Scoring Systems: At the Extreme of Interpretable Machine Learning

Cynthia Rudin

Professor of Computer Science, Electrical and Computer Engineering
Statistical Science, and Biostatistics & Bioinformatics

Duke University

Can a typographical error lead to years of extra prison time?

Can a typographical error lead to years of extra prison time?

The New York Times

OP-ED CONTRIBUTOR

When a Computer Program Keeps You in Jail

By Rebecca Wexler

June 13, 2017

232

Glenn Rodriguez was denied parole because of a miscalculated “COMPAS” score.

- A black box model is a formula that is either too complicated for any human to understand or is proprietary.
- An interpretable machine learning model obeys a domain-specific set of constraints so that humans can better understand it.

- High-stakes decisions or troubleshooting
  - Criminal justice models, credit scoring, air pollution, airplane maintenance, many healthcare applications – anything high stakes

What happens when we use a black box?

THE SACRAMENTO BEE

How bad is Sacramento's air, exactly? Google results appear at odds with reality, some say

BY MICHAEL MCGOUGH
AUGUST 07, 2018 09:26 AM, UPDATED AUGUST 07, 2018 09:26 AM

Smoke is affecting air quality all over California. Here's what it looks like at the Carr Fire, north of Redding, on July 31, 2018.
BY PAUL KITAGAKI JR.

Innovate Healthcare
HealthImaging
INSIGHTS IN IMAGING & INFORMATICS

NEWS TOPICS CONFERENCES EXPERIENCE STORIES SUBSCRIBE

Algorithm's 'unexpected' weakness raises larger concerns about AI's potential in broader populations

Matt O'Connor | April 05, 2021 | Artificial Intelligence

Deep learning detects intercranial hemorrhages

And this is the tip of the iceberg…

The New York Times

OP-ED CONTRIBUTOR

When a Computer Program Keeps You in Jail

By Rebecca Wexler

June 13, 2017

232

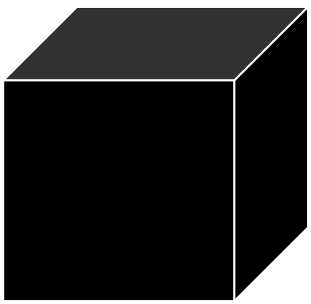
Glenn Rodriguez was denied parole because of a miscalculated “COMPAS” score.

How accurate is COMPAS? Data from Florida can tell us…

COMPAS vs. CORELS

COMPAS: (Correctional Offender Management Profiling for Alternative Sanctions)

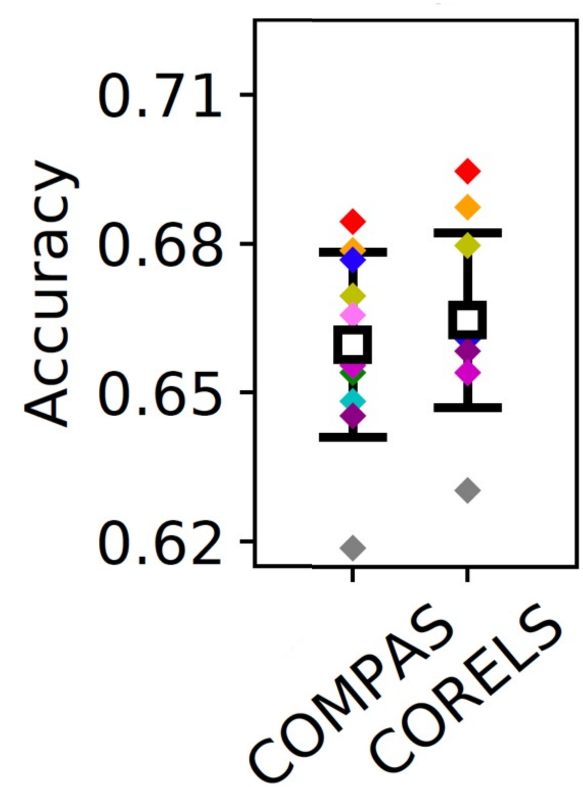
CORELS: (Certifiably Optimal RulE ListS, with Elaine Angelino, Nicholas Larus-Stone, Daniel Alabi, and Margo Seltzer, KDD 2017 & JMLR 2018)

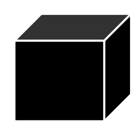


Here is the machine learning model:

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

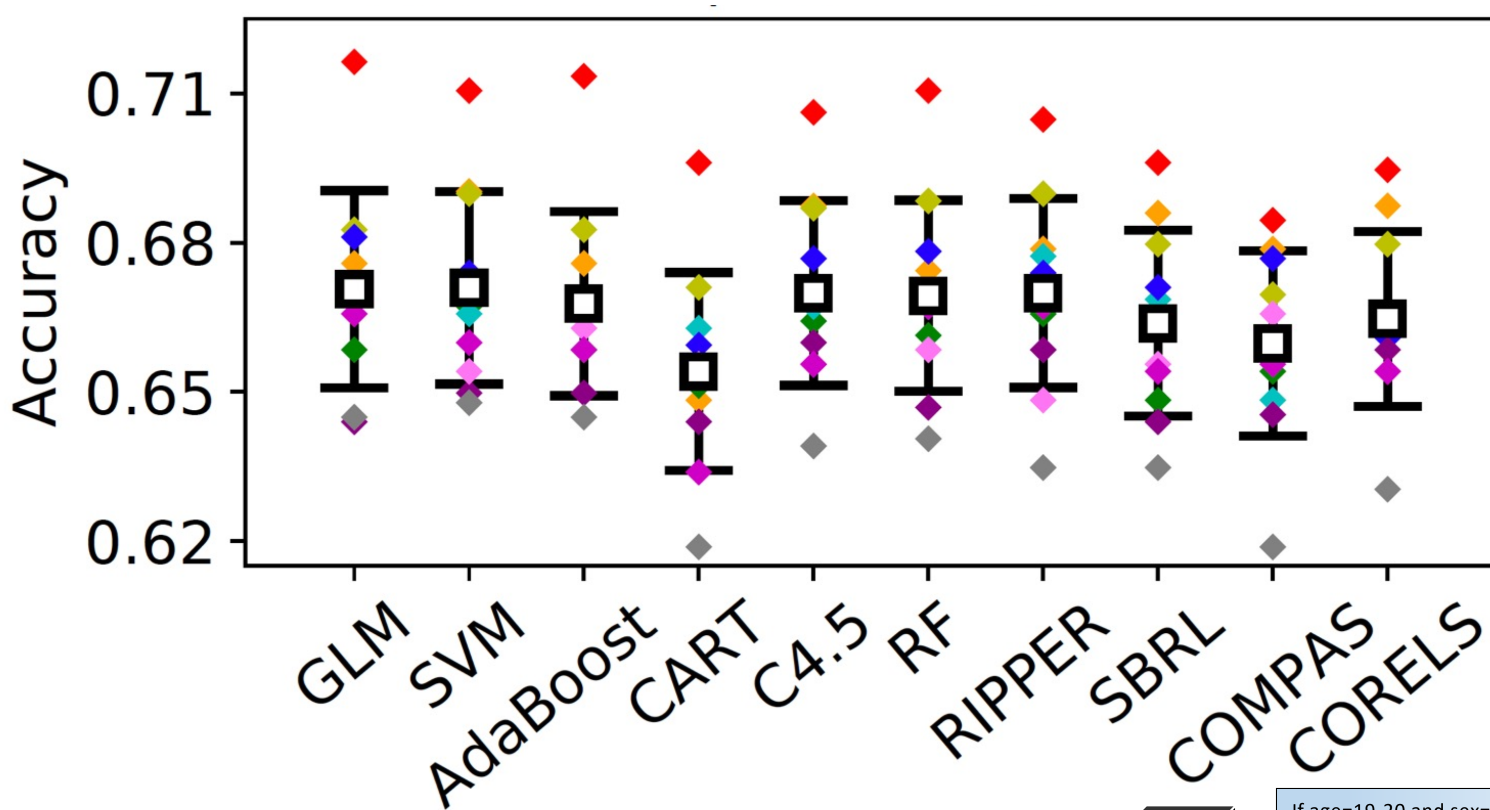
Prediction of re-arrest within 2 years

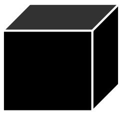




If age=19-20 and sex=male, then predict arrest
else if age=21-22 and priors=2-3 then predict arrest
else if priors >3 then predict arrest
else predict no arrest

Prediction of re-arrest within 2 years



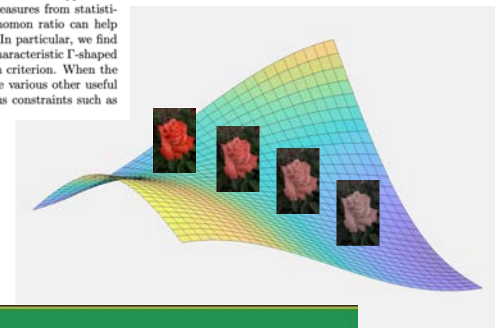


If age=19-20 and sex=male, then predict arrest
else if age=21-22 and priors=2-3 then predict arrest
else if priors >3 then predict arrest
else predict no arrest

Problem spectrum

age   45
congestive heart failure?   yes
takes aspirin
smoking? no
gender   M
exercise? yes
allergies? no
number of past strokes   2
diabetes? yes

The *Rashomon effect* occurs when many different explanations exist for the same phenomenon. In machine learning, Leo Breiman used this term to characterize problems where many accurate-but-different models exist to describe the same data. In this work, we study how the Rashomon effect can be useful for understanding the relationship between training and test performance, and the possibility that simple-yet-accurate models exist for many problems. We consider the *Rashomon set*—the set of almost-equally-accurate models for a given problem—and study its properties and the types of models it could contain. We present the *Rashomon ratio* as a new measure related to simplicity of model classes, which is the ratio of the volume of the set of accurate models to the volume of the hypothesis space; the Rashomon ratio is different from standard complexity measures from statistical learning theory. For a hierarchy of hypothesis spaces, the Rashomon ratio can help modelers to navigate the trade-off between simplicity and accuracy. In particular, we find empirically that a plot of empirical risk vs. Rashomon ratio forms a characteristic Γ-shaped *Rashomon curve*, whose elbow seems to be a reliable model selection criterion. When the Rashomon set is large, models that are accurate—but that also have various other useful properties—can often be obtained. These models might obey various constraints such as interpretability, fairness, or monotonicity.



Tabular:

- many problems in criminal justice, healthcare, social sciences, equipment reliability & maintenance, etc.
- features include counts, categorical data

Raw:

- pixels/voxels, words, parts of sound waves

Problem spectrum

Very sparse models (trees, scoring systems)

With minor pre-processing, all methods have similar performance

Neural networks

Tabular:

- many problems in criminal justice, healthcare, social sciences, equipment reliability & maintenance, etc.
- features include counts, categorical data

Raw:

- pixels/voxels, words, parts of sound waves

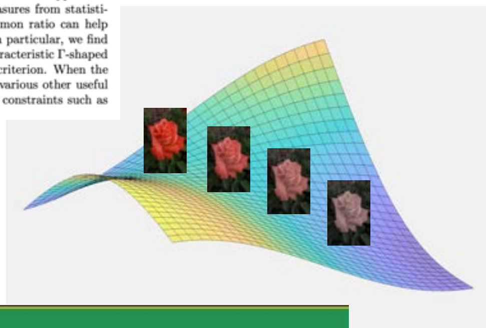
Problem spectrum

age 45
congestive heart failure? yes
takes aspirin
smoking? no
gender M
exercise? yes
allergies? no
number of past strokes 2
diabetes? yes

Tabular:

- many problems in criminal justice, healthcare, social sciences, equipment reliability & maintenance, etc.
- features include counts, categorical data

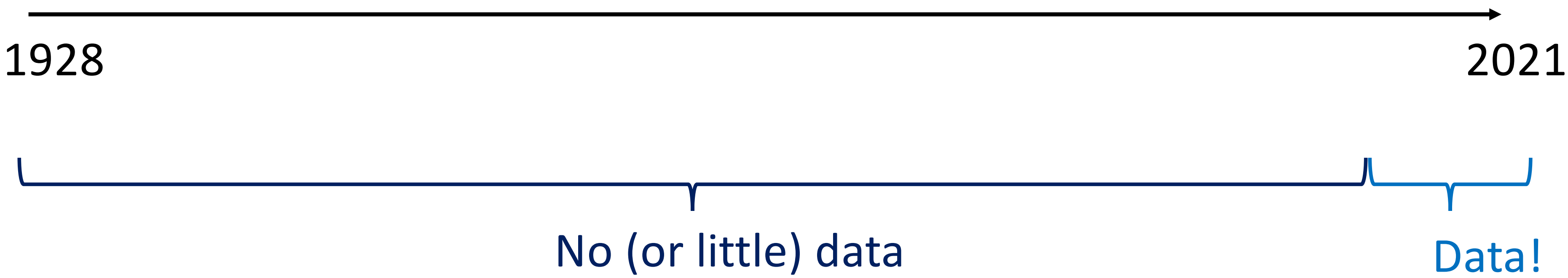
The *Rashomon effect* occurs when many different explanations exist for the same phenomenon. In machine learning, Leo Breiman used this term to characterize problems where many accurate-but-different models exist to describe the same data. In this work, we study how the Rashomon effect can be useful for understanding the relationship between training and test performance, and the possibility that simple-yet-accurate models exist for many problems. We consider the *Rashomon set*—the set of almost-equally-accurate models for a given problem—and study its properties and the types of models it could contain. We present the *Rashomon ratio* as a new measure related to simplicity of model classes, which is the ratio of the volume of the set of accurate models to the volume of the hypothesis space; the Rashomon ratio is different from standard complexity measures from statistical learning theory. For a hierarchy of hypothesis spaces, the Rashomon ratio can help modelers to navigate the trade-off between simplicity and accuracy. In particular, we find empirically that a plot of empirical risk vs. Rashomon ratio forms a characteristic Γ-shaped *Rashomon curve*, whose elbow seems to be a reliable model selection criterion. When the Rashomon set is large, models that are accurate—but that also have various other useful properties—can often be obtained. These models might obey various constraints such as interpretability, fairness, or monotonicity.



Raw:

- pixels/voxels, words, parts of sound waves

Predictive modeling over the last century



Scoring systems

CHADS2 Score (Gage et al., 2001)

The most widely-used predictive model in healthcare?

Not an ML model

1.	*Congestive Heart Failure*	1 point		$\cdots$
2.	*Hypertension*	1 point	+	$\cdots$
3.	*Age* ≥ 75	1 point	+	$\cdots$
4.	*Diabetes Mellitus*	1 point	+	$\cdots$
5.	*Prior Stroke or Transient Ischemic Attack*	2 points	+	$\cdots$
	ADD POINTS FROM ROWS 1–5	**SCORE**	=	$\cdots$

SCORE	0	1	2	3	4	5	6
STROKE RISK	1.9%	2.8%	4.0%	5.9%	8.5%	12.5%	18.2%

Burgess. Factors determining success or failure on parole. 1928

Accordingly, twenty-one factors were selected by which each man was graded, in comparison with the average for the 1,000 cases, upon the probabilities of making good or of failing upon parole.

