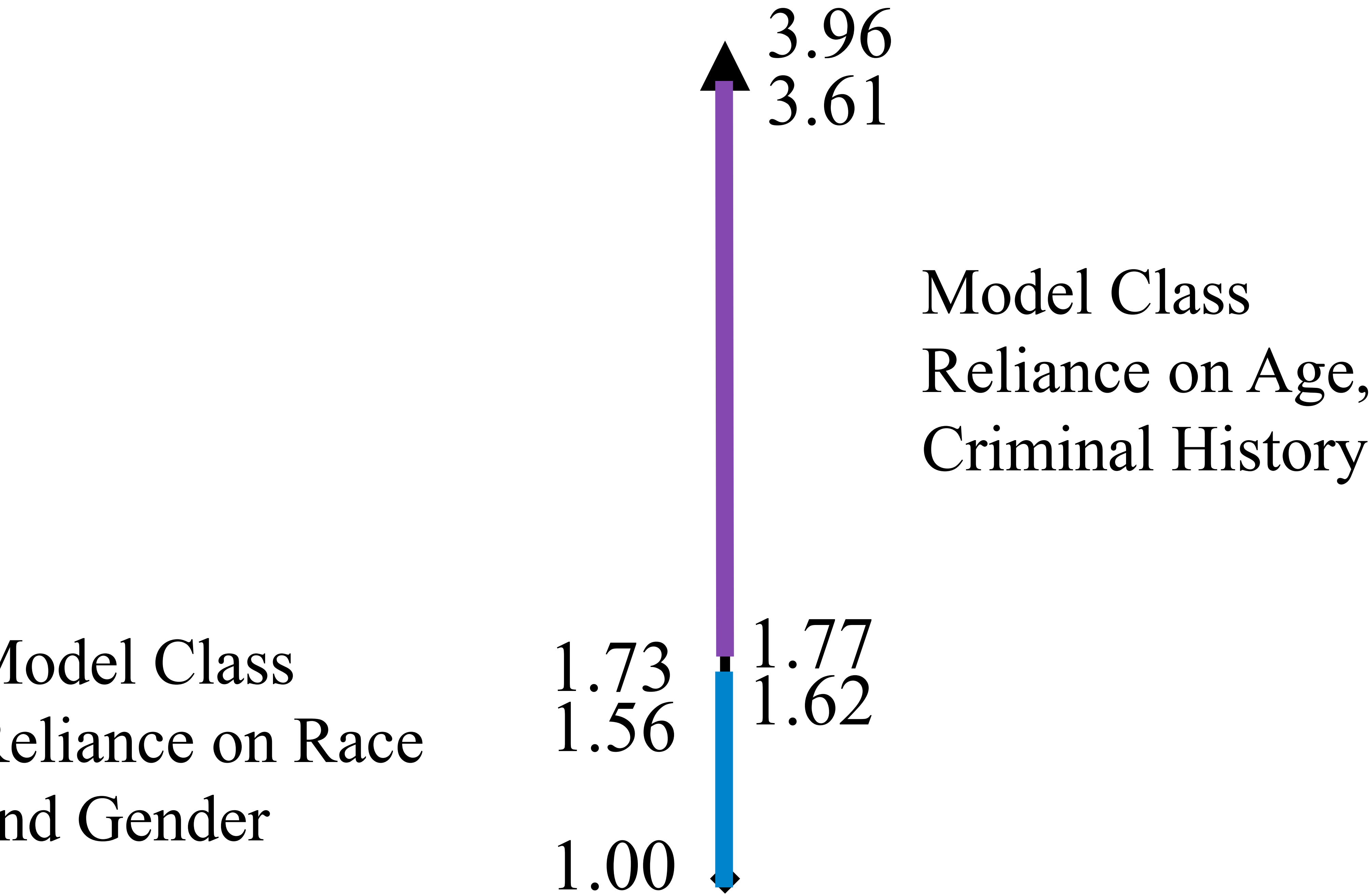
Choose a really flexible model class. Use age, criminal history, gender and race as regressors.

Choices:

- Kernel regression,
- Gaussian kernels with σ cross-validated on a training set,
- Regularized kernel weights (with parameter cross-validated)
- ε for the Rashomon Set as 0.1 × minimum cross-validated loss.

Calculate Model Class Reliance on race and gender.