1 point if person has **social type** with **below average** parole violation rate

Social Type	Violation Rate
All persons	26.5%
Ne'er-do-well	25.6
Mean citizen	30.0
Drunkard	38.9
Gangster	23.2
Recent immigrant	16.7
Farm boy	10.2
Drug addict	66.7

total score over all 21 significant factors predicts **success at parole**

Points for Number of Factors	Per Cent Non-violators of Parole
16-21	98.5
14-15	97.8
13	91.2
12	84.9
11	77.3
10	65.9
7-9	56.1
5-6	32.9
2-4	24.0

Burgess. Factors determining success or failure on parole. 1928

Pennsylvania Commission on Sentencing, 2013

FACTOR	Score *
Gender	
Female	0
Male	1
Age	
Less than 24	3
24-29	2
30-49	1
50+	0
County	
Rural counties	0
Smaller, urban count	1
Allegheny and Philadelphia Counties	2
Total number of prior arrests	
0	0
1	1
2 to 4	2
5 to 12	3
13+	4
Prior property arrests	
No	0
Yes	1
Prior drug arrests	
No	0
Yes	1
Property offender	
No	0
Yes	1
Offense gravity score (OGS)	
4+	0

Risk score	N	% Arrested
0	3	0.0
1	47	17.0
2	181	9.9
3	436	23.6
4	737	24.8
5	1,036	32.4
6	1,067	40.7
7	1,434	47.2
8	1,934	55.5
9	2,103	62.3
10	1,829	69.9
11	1,098	72.2
12	278	79.1
13	25	80.0
14	3	66.7

Violence Risk Appraisal Guide (Quinsey et al, 2006)

1. Lived with both biological parents to age 16 (except for death of parent):
Yes .... -2
No .... +3
Evidence:

2. Elementary School Maladjustment:
No Problems.. -1
Slight (Minor discipline or attendance) or Moderate Problems............................ +2
Severe Problems (Frequent disruptive behavior and/or attendance or behavior resulting in expulsion or serious suspensions) .... +5
(Same as CATS Item)

3. History of alcohol problems *(Check if present):*
˜ Parental Alcoholism ˜ Teenage Alcohol Problem
˜ Adult Alcohol Problem ˜ Alcohol involved in prior offense
˜ Alcohol involved in index offense
No boxes checked................................... -1
1 or 2 boxes checked 0
3 boxes checked .... +1
4 or 5 boxes checked .... +2
Evidence:

4. Marital status (at the time of or prior to index offense):
Ever married (or lived common law in the same home for at least six months) .... -2
Never married... +1
Evidence:

5. Criminal history score for nonviolent offenses prior to the index offense
Score 0 .... -2
Score 1 or 2.. 0
Score 3 or above .... +3
(from the Cormier-Lang system, see below)

6. Failure on prior conditional release (includes parole or probation violation or revocation, failure to comply, bail violation, and any new arrest while on conditional release):
No..0
Yes .... +3
Evidence:

7. Age at index offense
Enter Date of Index Offense: ___/___/_____
Enter Date of Birth: ___/___/_____
Subtract to get Age:
39 or over .... -5
34 - 38 .... -2
28 - 33 .... -1
27 0
26 or less.. +2

8. Victim Injury (for index offense; the most serious is scored):
Death... -2
Hospitalized..0
Treated and released.............................. +1
None or slight (includes no victim).......... +2
Note: admission for the gathering of forensic evidence only is NOT considered as either treated or hospitalized; ratings should be made based on the degree of injury.
Evidence:

9. Any female victim (for index offense)
Yes .... -1
No (includes no victim)............................ +1
Evidence:

10. Meets DSM criteria for any personality disorder (must be made by appropriately licensed or certified professional)
No... -2
Yes .... +3
Evidence:

11. Meets DSM criteria for schizophrenia (must be made by appropriately licensed or certified professional)
Yes .... -3
No .... +1
Evidence:

12. a. Psychopathy Checklist score (if available, otherwise use item 12.b. CATS score)........
4 or under .... -3
5 – 9... -3
10-14 .... -1
15-24 0
25-34 .... +4
35 or higher .... +12
Note: If there are two or more PCL scores, average the scores.
Evidence:

12. b. CATS score (from the CATS worksheet)
0 or 1 .... -3
2 or 3 0
4 ....+2
5 or higher .... +3

12. WEIGHT (Use the highest circled weight from 12 a. or 12 b.) .... _____

TOTAL VRAG SCORE (SUM CIRCLED SCORES FOR ITEMS 1 – 11 PLUS THE WEIGHT FOR ITEM 12): _________

VRAG Score	Category of Risk
-24	Low
-23	Low
-22	Low
-20	Low
-19	Low
-18	Low
-17	Low
-16	Low
-15	Low
-14	Low
-13	Low
-12	Low
-11	Low
-10	Low
-9	Low
-8	Low
-7	Medium
-6	Medium
-5	Medium
-4	Medium
-3	Medium
-2	Medium
-1	Medium
0	Medium
1	Medium
2	Medium
3	Medium
4	Medium
5	Medium
6	Medium
7	Medium
8	Medium
9	Medium
10	Medium
11	Medium
2	Medium
13	Medium
14	High
15	High
16	High
17	High
18	High
19	High
20	High
21	High
22	High
23	High
24	High
25	High
26	High
28	High
32	High

Violence Risk Appraisal Guide
(Quinsey et al, 2006)

1. Lived with both biological parents to age 16 (except for death of parent):
Yes .... -2
No .... +3
Evidence:

2. Elementary School Maladjustment:
No Problems... -1
Slight (Minor discipline or attendance) or Moderate Problems............................ +2
Severe Problems (Frequent disruptive behavior and/or attendance or behavior resulting in expulsion or serious suspensions) .... +5
(Same as CATS Item)

3. History of alcohol problems *(Check if present):*
˜ Parental Alcoholism ˜ Teenage Alcohol Problem
˜ Adult Alcohol Problem ˜ Alcohol involved in prior offense
˜ Alcohol involved in index offense
No boxes checked.................................. -1
1 or 2 boxes checked 0
3 boxes checked .... +1
4 or 5 boxes checked .... +2
Evidence:

4. Marital status (at the time of or prior to index

8. Victim Injury (fo
serious is scored):
Death...................
Hospitalized.........
Treated and releas
None or slight (inc
Note: admission for
evidence only is NO
treated or hospitalize
made based on the
Evidence:

9. Any female victi
Yes ....
No (includes no vi
Evidence:

10. Meets DSM cri
disorder (must be
licensed or certifie
No.........................
Yes ....
Evidence:

11. Meets DSM cri
be made by appro
certified professio

VRAG Score	Category of Risk
-24	Low
-23	Low
-22	Low
-20	Low
-19	Low
-18	Low
-17	Low
-16	Low
-15	Low
-14	Low
-13	Low
-12	Low
-11	Low
-10	Low
-9	Low
-8	Low
-7	Medium
-6	Medium
-5	Medium
-4	Medium
-3	Medium
-2	Medium

Violence Risk Appraisal Guide (Quinsey et al, 2006)

1. Lived with both biological parents to age 16 (except for death of parent):
Yes .... -2
No .... +3
Evidence:

2. Elementary School Maladjustment:
No Problems.. -1
Slight (Minor discipline or attendance) or Moderate Problems............................ +2
Severe Problems (Frequent disruptive behavior and/or attendance or behavior resulting in expulsion or serious suspensions) .... +5
(Same as CATS Item)

3. History of alcohol problems *(Check if present):*
˜ Parental Alcoholism ˜ Teenage Alcohol Problem
˜ Adult Alcohol Problem ˜ Alcohol involved in prior offense
˜ Alcohol involved in index offense
No boxes checked................................... -1
1 or 2 boxes checked 0
3 boxes checked .... +1
4 or 5 boxes checked .... +2
Evidence:

4. Marital status (at the time of or prior to index offense):
Ever married (or lived common law in the same home for at least six months) .... -2
Never married... +1
Evidence:

5. Criminal history score for nonviolent offenses prior to the index offense
Score 0 .... -2
Score 1 or 2.. 0
Score 3 or above .... +3
(from the Cormier-Lang system, see below)

6. Failure on prior conditional release (includes parole or probation violation or revocation, failure to comply, bail violation, and any new arrest while on conditional release):
No...0
Yes .... +3
Evidence:

7. Age at index offense
Enter Date of Index Offense: ___/___/_____
Enter Date of Birth: ___/___/_____
Subtract to get Age:
39 or over .... -5
34 - 38 .... -2
28 - 33 .... -1
27 0
26 or less.. +2

8. Victim Injury (for index offense; the most serious is scored):
Death... -2
Hospitalized..0
Treated and released.............................. +1
None or slight (includes no victim).......... +2
Note: admission for the gathering of forensic evidence only is NOT considered as either treated or hospitalized; ratings should be made based on the degree of injury.
Evidence:

9. Any female victim (for index offense)
Yes .... -1
No (includes no victim)............................ +1
Evidence:

10. Meets DSM criteria for any personality disorder (must be made by appropriately licensed or certified professional)
No... -2
Yes .... +3
Evidence:

11. Meets DSM criteria for schizophrenia (must be made by appropriately licensed or certified professional)
Yes .... -3
No .... +1
Evidence:

12. a. Psychopathy Checklist score (if available, otherwise use item 12.b. CATS score)........
4 or under .... -3
5 – 9... -3
10-14 .... -1
15-24 0
25-34 .... +4
35 or higher .... +12
Note: If there are two or more PCL scores, average the scores.
Evidence:

12. b. CATS score (from the CATS worksheet)
0 or 1 .... -3
2 or 3 0
4 ....+2
5 or higher .... +3

12. WEIGHT (Use the highest circled weight from 12 a. or 12 b.) .... _____

TOTAL VRAG SCORE (SUM CIRCLED SCORES FOR ITEMS 1 – 11 PLUS THE WEIGHT FOR ITEM 12): _________

VRAG Score	Category of Risk
-24	Low
-23	Low
-22	Low
-20	Low
-19	Low
-18	Low
-17	Low
-16	Low
-15	Low
-14	Low
-13	Low
-12	Low
-11	Low
-10	Low
-9	Low
-8	Low
-7	Medium
-6	Medium
-5	Medium
-4	Medium
-3	Medium
-2	Medium
-1	Medium
0	Medium
1	Medium
2	Medium
3	Medium
4	Medium
5	Medium
6	Medium
7	Medium
8	Medium
9	Medium
10	Medium
11	Medium
2	Medium
13	Medium
14	High
15	High
16	High
17	High
18	High
19	High
20	High
21	High
22	High
23	High
24	High
25	High
26	High
28	High
32	High

Drugs & Diseases

Calculators

By Category Alphabetically

Addiction Medicine

Anesthesiology

Cardiac Surgery

Cardiology

COVID-19

Critical Care

Emergency

- Intracerebral Hemorrhage
- Ischemic Stroke
- Movement Disorder
- Multiple Sclerosis & Demyelinating Disease
- Neurophysiology
- Seizure
  - 2HELPS2B Score
  - Phenytoin Adjustment in Renal Failure
  - Seizure vs Syncope
- Subarachnoid Hemorrhage

Obstetrics & Gynecology

Oncology

Orthopedics

Otolaryngology (ENT)

2HELPS2B Score

Estimate duration of EEG monitoring needed to detect 95% of seizures

Si | US

Calculator | References/About

1. Frequency of any periodic or rhythmic pattern of more than 2 Hz except generalized rhythmic delta activity?
2. Independent sporadic epileptiform discharges?
3. Lateralized Periodic Discharges (LPDs), Bilateral Independent Periodic Discharges (BIPDs), or Lateralized Rhythmic Delta Activity (LRDA)?
4. "Plus" features: superimposed rhythmic, fast, or sharp activity only on LRDA, LPDs, or BIPDs?
5. Prior seizure: a history of epilepsy or recent events suspicious for clinical seizures?
6. BIRD: Brief potentially Ictal Rhythmic Discharges?

1. Frequency of any periodic or rhythmic pattern of more than 2 Hz except generalized rhythmic delta activity?

Yes

No

Next Question →

Created by QxMD

0/6 completed

Preventing Brain Damage in Critically Ill Patients

CT-angiography, Anterior Communicating Saccular Aneurysm

Head CT without contrast showing Subarachnoid Hemorrhage

- Seizure are common (20%)
- Seizure→ Brain Damage
- Need EEG to detect seizures

Need to use EEG data to predict seizures, determine EEG duration

EEG is expensive and limited: 24hrs of monitoring is $1600-$4000

- 2HELPS2B was not created by doctors
- It is a ML model
- It is just as accurate as black box models.
- Doctors can decide themselves whether to trust it
- Doctors can calibrate the score with information not in the database
- Score can be explained to non-physicians

2HELPS2B

1.	Any cEEG Pattern with Frequency **2 H**z	1 point		· · ·
2.	**E**pileptiform Discharges	1 point	+	· · ·
3.	Patterns include [**L**PD, LRDA, BIPD]	1 point	+	· · ·
4.	**P**atterns Superimposed with Fast or Sharp Activity	1 point	+	· · ·
5.	Prior **S**eizure	1 point	+	· · ·
6.	**B**rief Rhythmic Discharges	**2** points	+	· · ·
		SCORE	=	· · ·

SCORE	0	1	2	3	4	5	6+
RISK	<5%	11.9%	26.9%	50.0%	73.1%	88.1%	95.3%

There are many variables in the database.

Variable
PDR
BRDs
Unreactive background
Prior Sz
GRDA
LRDA
GPDs
LPDs
BIPDs
Infection
Inflammation
Neoplasm
ICH
Metabolic encephalopathy
Stroke
SAH
SDH
TBI
Hypoxic/ischemic
IVH
Hydrocephalus
Discharges
Frequency (>2Hz)[c]

Designing an optimal scoring system is not easy

Key challenges:

- Accuracy
- Sparsity
- Constraints (e.g., FP<20%, fairness, etc.)
- Integer coefficients

CHADS2 Score

1.	*Congestive Heart Failure*	1 point		⋯
2.	*Hypertension*	1 point	+	⋯
3.	*Age* $\geq$ *75*	1 point	+	⋯
4.	*Diabetes Mellitus*	1 point	+	⋯
5.	*Prior Stroke or Transient Ischemic Attack*	2 points	+	⋯
	ADD POINTS FROM ROWS 1–5	**SCORE**	=	⋯

SCORE	0	1	2	3	4	5	6
STROKE RISK	1.9%	2.8%	4.0%	5.9%	8.5%	12.5%	18.2%

Typical approaches:

panel of experts: (Gage et al., 2001), CHADS2 score for stroke prediction

ad hoc: feature selection, followed by logistic regression with the chosen features, scaling, and rounding (Antman et al., 2000), TIMI risk score for unstable angina/non-ST elevation MI

Elastic Net

SCORE	=	1.42	Rhythmic Patterns Include [BiPD, LRDA, LPD]
		+ 0.31	Prior Seizure
		+ 0.21	Epileptiform Discharges
		+ 0.26	Patterns Superimposed with Fast or Sharp Activity
		+ 0.25	Brief Rhythmic Discharges
		− 2.54	

Elastic Net + Rounding

SCORE	=	1	Rhythmic Patterns Include [BiPD, LRDA, LPD]
		+ 0	~~Prior Seizure~~
		+ 0	~~Epileptiform Discharges~~
		+ 0	~~Patterns Superimposed with Fast or Sharp Activity~~
		+ 0	~~Brief Rhythmic Discharges~~
		− 3	

Elastic Net

SCORE	=	1.42	Rhythmic Patterns Include [BiPD, LRDA, LPD]
		+ 0.31	Prior Seizure
		+ 0.21	Epileptiform Discharges
		+ 0.26	Patterns Superimposed with Fast or Sharp Activity
		+ 0.25	Brief Rhythmic Discharges
		− 2.54	

Elastic Net + Scaling + Rounding

SCORE	=	14	Rhythmic Patterns Include [BiPD, LRDA, LPD]
		+ 3	Prior Seizure
		+ 2	Epileptiform Discharges
		+ 3	Patterns Superimposed with Fast or Sharp Activity
		+ 3	Brief Rhythmic Discharges
		− 25	

Elastic Net + Scaling + Rounding

SCORE	=	14	Rhythmic Patterns Include [BiPD, LRDA, LPD]
		+ 3	Prior Seizure
		+ 2	Epileptiform Discharges
		+ 3	Patterns Superimposed with Fast or Sharp Activity
		+ 3	Brief Rhythmic Discharges
		− 25	

2HELPS2B

1.	Any cEEG Pattern with Frequency **2 H**z	1 point		···
2.	**E**pileptiform Discharges	1 point	+	···
3.	Patterns include [**L**PD, LRDA, BIPD]	1 point	+	···
4.	**P**atterns Superimposed with Fast or Sharp Activity	1 point	+	···
5.	Prior **S**eizure	1 point	+	···
6.	**B**rief Rhythmic Discharges	**2** points	+	···
		SCORE	=	···

SCORE	0	1	2	3	4	5	6+
RISK	<5%	11.9%	26.9%	50.0%	73.1%	88.1%	95.3%

Risk-Calibrated Supersparse Linear Integer Models (Risk-SLIM)

(Ustun and Rudin, Optimal Scoring Systems, Journal of Machine Learning Research, 2019)

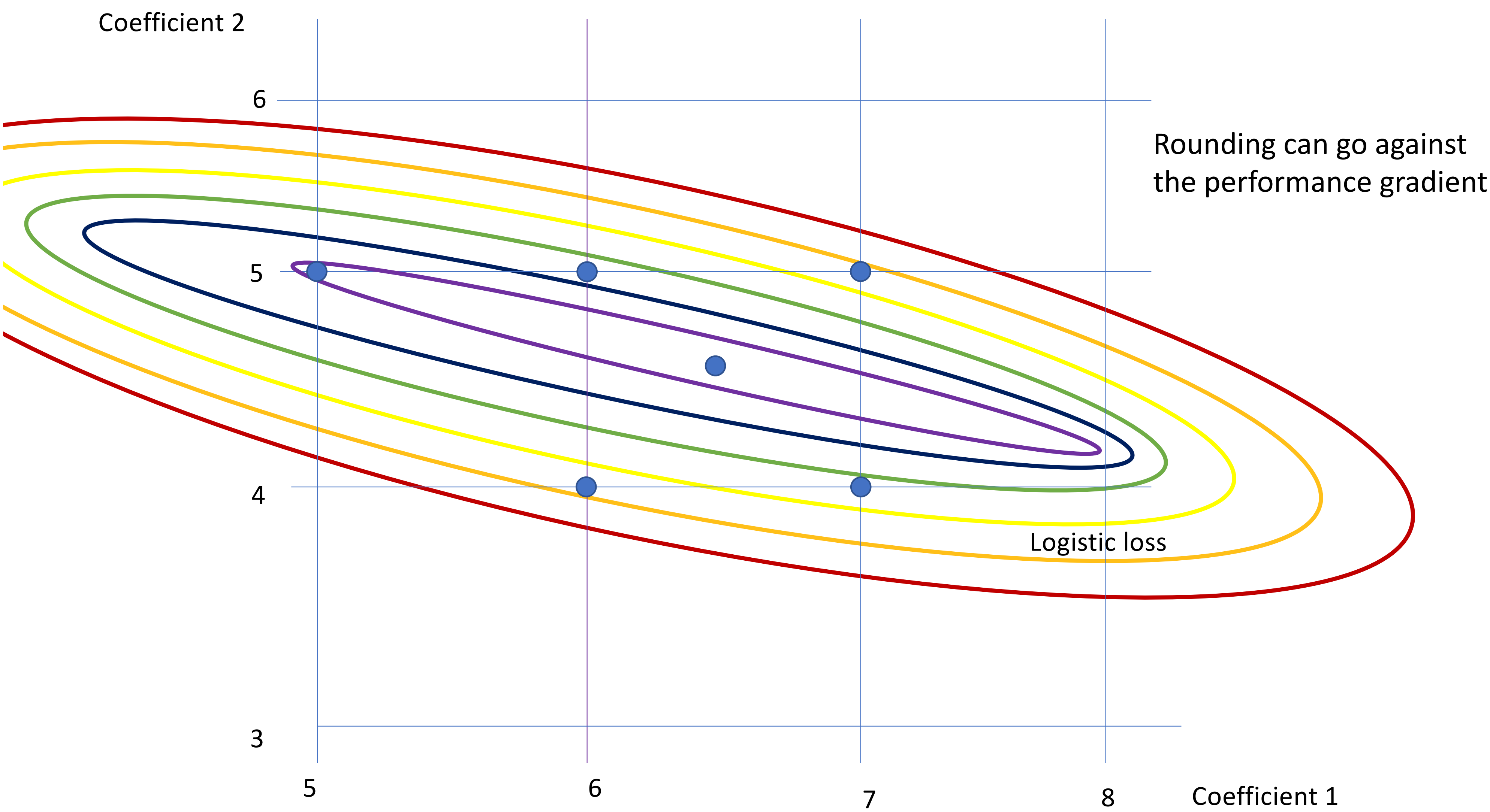
$$\min_{\lambda\in L} \overbrace{\sum_{i=1}^{n} \log\left(1+e^{-y_i x_i^{\mathsf{T}}\lambda}\right)}^{\text{Logistic Loss}} + \overbrace{C\|\lambda\|_0}^{\text{Model Size}}$$

MINLP – really hard…

$$\lambda \in L \text{ means that } \forall j,\ \lambda_j \in \underbrace{\{-10,-9,\ldots,0,\ldots,9,10\}}_{\text{Small Integer Coefficients}}$$

(optional: additional constraints)

Solution uses our *Lattice Cutting Plane* Algorithm, discussed later.



Coefficient 2
6
5
4
3
5
6
7
8
Coefficient 1
Rounding can go against
the performance gradient
Logistic loss

2HELPS2B

Original Investigation

FREE

December 2017

Association of an Electroencephalography-Based Risk Score With Seizure Probability in Hospitalized Patients

Aaron F. Struck, MD[1]; Berk Ustun, PhD[2]; Andres Rodriguez Ruiz, MD[3]; Jong Woo Lee, MD, PhD[4]; Suzette M. LaRoche, MD[3,5]; Lawrence J. Hirsch, MD[6]; Emily J. Gilmore, MD[6]; Jan Vlachy, MS[7]; Hiba Arif Haider, MD[3]; Cynthia Rudin, PhD[8]; M. Brandon Westover, MD, PhD[9]

Author Affiliations | Article Information

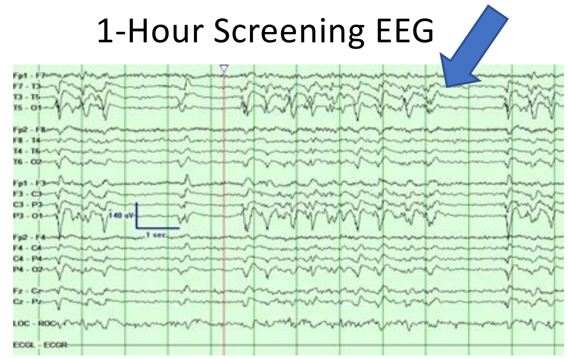
JAMA Neurol. 2017;74(12):1419-1424. doi:10.1001/jamaneurol.2017.2459

Preventing Brain Damage in Critically Ill Patients

CT-angiography, Anterior Communicating Saccular Aneurysm

Head CT without contrast showing Subarachnoid Hemorrhage

1-Hour Screening EEG



2HELPS2B=3 (high-risk)

- Placed on Continuous EEG for >72H
- Start on preventative medications

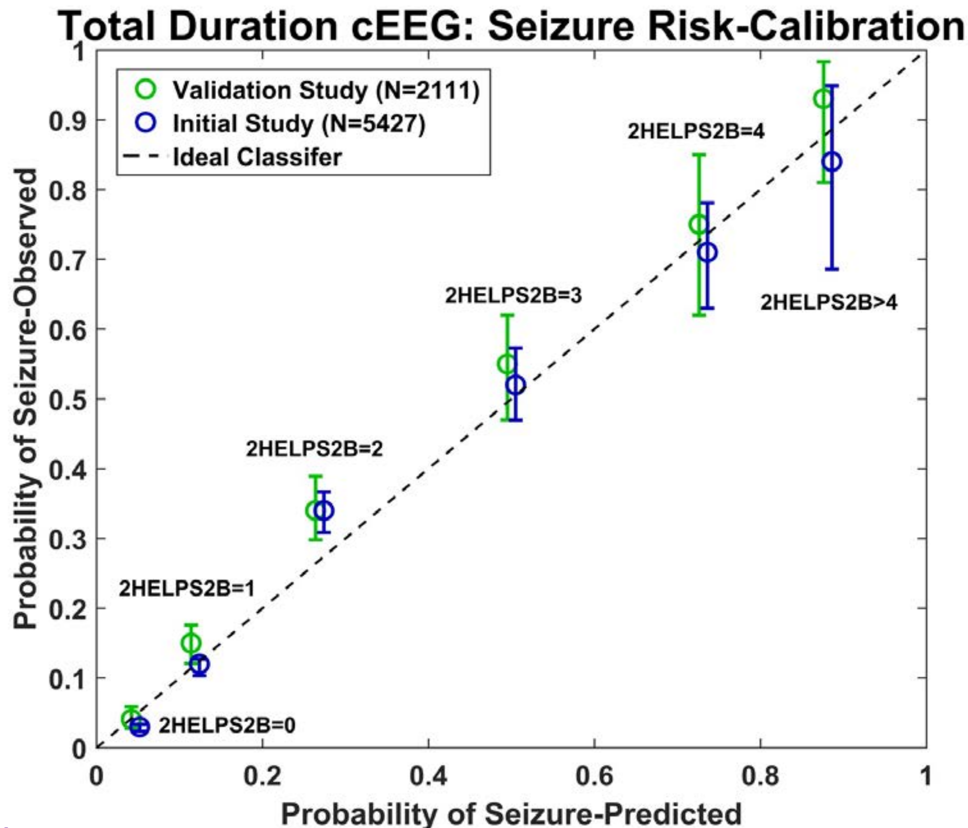
So far…

- 2HELPS2B validated on independent multicenter cohort (Struck et al. 2021, N=2111)


Original Investigation
January 13, 2020
Assessment of the Validity of the 2HELPS2B Score for Inpatient Seizure Risk Prediction

- Implemented: University of Wisconsin, Massachusetts General Hospital/Harvard Medical School
- Ongoing implementation: Emory University, Duke University, Medical University of South Carolina, Free University of Brussels (Belgium)

- Resulted in **63.6%** reduction in duration of EEG monitoring per patient
  - $1,134.831 saving per patient[1]
- **2.82 X** More Patients Monitored
- **>$6.1M** estimated savings in FY 2018 at MGH,UW



[1]2016 Medicare Reimbursement Most Common Professional Code

Risk-Calibrated Supersparse Linear Integer Models (Risk-SLIM)

(Ustun, R, 2019)

MINLP – really hard...

$$\min_{\lambda \in L} \overbrace{\sum_{i=1}^{n} \log\left(1 + e^{-y_i \boldsymbol{x}_i^{\intercal} \lambda}\right)}^{\text{Logistic Loss}} + \overbrace{C\|\lambda\|_0}^{\text{Model Size}}$$

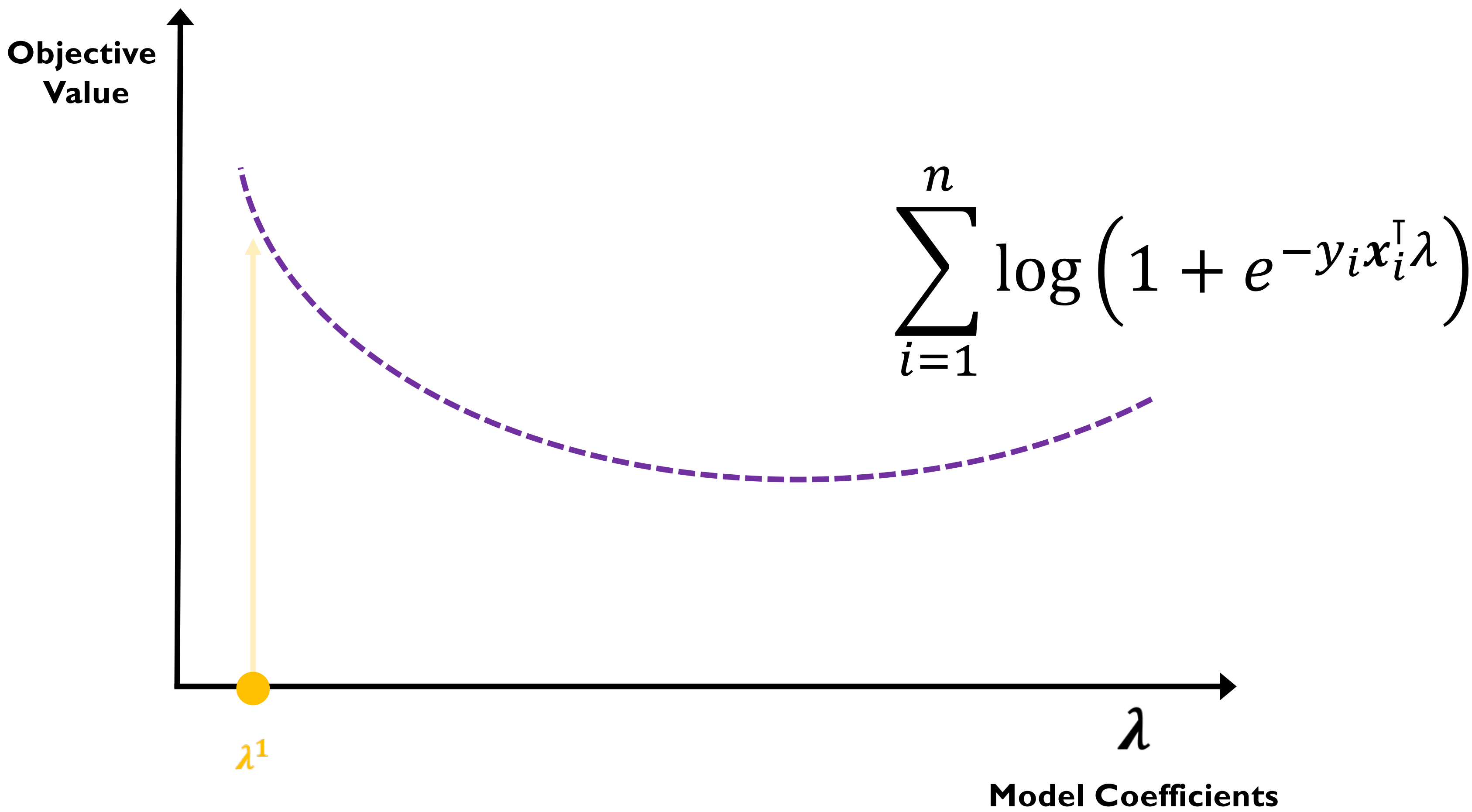
$$\lambda \in L \text{ means that } \forall j,\ \lambda_j \in \underbrace{\{-10,-9,\ldots,0,\ldots,9,10\}}_{\text{Small Integer Coefficients}}$$

(optional: additional constraints)

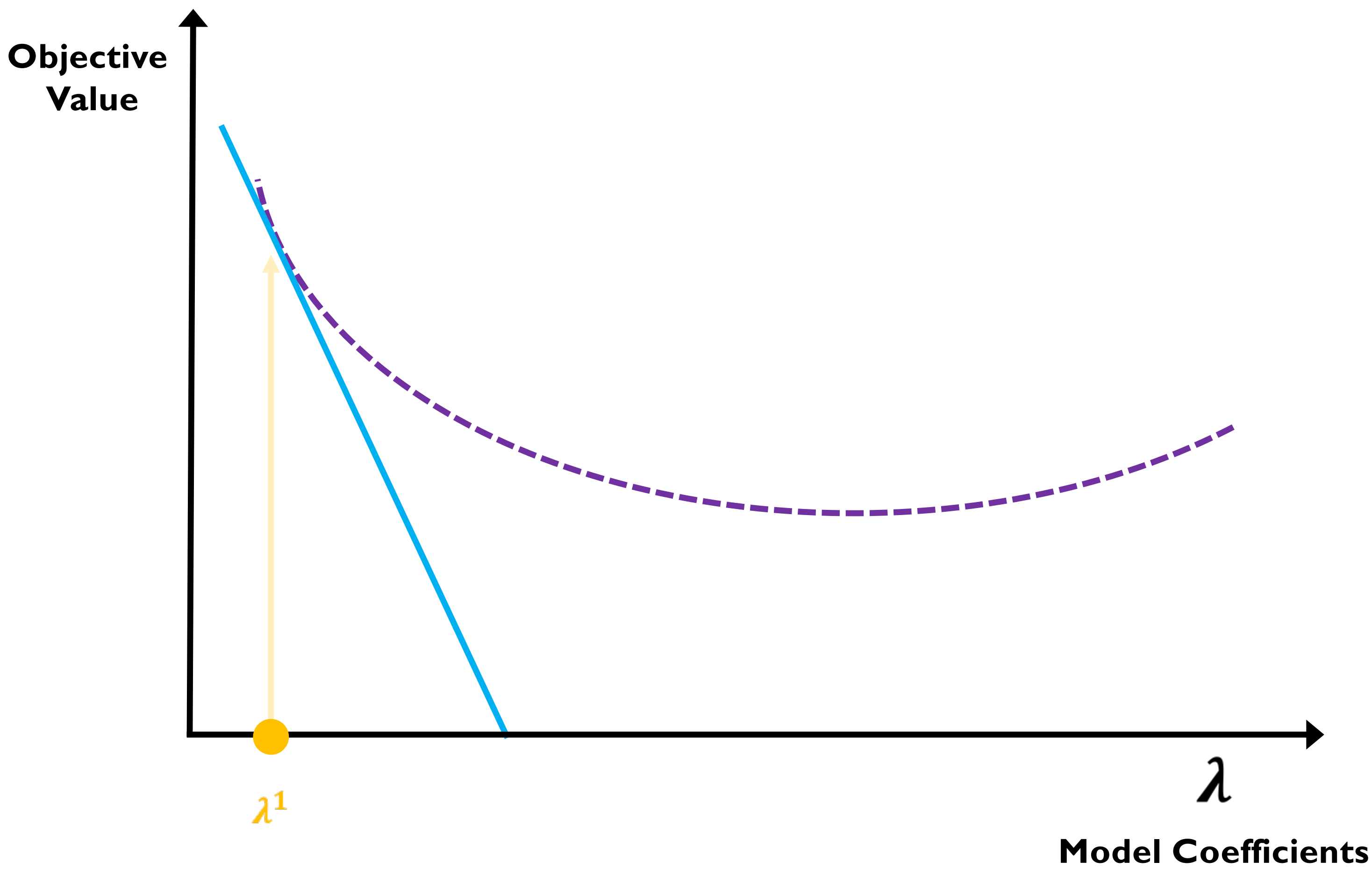
Cutting Planes (Traditional)

$$\min_{\lambda} \sum_{i=1}^{n} \log\left(1 + e^{-y_i \boldsymbol{x}_i^\intercal \lambda}\right)$$