Rudin, Dominici. All Models are Wrong but many are Useful: Learning a Variable's Importance by Studying a
Class of Prediction Models Simultaneously. 2018


3.96
3.61
Model Class
Reliance on Age,
Criminal History
1.77
1.73
1.62
1.56
Model Class
Reliance on Race
and Gender
1.00

World vs. Zeng et al., JRSS, 2016

Government/COMPAS: Black box is necessary.

Interpretable models are just as good

Propublica: COMPAS is racially biased.

There’s no need to use race after conditioning on age and criminal history, so any decent model wouldn’t use it.

?

The New York Times

OP-ED CONTRIBUTOR

When a Computer Program Keeps You in Jail

By Rebecca Wexler

June 13, 2017

232

to complicated proprietar

d yet.

:h survey

least one typo

I NEEVR

MAKE TYPO

e	COMPAS Violence Decile	# Priors	Selected Prior Charges	Selected Subsequent Charges
a pa	1	4	Aggravated Battery (F,1), Child Abuse (F,1), Resist Officer w/Violence (F,1)	
d er	1	14	Battery on Law Enforc Officer (F,3), Aggravated Assault W/Dead Weap (F,1), Aggravated Battery (F,1), Resist/obstruct Officer W/viol (F,1)	
y ers	1	15	Attempted Murder 1st Degree (F,1), Resist/obstruct Officer W/viol (F,1), Agg Battery Grt/Bod/Harm (F,1), Carrying Concealed Firearm (F,1)	Armed Sex Batt/vict 12 Yrs + (F,2), Aggravated Assault W/dead Weap (F,3 Kidnapping (F,1)
ando ker	1	22	Aggrav Battery w/Deadly Weapon (F,1), Driving Under The Influence (M,2), Carrying Concealed Firearm (F,1)	
en er	1	28	Robbery / Deadly Weapon (F,11), Poss Firearm Commission Felony (F,7)	
s son	1	40	Resist/obstruct Officer W/viol (F,3), Battery on Law Enforc Officer (F,2), Attempted Robbery Deadly Weapo (F,1), Robbery 1 / Deadly Weapon (F,1)	
uel zalez	2	6	Murder in the First Degree (F,1), Aggrav Battery w/Deadly Weapon (F,1), Carrying Concealed Firearm (F,1)	

Name	COMPAS Violence Decile	# Priors	Selected Prior Charges	Selected Subsequent Charges
Vilma Dieppa	1	4	Aggravated Battery (F,1), Child Abuse (F,1), Resist Officer w/Violence (F,1)	
David Selzer	1	14	Battery on Law Enforc Officer (F,3), Aggravated Assault W/Dead Weap (F,1), Aggravated Battery (F,1), Resist/obstruct Officer W/viol (F,1)	
Berry Sanders	1	15	Attempted Murder 1st Degree (F,1), Resist/obstruct Officer W/viol (F,1), Agg Battery Grt/Bod/Harm (F,1), Carrying Concealed Firearm (F,1)	Armed Sex Batt/vict 12 Yrs + (F,2), Aggravated Assault W/dead Weap (F,3), Kidnapping (F,1)
Fernando Walker	1	22	Aggrav Battery w/Deadly Weapon (F,1), Driving Under The Influence (M,2), Carrying Concealed Firearm (F,1)	
Steven Glover	1	28	Robbery / Deadly Weapon (F,11), Poss Firearm Commission Felony (F,7)	
Rufus Jackson	1	40	Resist/obstruct Officer W/viol (F,3), Battery on Law Enforc Officer (F,2), Attempted Robbery Deadly Weapo (F,1), Robbery 1 / Deadly Weapon (F,1)	
Miguel Gonzalez	2	6	Murder in the First Degree (F,1), Aggrav Battery w/Deadly Weapon (F,1), Carrying Concealed Firearm (F,1)	
William Kelly	2	17	Aggravated Assault (F,5), Aggravated Assault W/dead Weap (F,2), Shoot/throw Into Vehicle (F,2), Battery Upon Detainee (F,1)	
Richard Campbell	2	21	Armed Trafficking In Cocaine (F,1), Poss Weapon Commission Felony (F,1), Carrying Concealed Firearm (F,1)	
John Coleman	2	25	Attempt Murder in the First Degree (F,1), Carrying Concealed Firearm (F,1), Felon in Pos of Firearm or Amm (F,1)	
Oscar Pope	2	38	Aggravated Battery (F,3), Robbery / Deadly Weapon (F,3), Kidnapping (F,1), Carrying Concealed Firearm (F,2)	Grand Theft in the 3rd Degree (F,3)
Travis Spencer	3	16	Aggravated Assault W/dead Weap (F,1), Burglary Damage Property>$1000 (F,1), Burglary Unoccupied Dwelling (F,1)	
Michael Avila	3	17	Aggravated Assault W/dead Weap (F,2), Aggravated Assault w/Firearm (F,2), Discharge Firearm From Vehicle (F,1), Home Invasion Robbery (F,1)	Fail Register Vehicle (M,2)