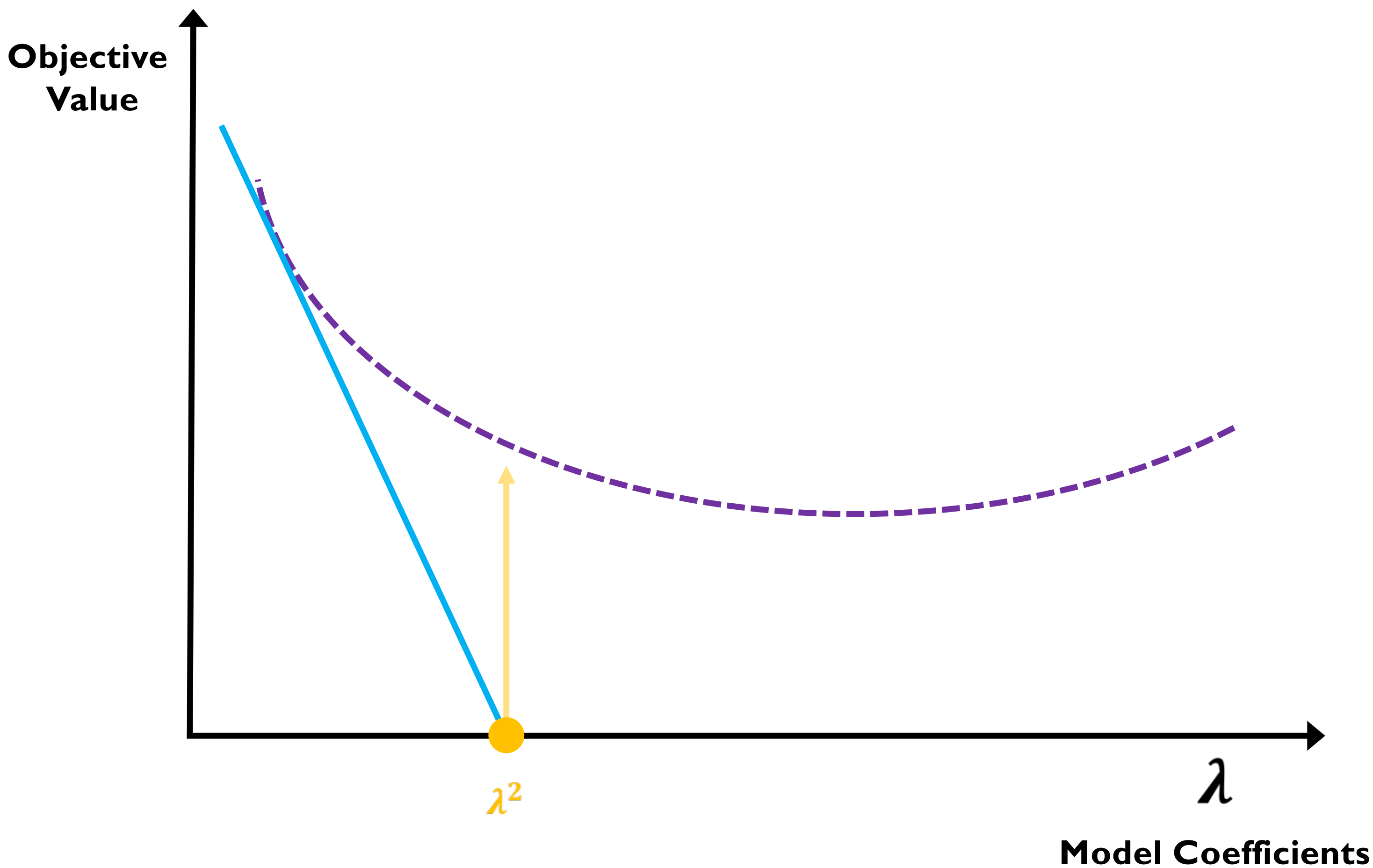
Traditional cutting planes



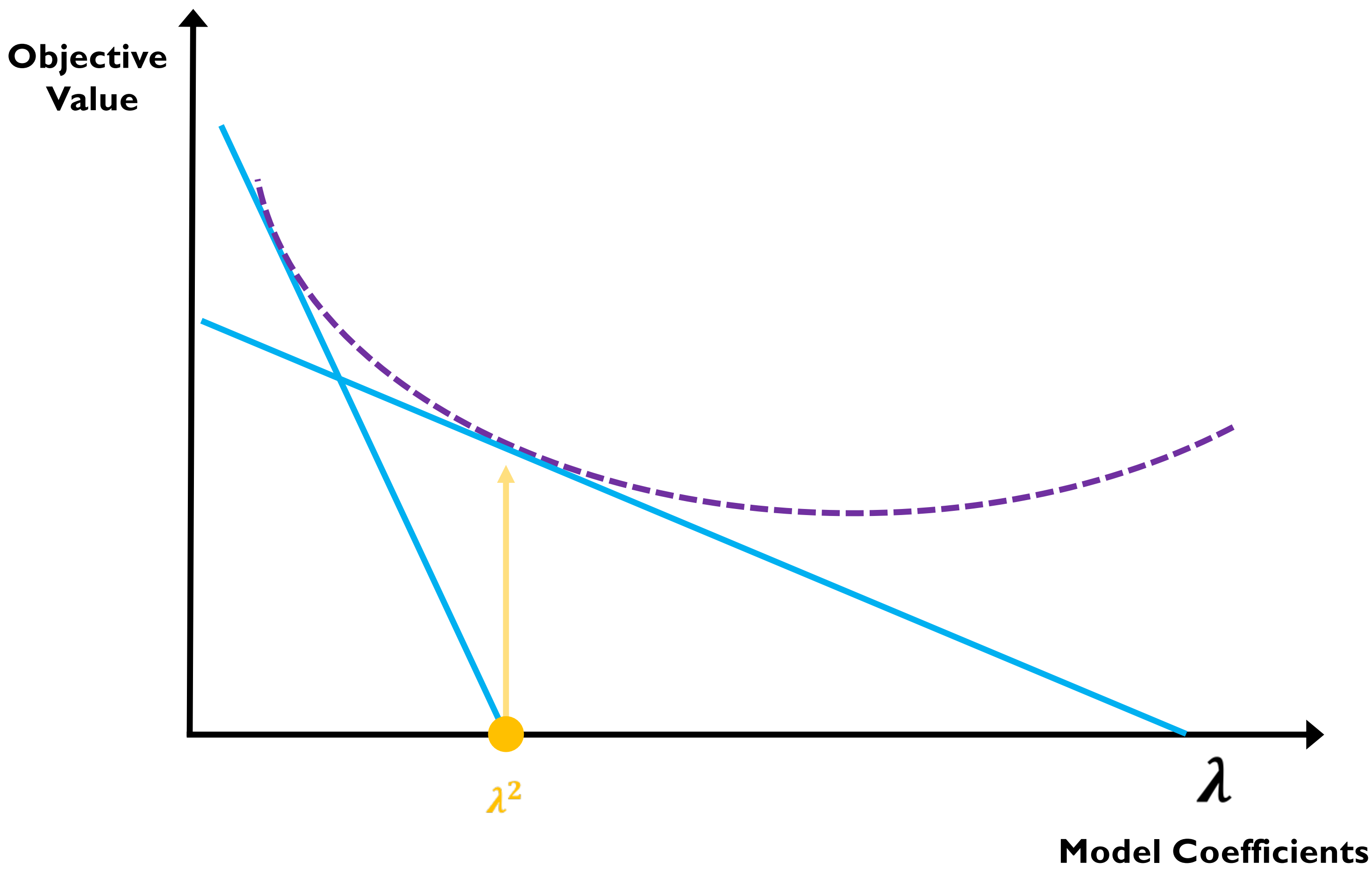
Traditional cutting planes



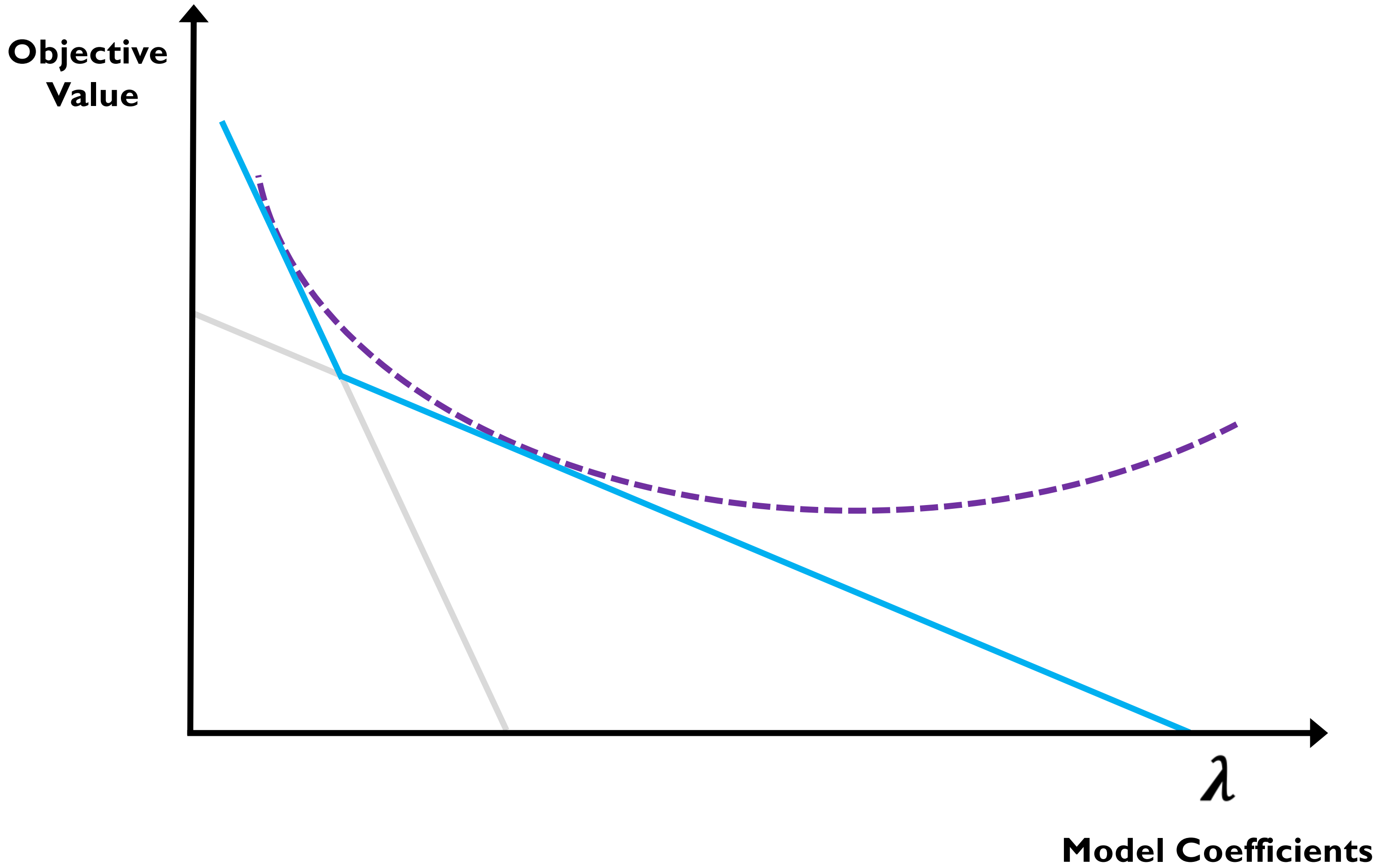
Traditional cutting planes



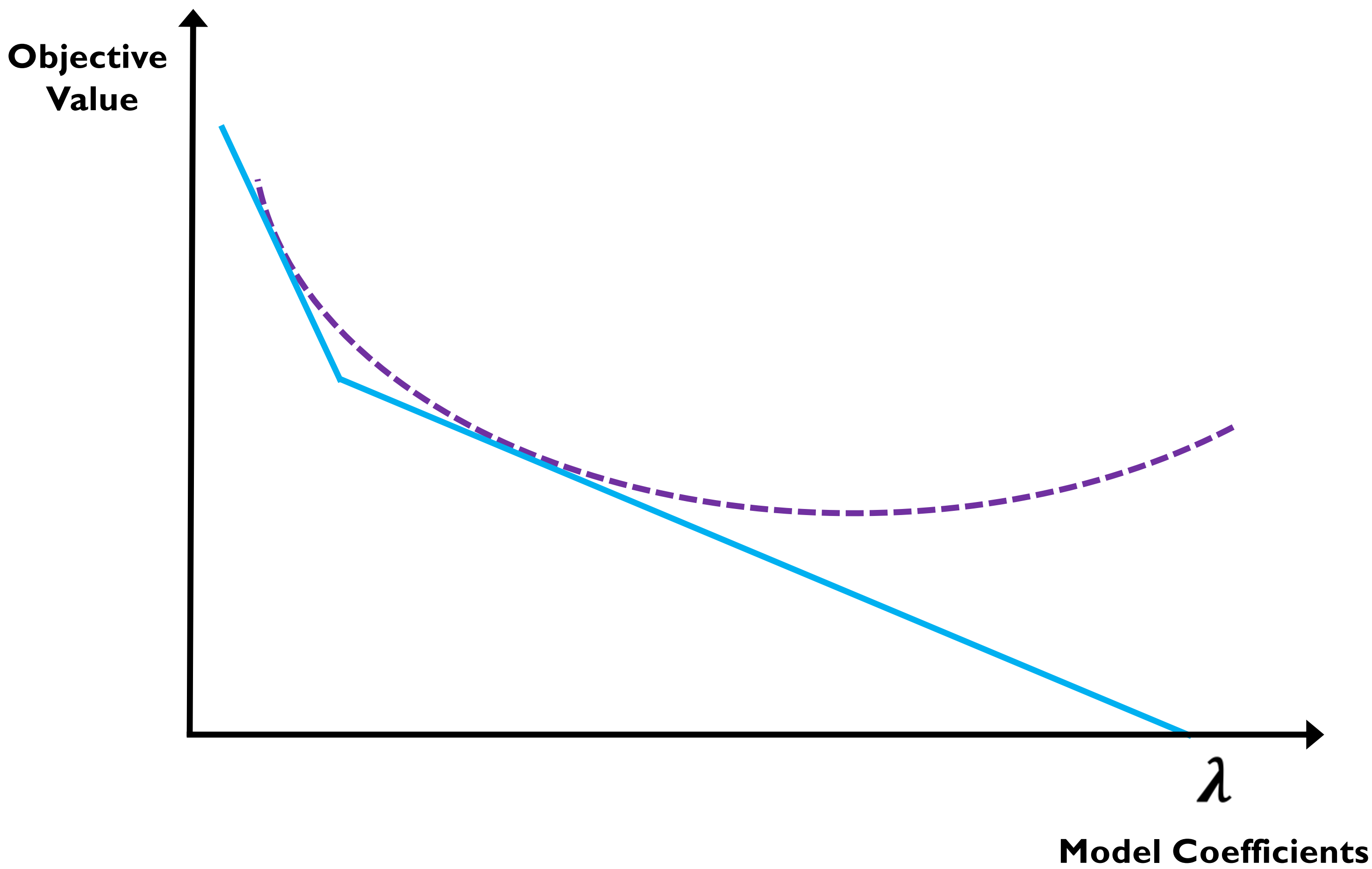
Traditional cutting planes



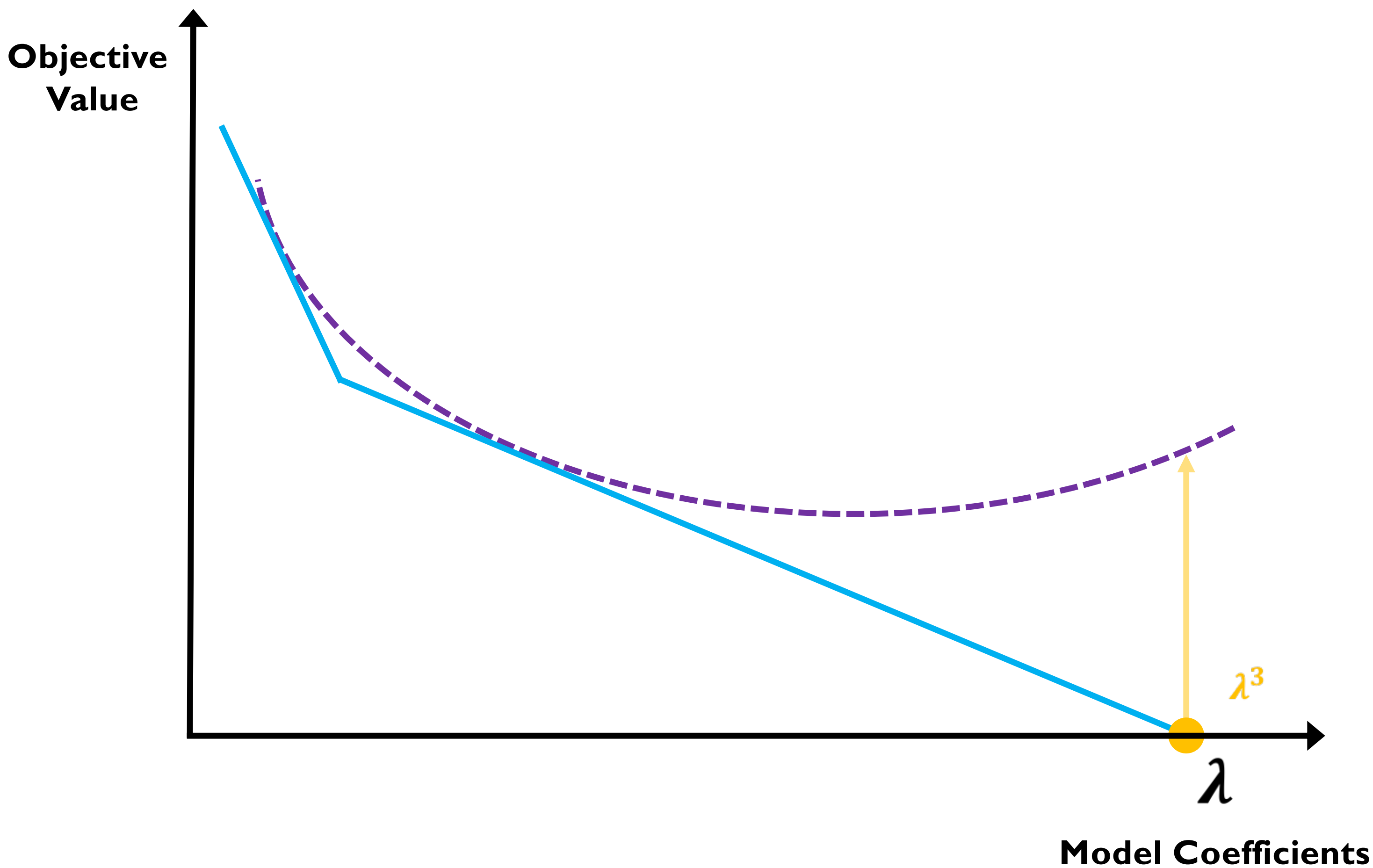
Traditional cutting planes



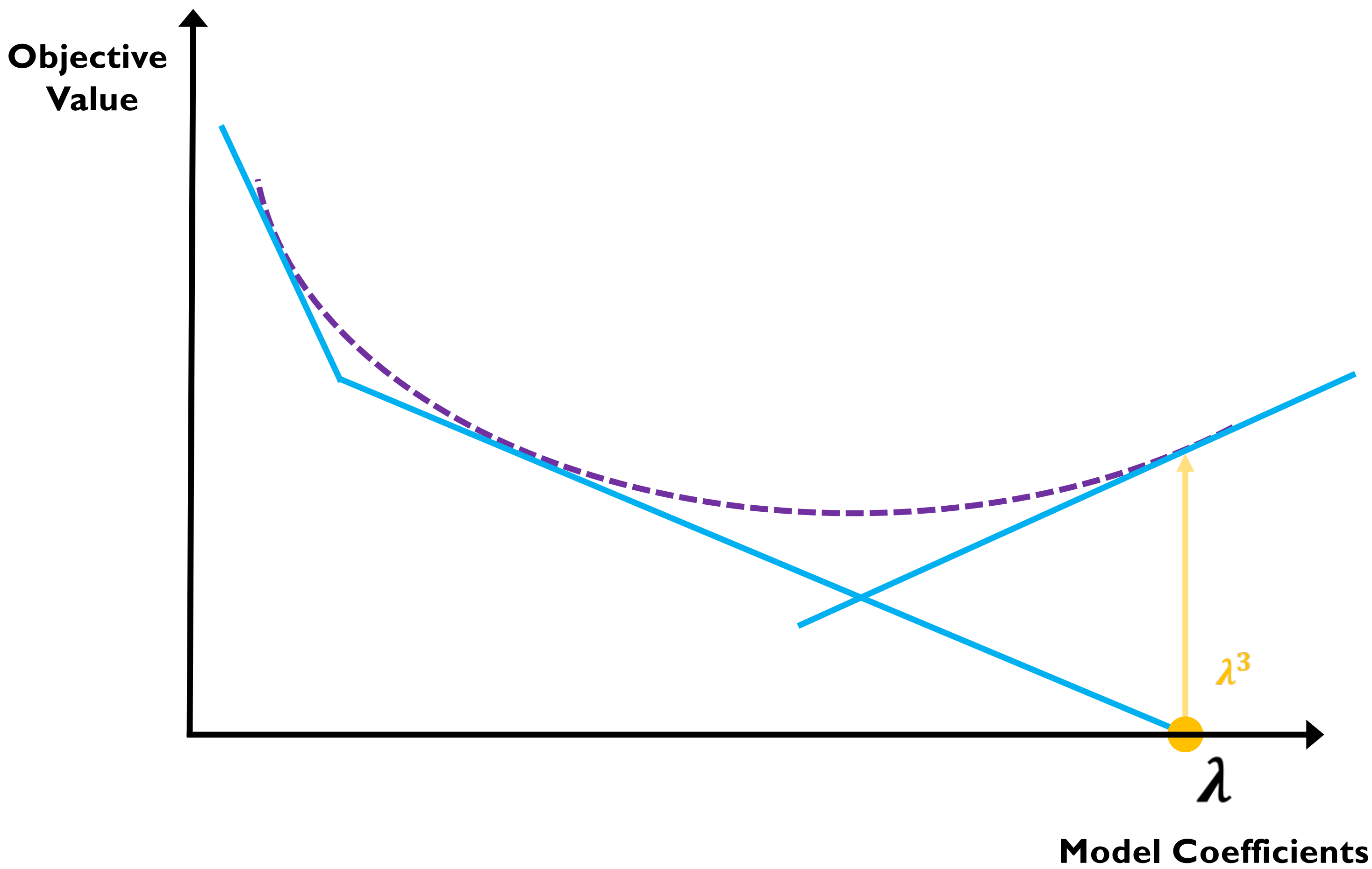
Traditional cutting planes



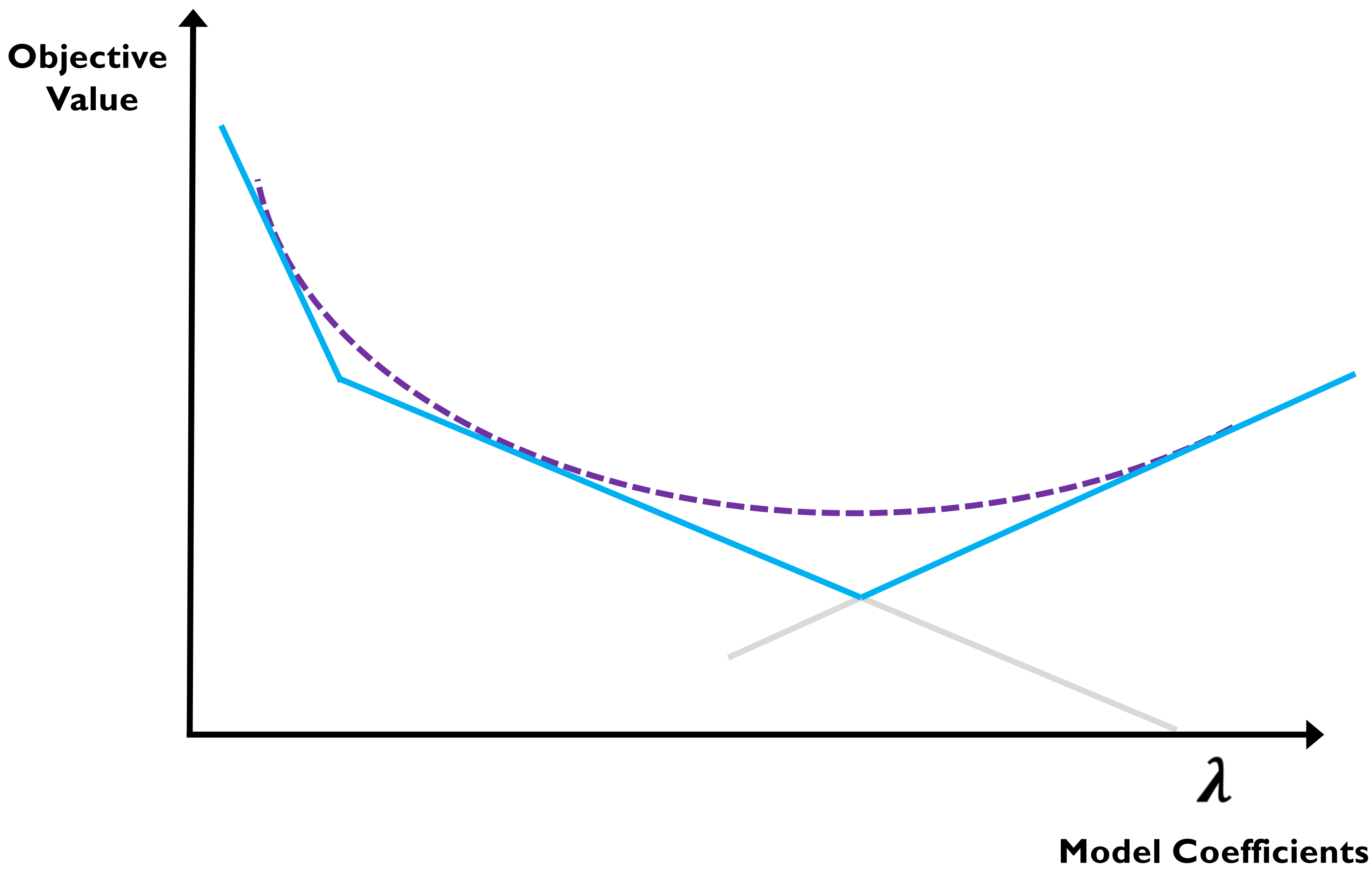
Traditional cutting planes



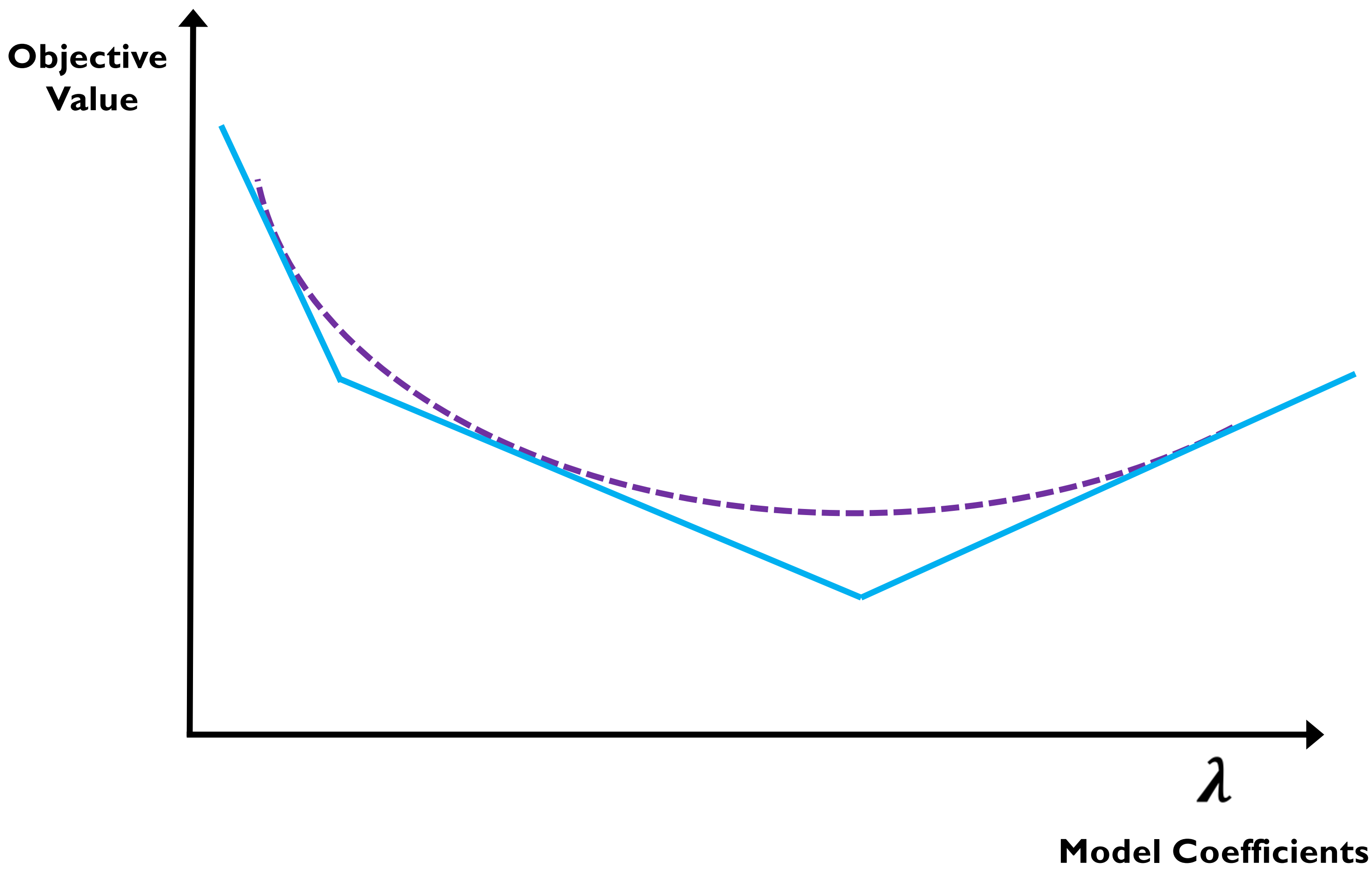
Traditional cutting planes



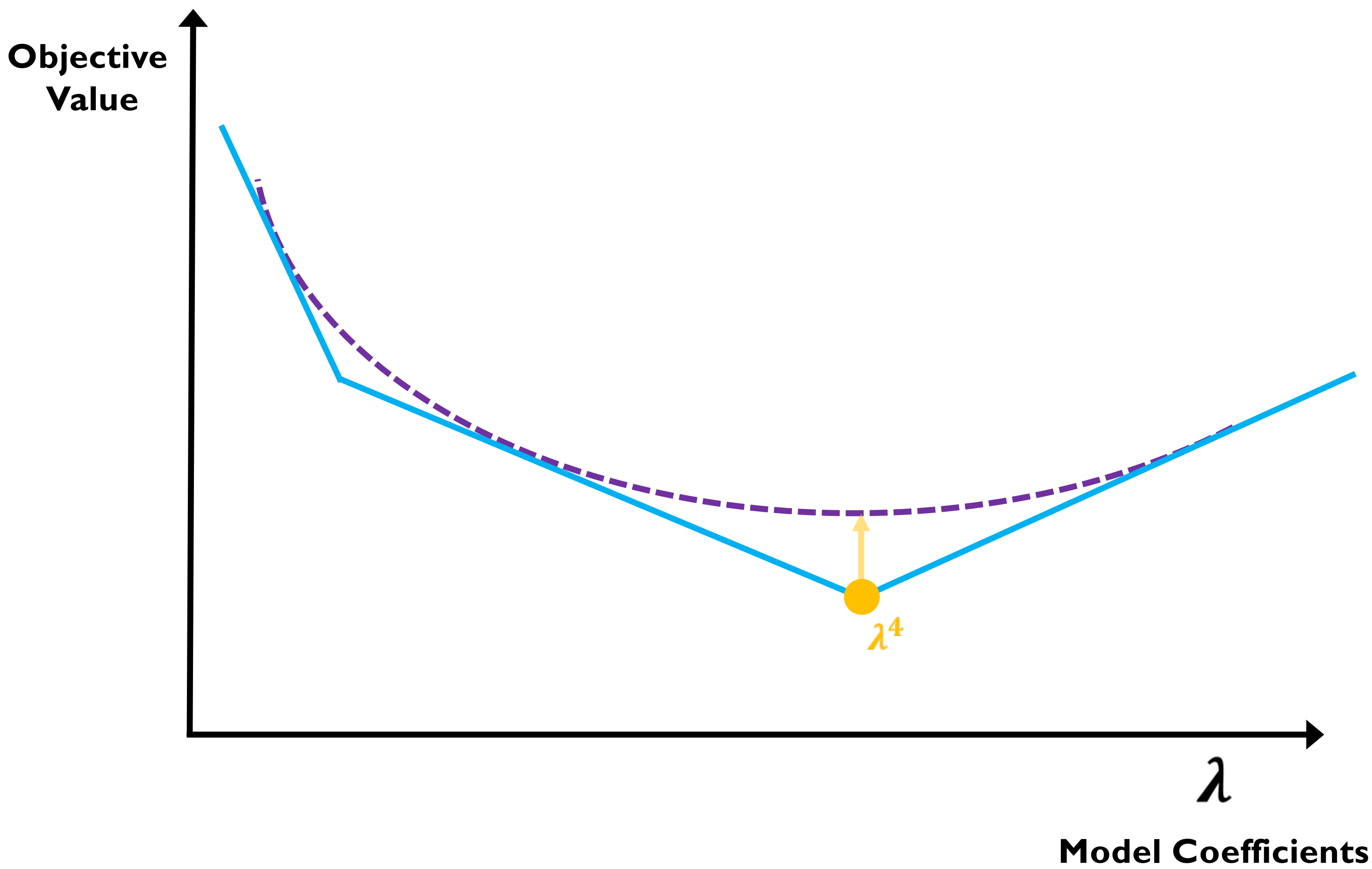
Traditional cutting planes



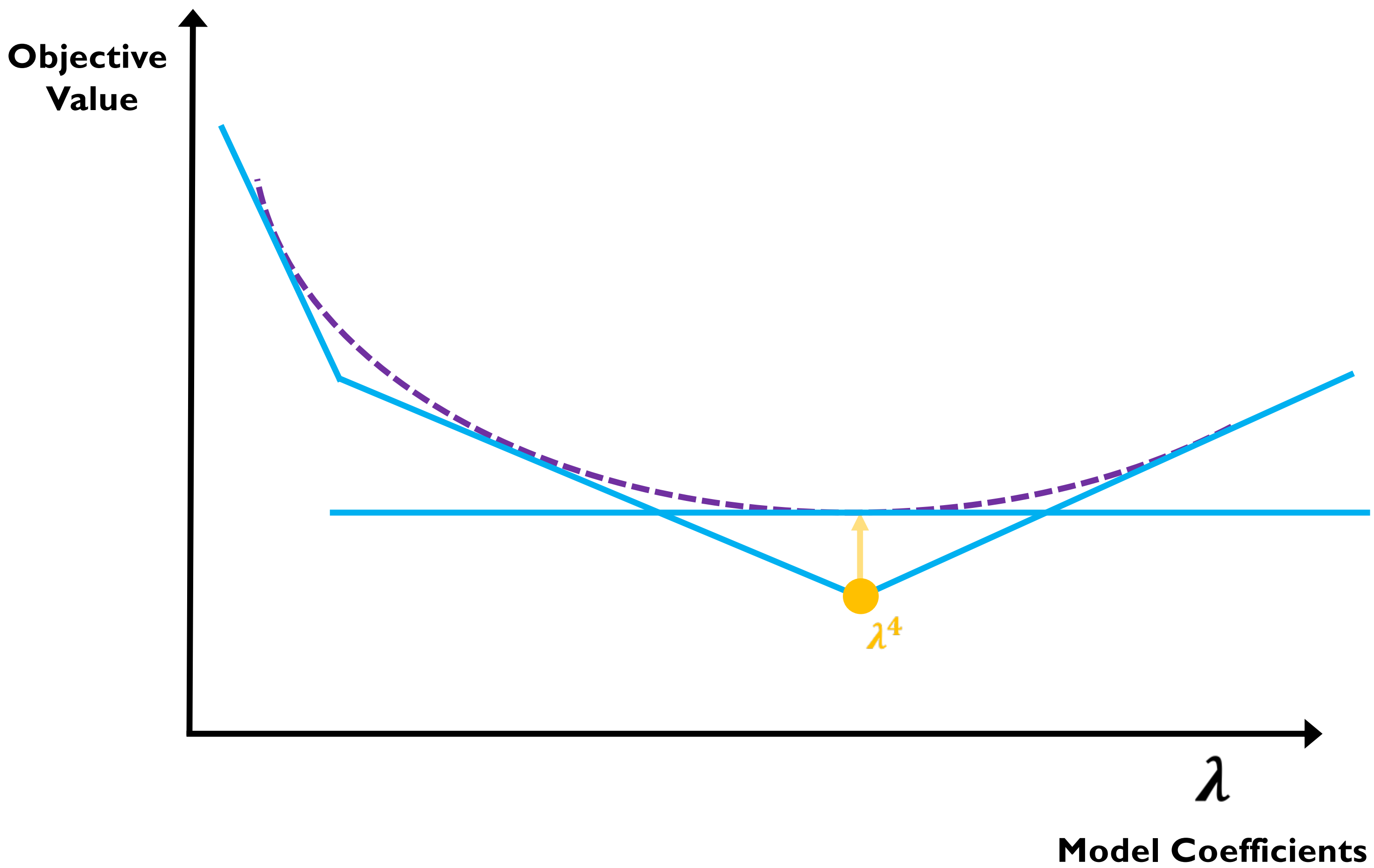
Traditional cutting planes



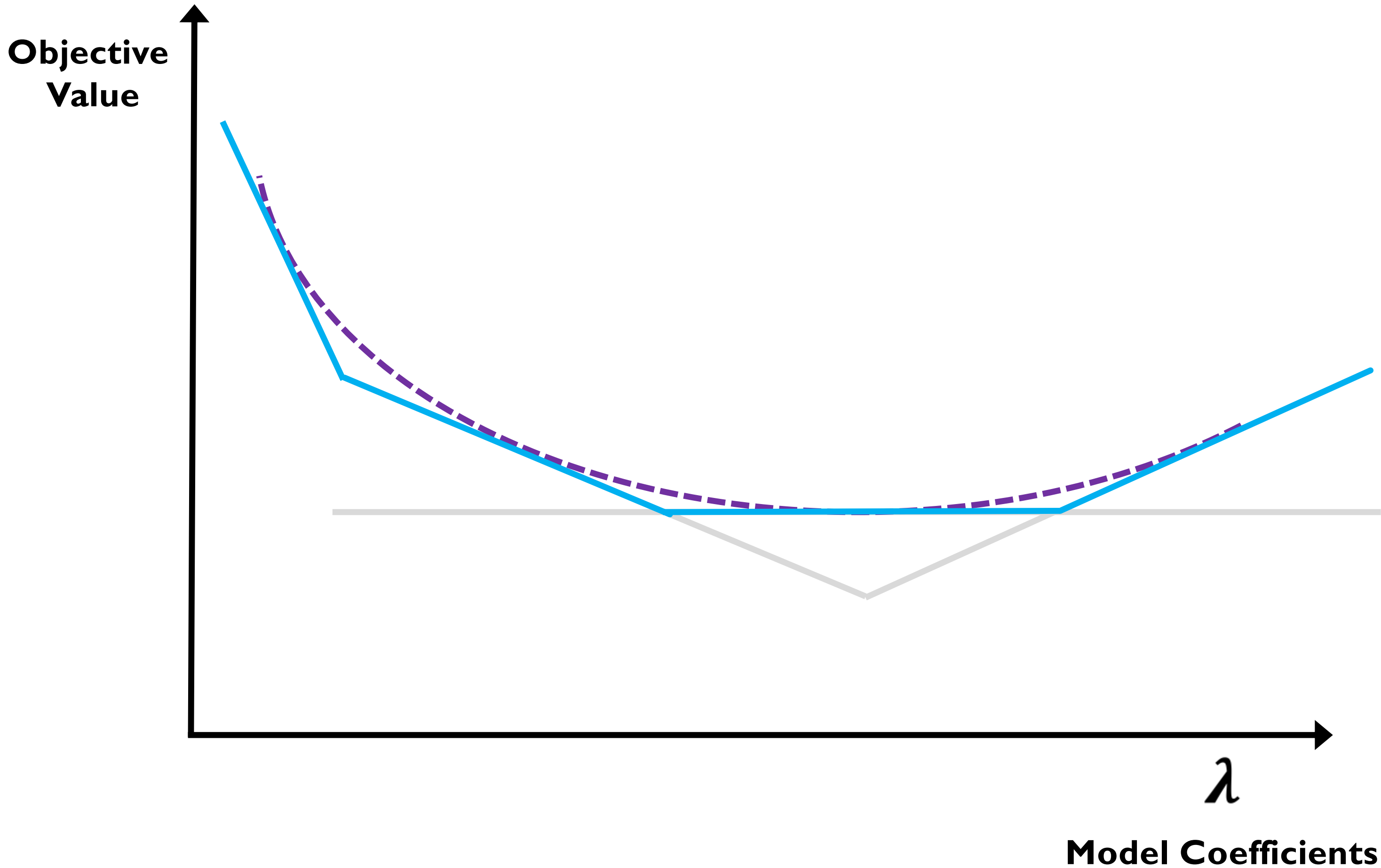