Richard Campbell	2	21	Armed Trafficking In Cocaine (F,1), Poss Weapon Commission Felony (F,1), Carrying Concealed Firearm (F,1)	
John Coleman	2	25	Attempt Murder in the First Degree (F,1), Carrying Concealed Firearm (F,1), Felon in Pos of Firearm or Amm (F,1)	
Oscar Pope	2	38	Aggravated Battery (F,3), Robbery / Deadly Weapon (F,3), Kidnapping (F,1), Carrying Concealed Firearm (F,2)	Grand Theft in the 3rd Degree (F,3)
Travis Spencer	3	16	Aggravated Assault W/dead Weap (F,1), Burglary Damage Property>$1000 (F,1), Burglary Unoccupied Dwelling (F,1)	
Michael Avila	3	17	Aggravated Assault W/dead Weap (F,2), Aggravated Assault w/Firearm (F,2), Discharge Firearm From Vehicle (F,1), Home Invasion Robbery (F,1)	Fail Register Vehicle (M,2)
Terrance Murphy	3	20	Solicit to Commit Armed Robbery (F,1), Armed False Imprisonment (F,1), Home Invasion Robbery (F,1)	Driving While License Revoked (F,3)
Anthony Hawthorne	3	25	Attempt Sexual Batt / Vict 12+ (F,1), Resist/obstruct Officer W/viol (F,1), Poss Firearm W/alter/remov Id# (F,1)	
Stephen Brown	3	36	Carrying Concealed Firearm (F,2), Battery On Law Enforce Officer (F,1), Kidnapping (F,1), Aggravated Battery (F,1)	Driving While License Revoked (F,3)
Samuel Walker	3	36	Murder in the First Degree (F,1), Poss Firearm Commission Felony (F,1), Solicit to Commit Armed Robbery (F,1)	Petit Theft 100−300 (M,1)
Jesse Bernstein	4	10	Aggravated Battery / Pregnant (F,1), Sex Battery Vict Mental Defect (F,1), Shoot/throw In Occupied Dwell (F,1)	Tresspass in Struct/Convey Occupy (M,1)
Shandedra Hardy	4	16	Aggrav Battery w/Deadly Weapon (F,1), Felon in Pos of Firearm or Amm (F,4)	Resist/Obstruct W/O Violence (M,1), Possess Drug Paraphernalia (M,1)

urrent State of Affairs

The Mercury News

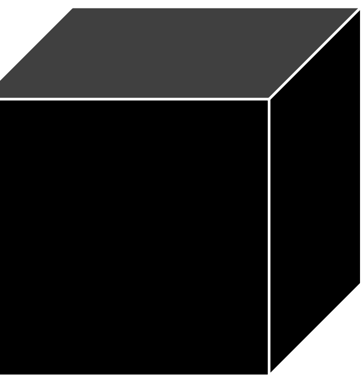
News > California News • News

Law ending cash bail in California halted after referendum qualifies for 2020 ballot

SB 10 won’t start in October as planned; final version divided legisla rights groups who initially supported it

- Iuman judges are biased black b
- COMPAS is a black box, possibl eliable. Financial incentives to u
- California is moving towards a n sing COMPAS.
- roPublica’s seriously flawed analysis still highly cited and espected
- Academic interest in fairness is huge, interest in explainability f black boxes is huge…
- ttle interest/expertise in interpretability

COMPAS:



CORELS Model:

If age=19-20 and sex=male, then predict arrest
else if age=21-22 and priors=2-3 then predict ar
else if priors >3 then predict arrest
else predict no arrest

Behind the scenes

Model Class Reliance

Optimal Decision Trees

efine the *Rashomon set* as the set of good models within F :

$$\{f: f \in F \text{ such that } \mathrm{Loss}(f, \mathrm{X}, \mathrm{Y}) \leq \epsilon\}.$$

fine *model reliance* of f on v:

$$\text{Model Reliance}(f, v) = \frac{\mathrm{Loss}(f, \mathrm{X}_{\mathrm{scramble}}, \mathrm{Y})}{\mathrm{Loss}(f, \mathrm{X}, \mathrm{Y})}$$

efine *model class reliance* of F on v:

$$\text{Model Class Reliance}_{+}(F, v) = \max_{f \in \text{Rashomon set }(F, \epsilon)} \text{Model Reliance}($$

$$\text{Model Class Reliance}_{-}(F, v) = \min_{f \in \text{Rashomon set }(F, \epsilon)} \text{Model Reliance}($$

er et al. (in progress, 2019) contains:

- Estimation using U-statistics
- Learning theoretic bounds
- How to compute MCR efficiently (linear, additive, reproducing kernel Hilbert space
- Connections to causal inference

efine *model class reliance* of F on v:

$$\text{Model Class Reliance}_{+}(F, v) = \max_{f \in \text{Rashomon set }(F, \epsilon)} \text{Model Reliance}($$

$$\text{Model Class Reliance}_{-}(F, v) = \min_{f \in \text{Rashomon set }(F, \epsilon)} \text{Model Reliance}($$

Behind the scenes

Model Class Reliance

Optimal Decision Trees

CORELS: Certifiably Optimal Rule Lists (JMLR, 2018)

Predictive model for 2 yr recidivism, built from data from Broward County Florida

age between 18-20 and sex is male	THEN predict arrest within 2 years
LSE IF age between 21-23 and 2-3 prior offenses	THEN predict arrest
LSE IF more than three priors	THEN predict arrest
LSE	predict no arrest.

in <10 seconds, certified to optimality in ~2 minutes, over the space of rule lists.

Server with two Intel Xeon E5-2699 v4 (55 MB cache, 2.20 GHz) processors and 448 GB RAM

$$R(d, \mathbf{x}, \mathbf{y}) = \underbrace{\frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{[\text{point } i \text{ is misclassified}]}}_{\text{Misclassification error}} + \lambda \underbrace{\#\text{Rules in list}}_{\text{Sparsity}}$$

Usually 0.01

IF age between 18-20 and sex is male THEN predict arrest within 2 yea
ELSE IF age between 21-23 and 2-3 prior offenses THEN predict arrest
ELSE IF more than three priors THEN predict arrest
ELSE predict no arrest.

Several theorems led to bounds

heorem 1: If a rule's support is less than λ, that rule cannot be in an optimal rule lis

heorem 2: If a rule in the list does not correctly classify at least λ fraction of
bservations, that rule cannot be in an optimal rule list.

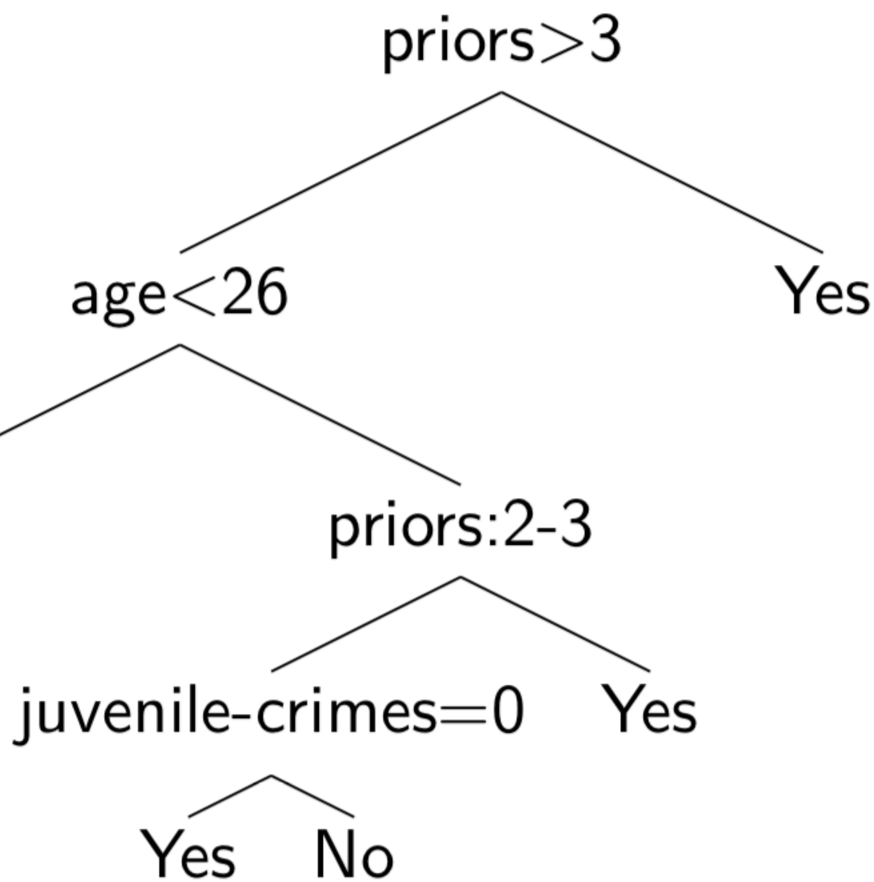
heorem 3: The length of an optimal rule list is bounded by a function of λ,
the accuracy of the current best model so far, and
the accuracy and length of our current prefix (partial rule list).

heorem 4: One-step-lookahead bound: If a prefix's lower bound is within λ of the
est current objective, adding any rules to it will lead to a non-optimal rule list.

eorem 5: Equivalent Points Bound: For every set of "equivalent" points, we will
ssify at least the minority label of them wrong.

heorem 6: Permutation Bound: Only an optimal permutation of a set of rules can be
tended to form an optimal rule list.

urrently…



Rudin, Seltzer. Optimal Sparse Decision Trees, NeurIPS (spotlight) 2019]

Behind the scenes

Model Class Reliance: https://github.com/aaronjfisher/mcr

, Rudin, Dominici. All Models are Wrong but many are Useful: Learning a Variable's Importance by Studying
Class of Prediction Models Simultaneously. https://arxiv.org/abs/1801.01489 , In Progress, 2019

Optimal Decision Trees: https://corels.eecs.harvard.edu

no et al. Certifiably Optimal Rule Lists for Categorical Data, Journal of Machine Learning Research, 2018.

udin, Seltzer. Optimal Sparse Decision Trees. https://arxiv.org/abs/1904.12847 , NeurIPS, 2019

ine Wang

au Coker

of Secrecy and
s in Recidivism
n, Harvard Data
eview, 2019

Interpretable Classification Models
For Recidivism Prediction, Journal of
the Royal Statistical Society, 2016

Jiaming Zeng

Berk Ustun

Elaine Angelino

Nicholas Laru

Daniel Alabi

Certifiably Optimal Rule Lists for
Categorical Data, Journal of
Machine Learning Research, 2018.

Aaron Fisher

Francesca Dominici

Margo

All Models are Wrong but many are Useful: Variable Importance for Black-Box,
Proprietary, or Misspecified Prediction Models, using Model Class Reliance, 2018

Systems Techniques

- Custom bit-vector library for rule list evaluation
- Computational reuse for evaluating multiple lists with similar prefixes
- Priority queue

Data structures: trie (prefix tree), symmetry-aware map, and queue

Mine all rules with sufficient support.

Start with rule lists of size 1.

While queue of rule lists is not empty:

Take current prefix from queue, consider each of its children and check:

- length bound
- rule accuracy
- one step-lookahead bound
- equivalent points bound
- symmetry-aware pruning

If lower bound is higher than current best, prefix is no good.

Otherwise add it into queue.

If current objective is lower than current best, update and store rule list.

End while

Output is optimal rule list (with certificate of optimality)

anks

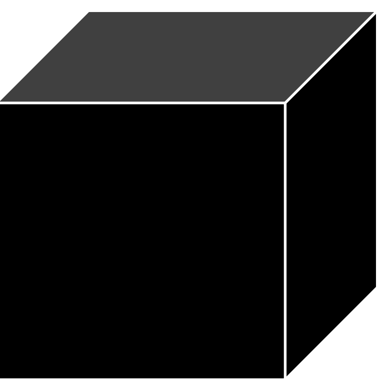
ia Rudin, Caroline Wang, and Beau Coker. The *Age* of Secrecy and Unfairness in Recidivism Prediction. 2018

, Rudin, Dominici. All Models are Wrong but many are Useful: Variable Importance for Black-Box, Proprietar
ecified Prediction Models, using Model Class Reliance. 2018

ng Zeng, Berk Ustun, Cynthia Rudin. Interpretable Classification Models For Recidivism Prediction. Journal o
Statistical Society, 2016.

Angelino, Nicholas Larus-Stone, Daniel Alabi, Margo Seltzer, and Cynthia Rudin, Certifiably Optimal Rule L
tegorical Data, Journal of Machine Learning Research, 2018.

COMPAS:



CORELS Model:

If age=19-20 and sex=male, then predict arrest
else if age=21-22 and priors=2-3 then predict ar
else if priors >3 then predict arrest
else predict no arrest