Traditional cutting planes



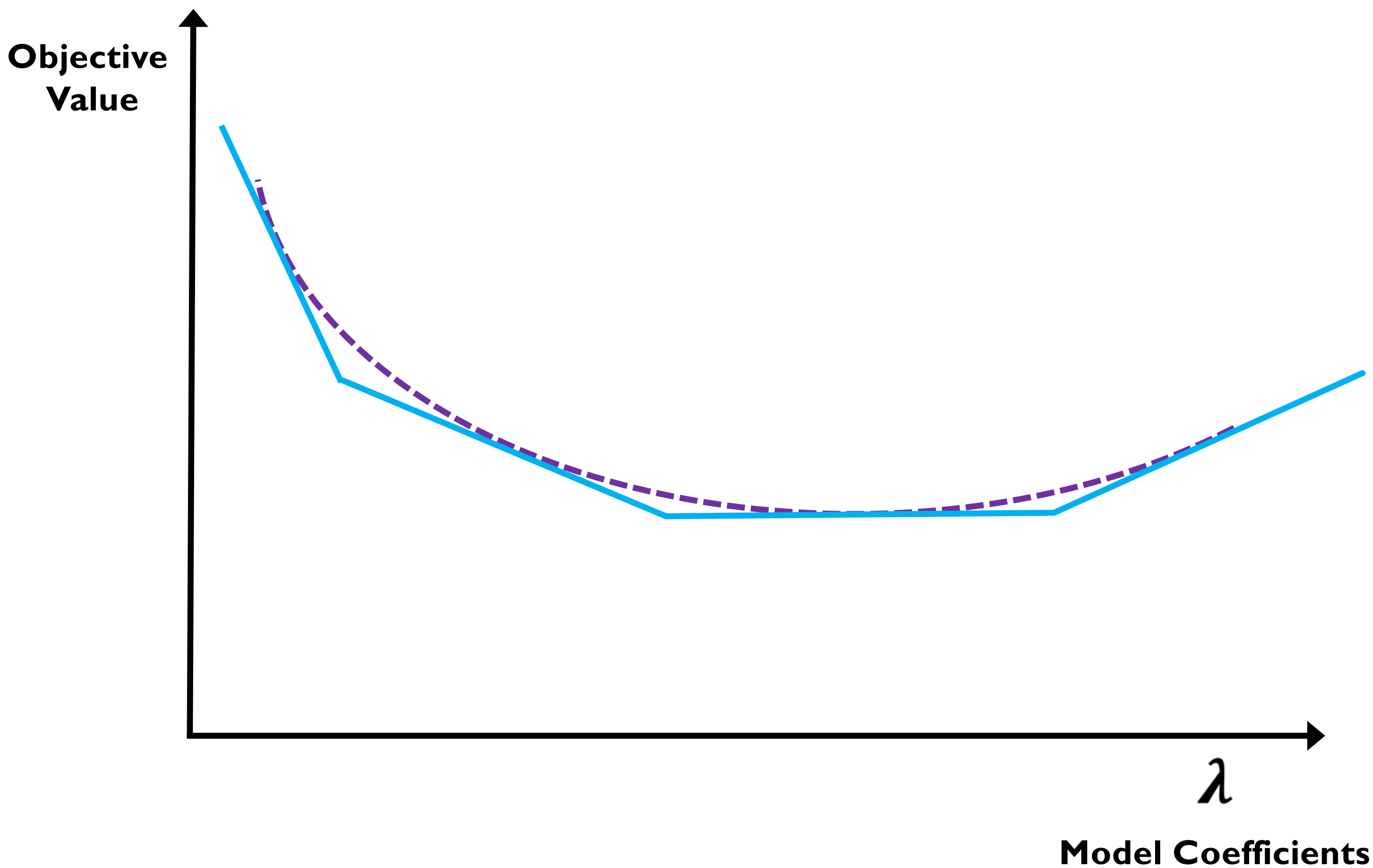
Traditional cutting planes



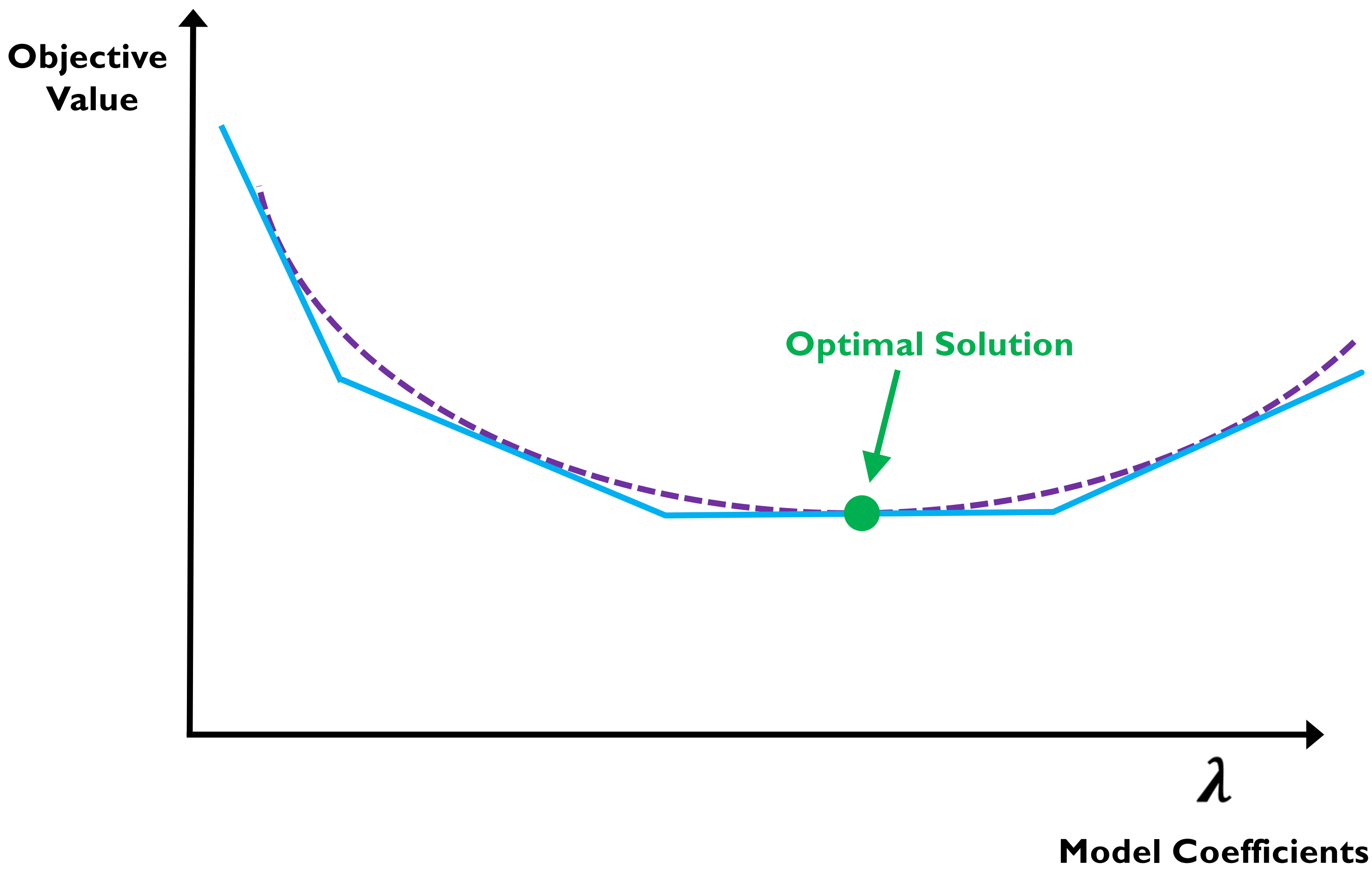
Traditional cutting planes



Traditional cutting planes

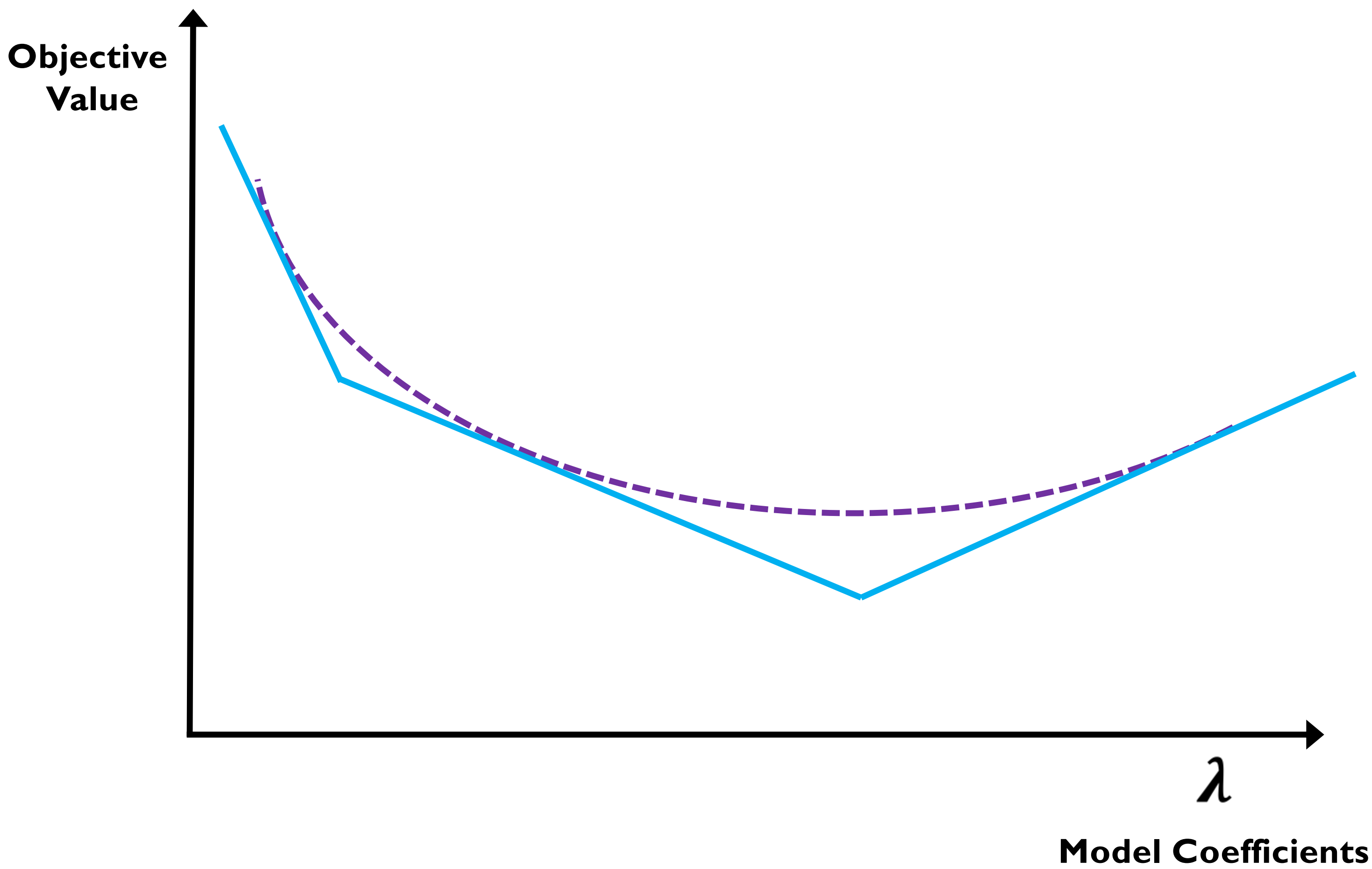


Traditional cutting planes

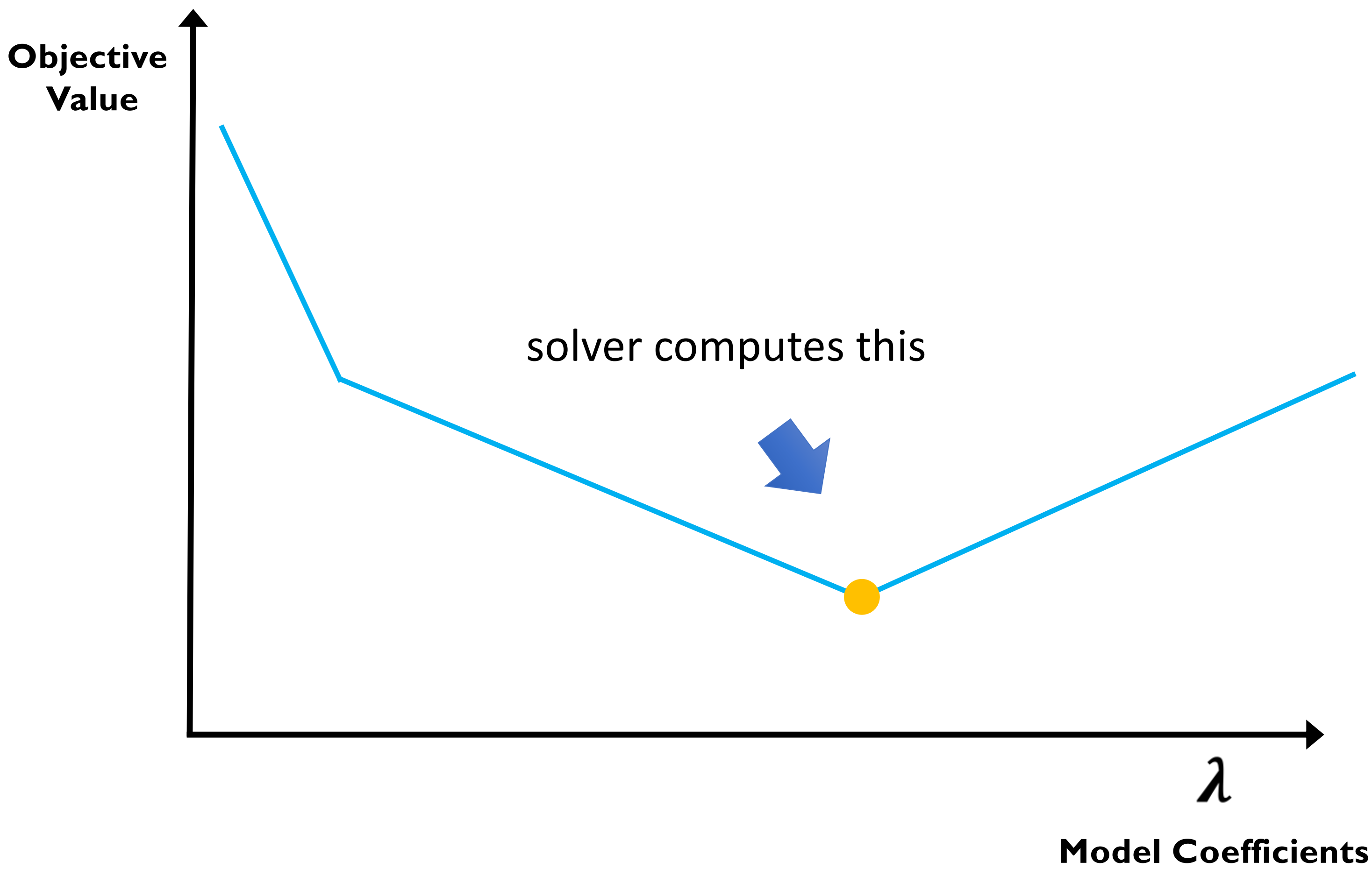


- Something goes wrong when creating models with integer coefficients.

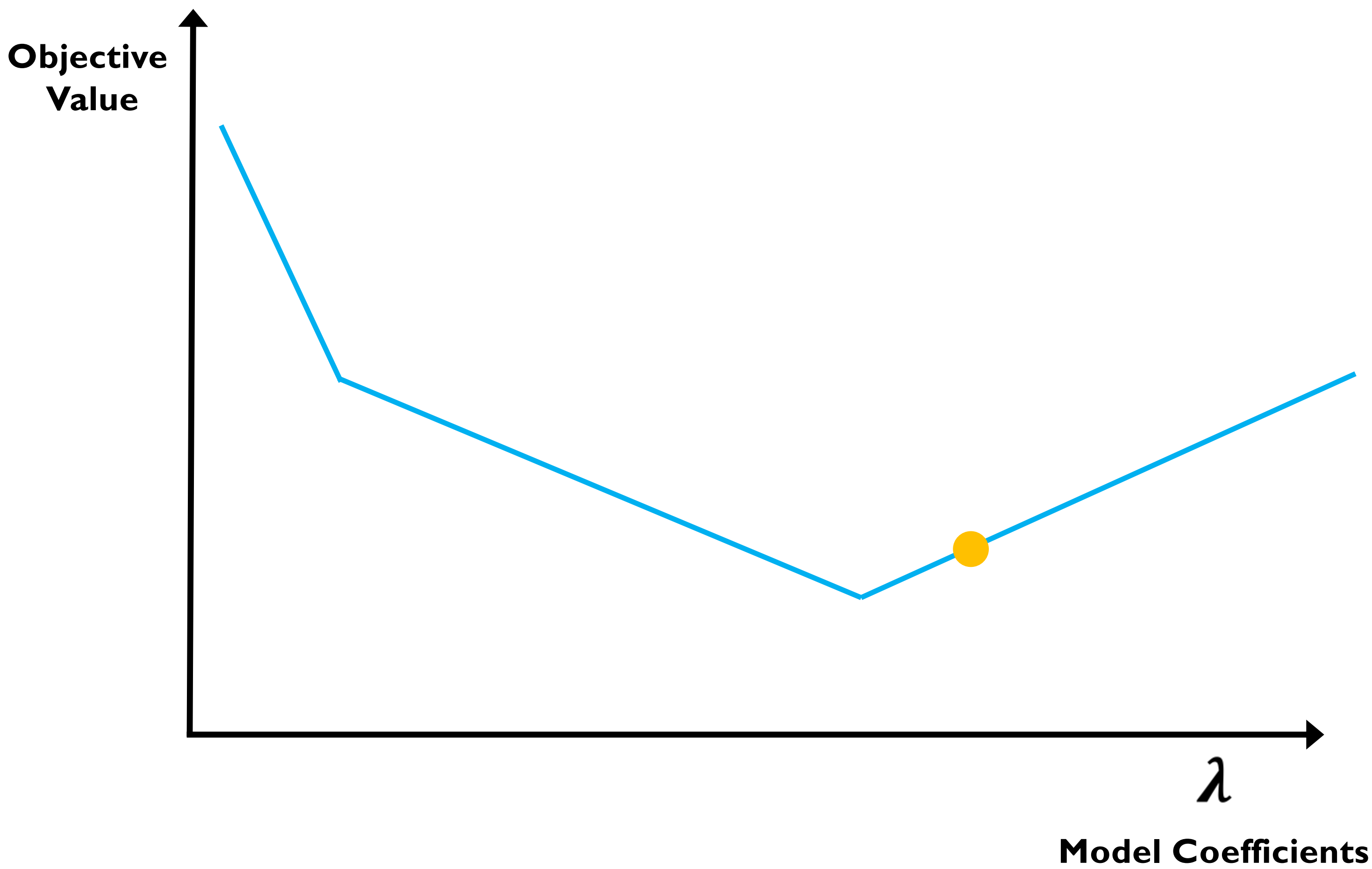
Traditional cutting planes



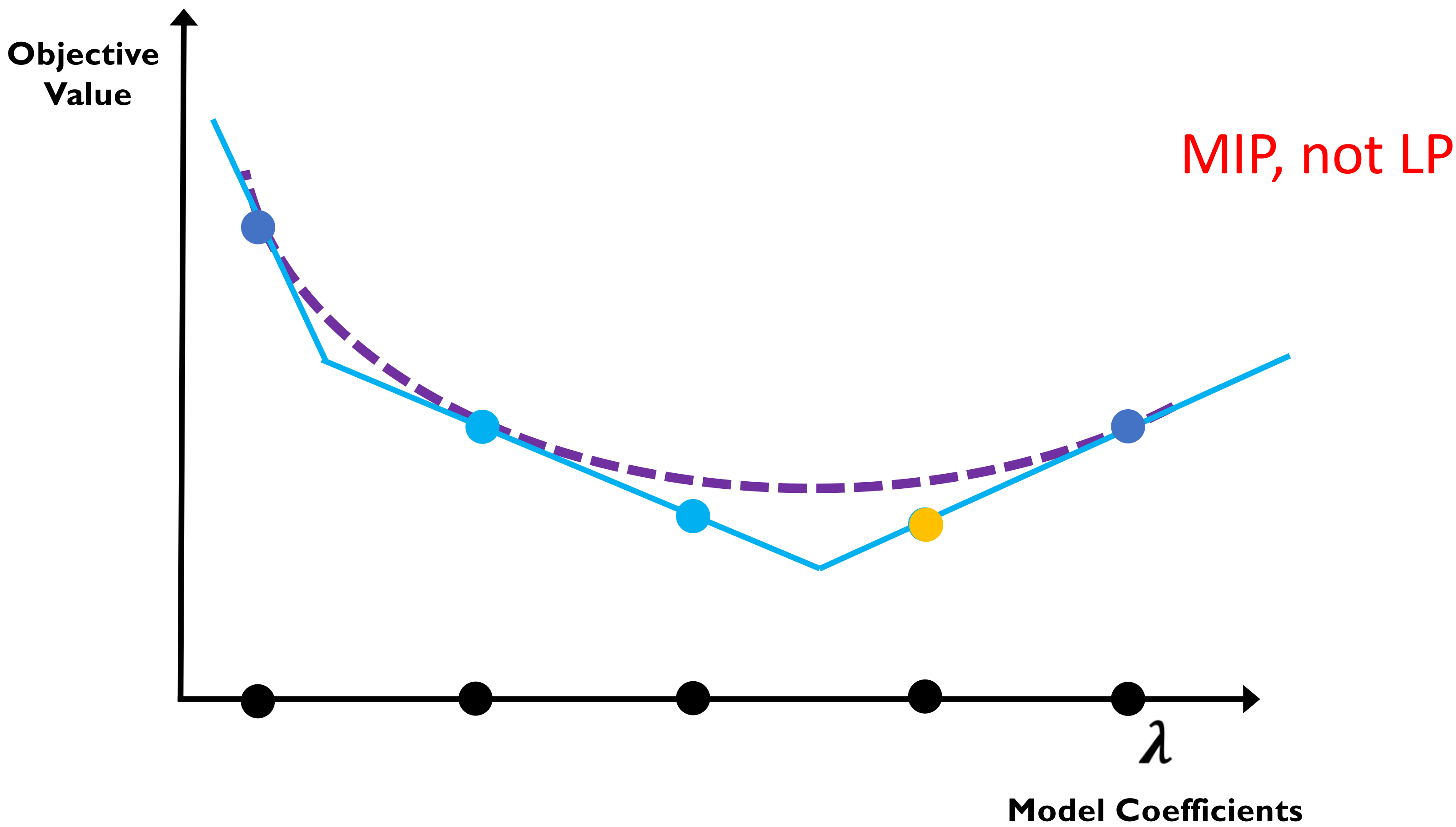
Traditional cutting planes



Traditional cutting planes



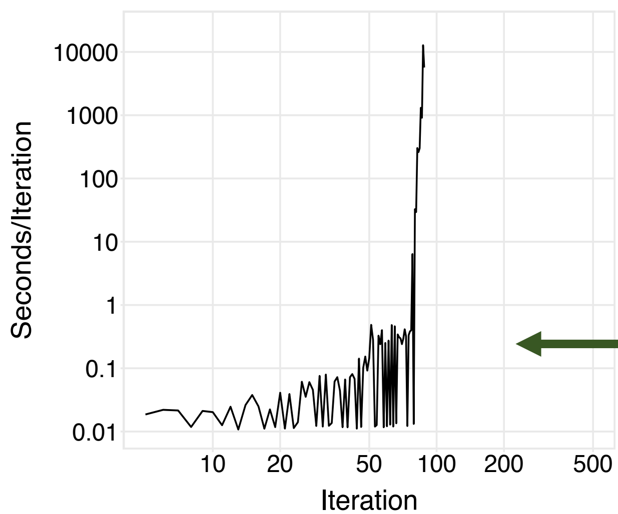
Traditional cutting planes



Stalling

$d = 20$

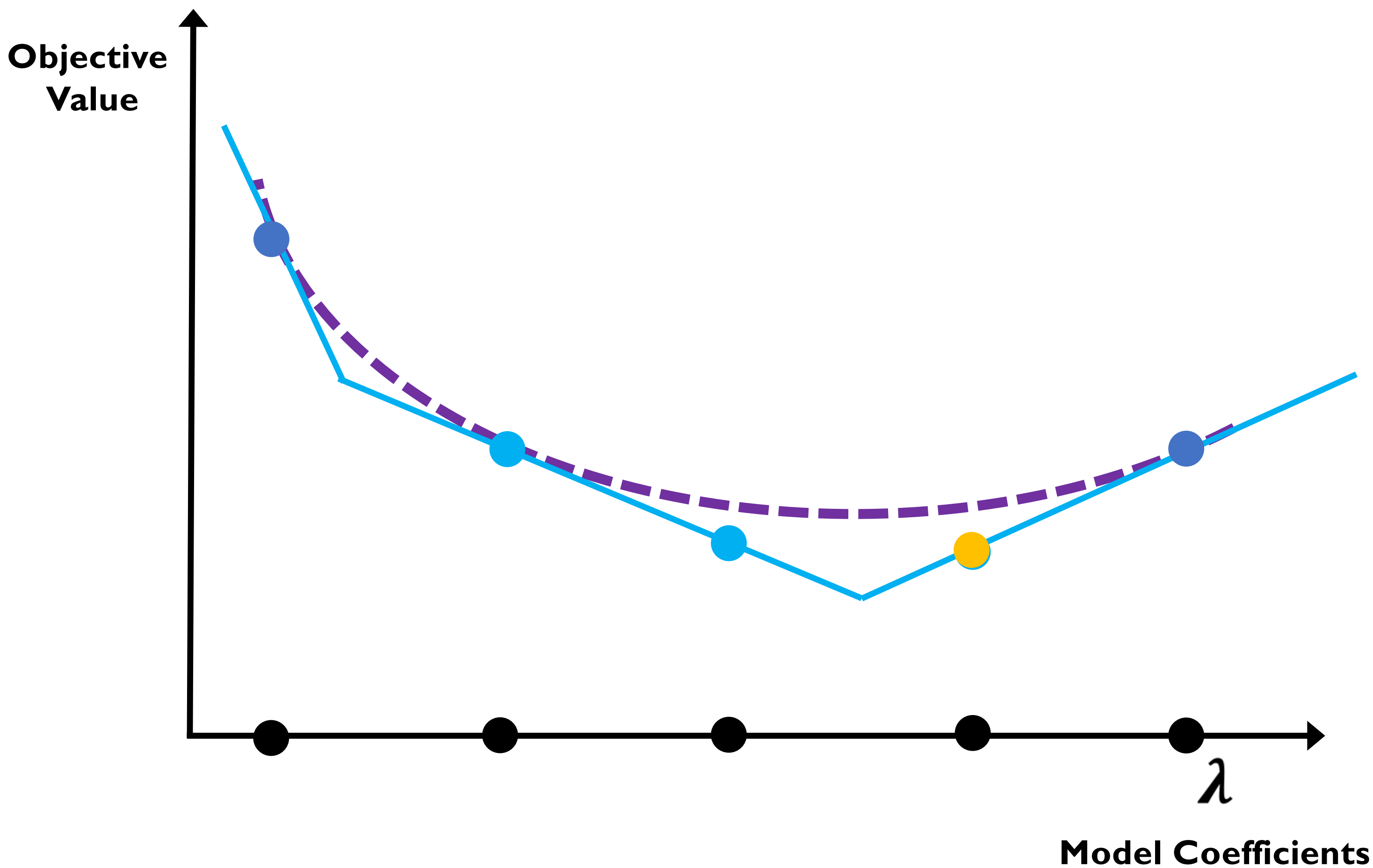
Seconds per iteration



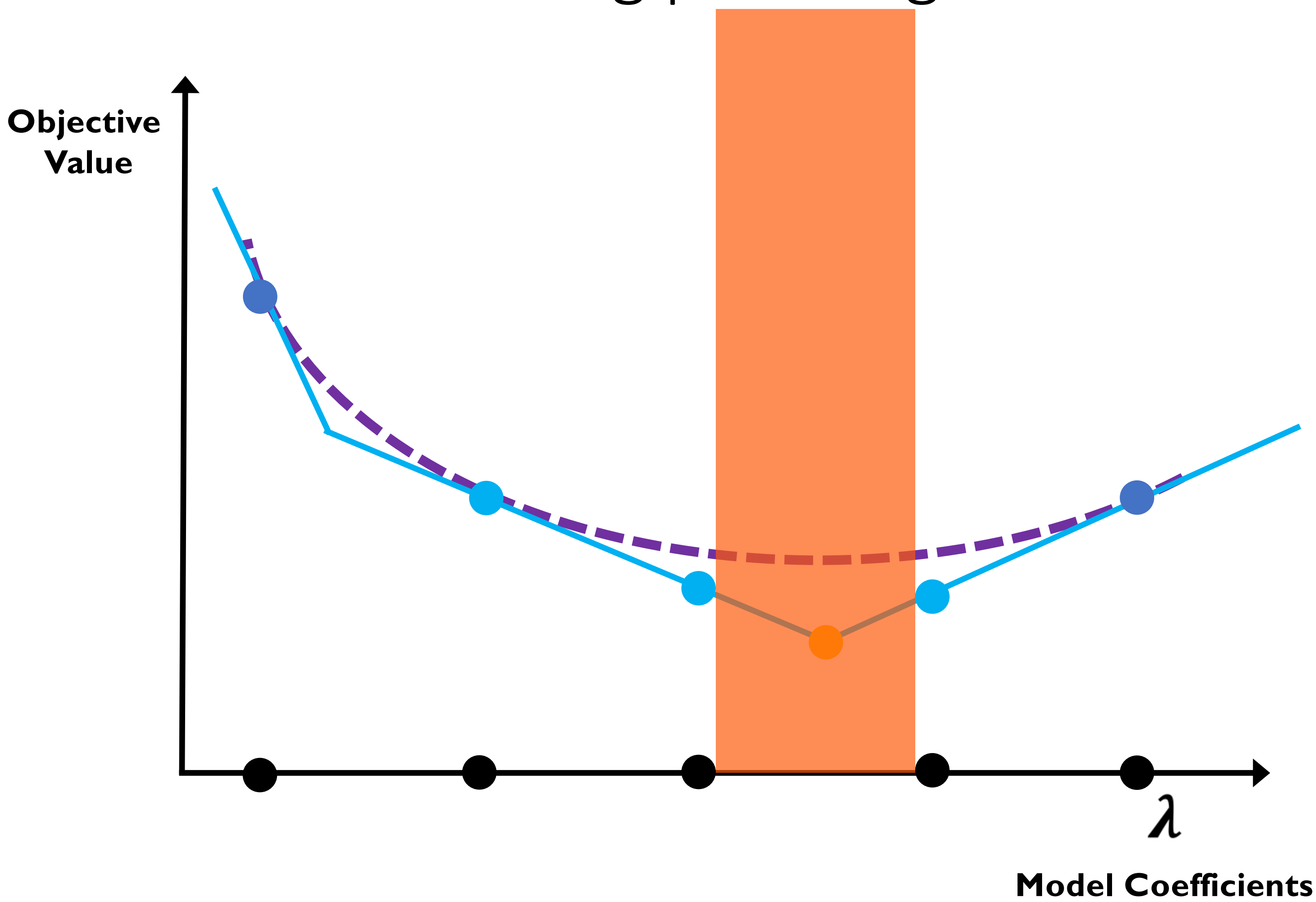
Stalling in traditional cutting planes

RiskSLIM's *Lattice Cutting Plane Algorithm*
(Ustun & Rudin, KDD 17)

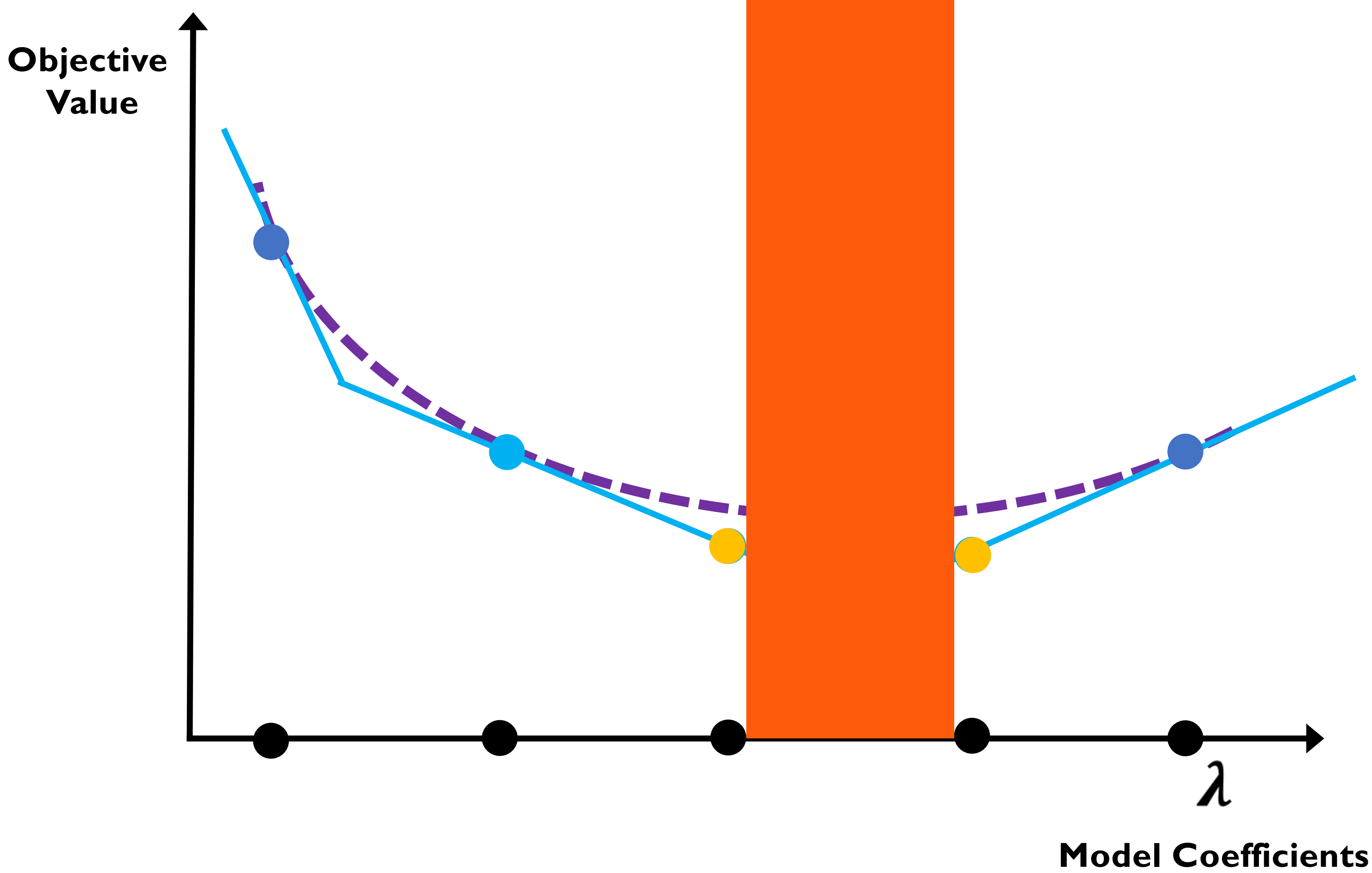
Lattice cutting plane algorithm



Lattice cutting plane algorithm

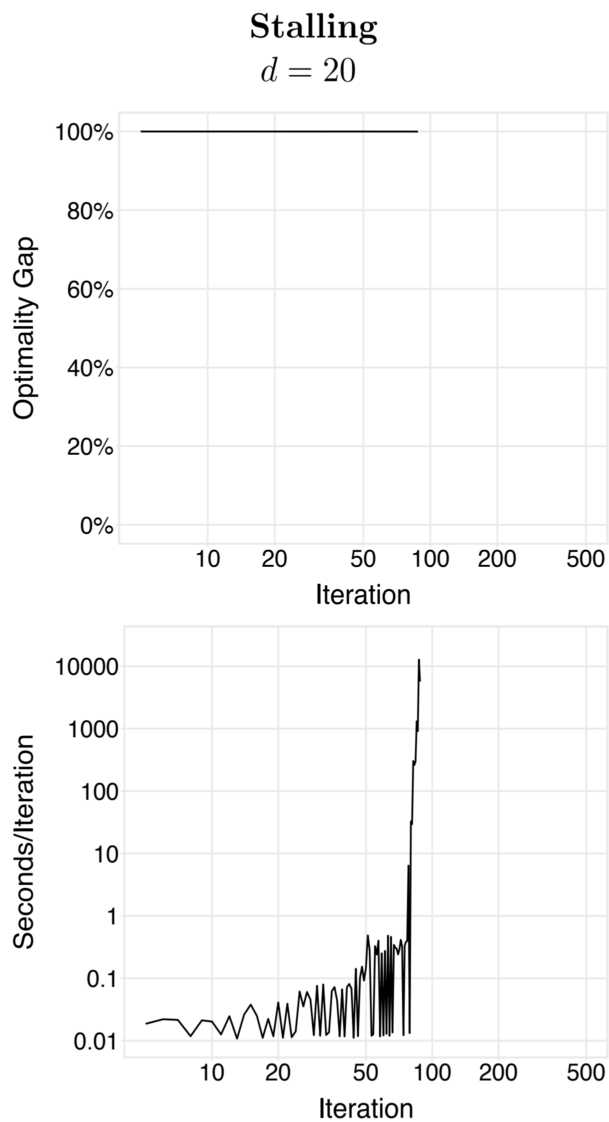


Lattice cutting plane algorithm



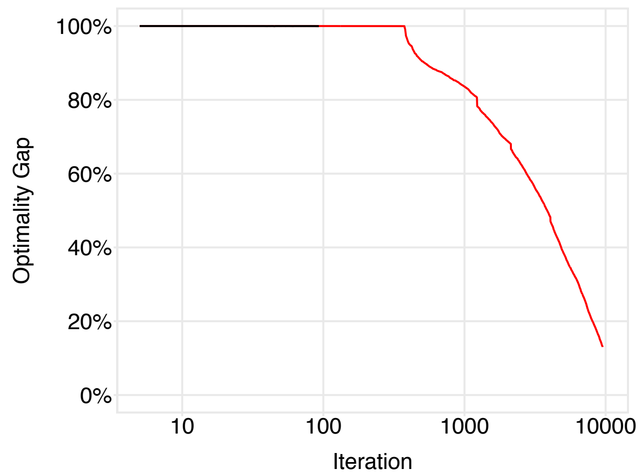
Optimality Gap

Seconds per iteration

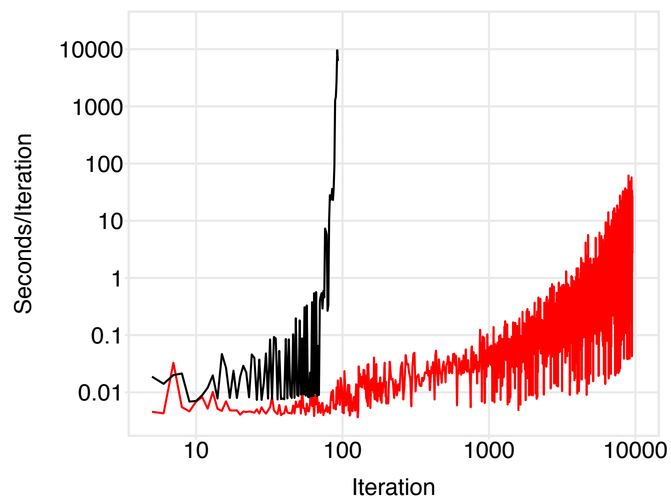


No Stalling

Optimality Gap



Seconds per iteration

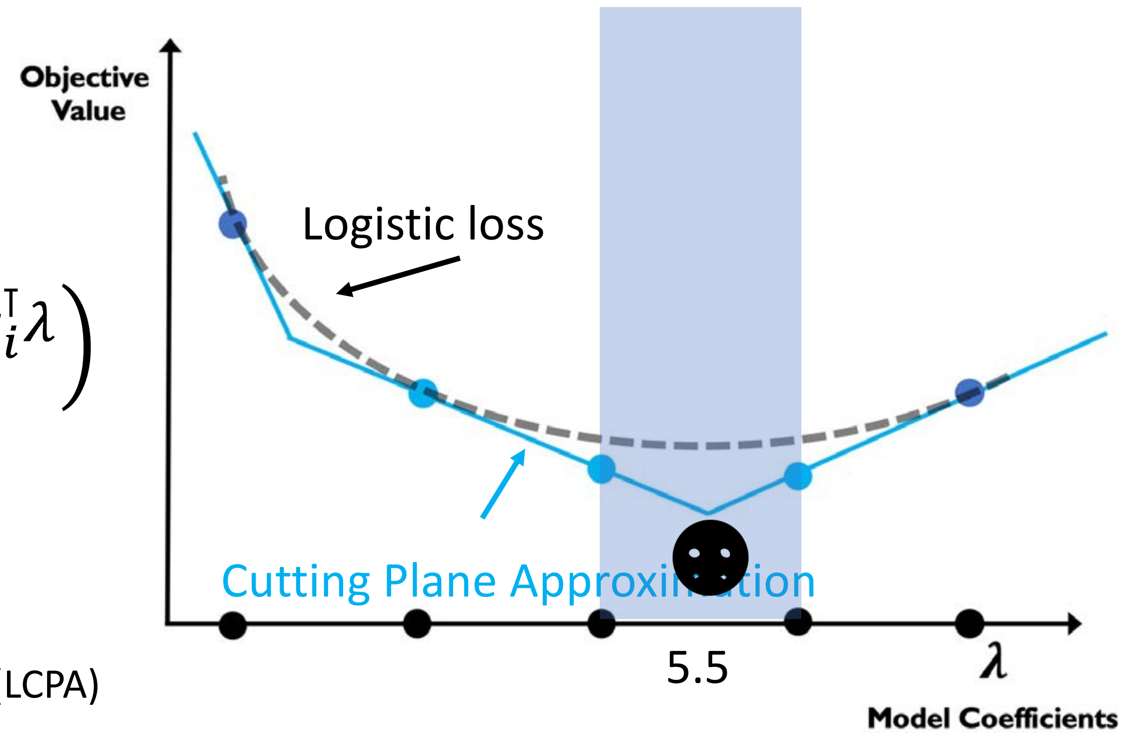


No Stalling for Lattice Cutting Plane Algorithm

Risk-SLIM

(Ustun, R, JMLR 2019)

$$\min_{\lambda\in L}\sum_{i=1}^{n}\log\left(1+e^{-y_i\boldsymbol{x}_i^{\top}\lambda}\right)$$



RiskSLIM's Lattice Cutting Plane Algorithm (LCPA)

1 2 1 5.5 6.3 3.8 1 0 9 7

≤ 5 ≥ 6

2 subproblems

If a subproblem leads to a feasible integer solution, add a cutting plane.

Otherwise split into 2 subproblems (linear programs).

If min cutting planes = objective, solved!

Risk-SLIM

(Ustun, R, JMLR 2019)

- LCPA is the only method that generates solutions within a reasonable time.
  - MINLP solvers don't work
  - standard cutting planes require solving larger and larger MIPs.

ADHD Screening

Clock Drawing Test

JAMA Psychiatry | Search All | Enter Search T

This Issue | Views 39,912 | Citations 82 | Altmetric 519

PDF | More | Cite | Permissions

Original Investigation | FREE

April 5, 2017

The World Health Organization Adult Attention-Deficit/Hyperactivity Disorder Self-Report Screening Scale for *DSM-5*

Berk Ustun, MS[1]; Lenard A. Adler, MD[2,3]; Cynthia Rudin, PhD[4,5]; et al

Learning Classification Models of Cognitive Conditions from Subtle Behaviors in the Digital Clock Drawing Test

William Souillard-Mandar · Randall Davis · Cynthia Rudin · Rhoda Au · David J. Libon · Rodney Swenson · Catherine C. Price · Melissa Lamar · Dana L. Penney ·

POPULAR SCIENCE | SCIENCE | TECH | DIY | REVIEWS

New Computer Tool Can Predict Dementia From Your Simple Drawings

An old test gets a techy update

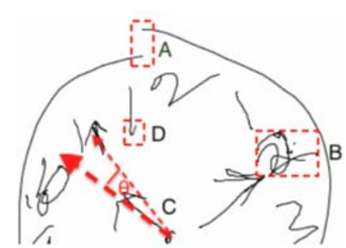
BY ALEXANDRA OSSOLA AUGUST 13, 2015

HEALTH

(b) Alzheimer's Disease

(c) Parkinson's Disease

The tests used to detect cognitive changes that often signal the onset of diseases like Parkinson's or Alzheimer's are surprisingly simple, usually only involving a pencil and paper. But they are very limited and not sensitive enough to pick up subtle neurological changes before disease fully sets in. Now researchers at MIT have created a model by using machine learning to assess the written tests so that clinicians can make diagnoses more quickly and objectively. The research was published recently in the journal *Machine Learning*.



Drawing elements recorded by the digital pen

Sleep Apnea Screening

npr | HEAR EVERY VOICE | WUNC

YOUR HEALTH

Do You Zone Out? Procrastinate? Might Be Adult ADHD

April 5, 2017 · 12:00 PM ET

REBECCA HERSHER

Do you pop up from your seat during meetings and finish other people's sentences? And maybe you also procrastinate, or find yourself zoning out in the middle of one-on-one conversations?

It's possible you have adult ADHD.

Six simple questions can reliably identify adults with attention-deficit/hyperactivity disorder, according to a World Health Organization advisory group working with two additional psychiatrists.

> J Clin Sleep Med. 2016 Feb;12(2):161-8. doi: 10.5664/jcsm.5476.

Clinical Prediction Models for Sleep Apnea: The Importance of Medical History over Symptoms

Berk Ustun [1], M Brandon Westover [2], Cynthia Rudin [3], Matt T Bianchi [2] [4]

Affiliations + expand

PMID: 26350602 PMCID: PMC4751423 DOI: 10.5664/jcsm.5476

Free PMC article

Journal of the Royal Statistical Society, 2017

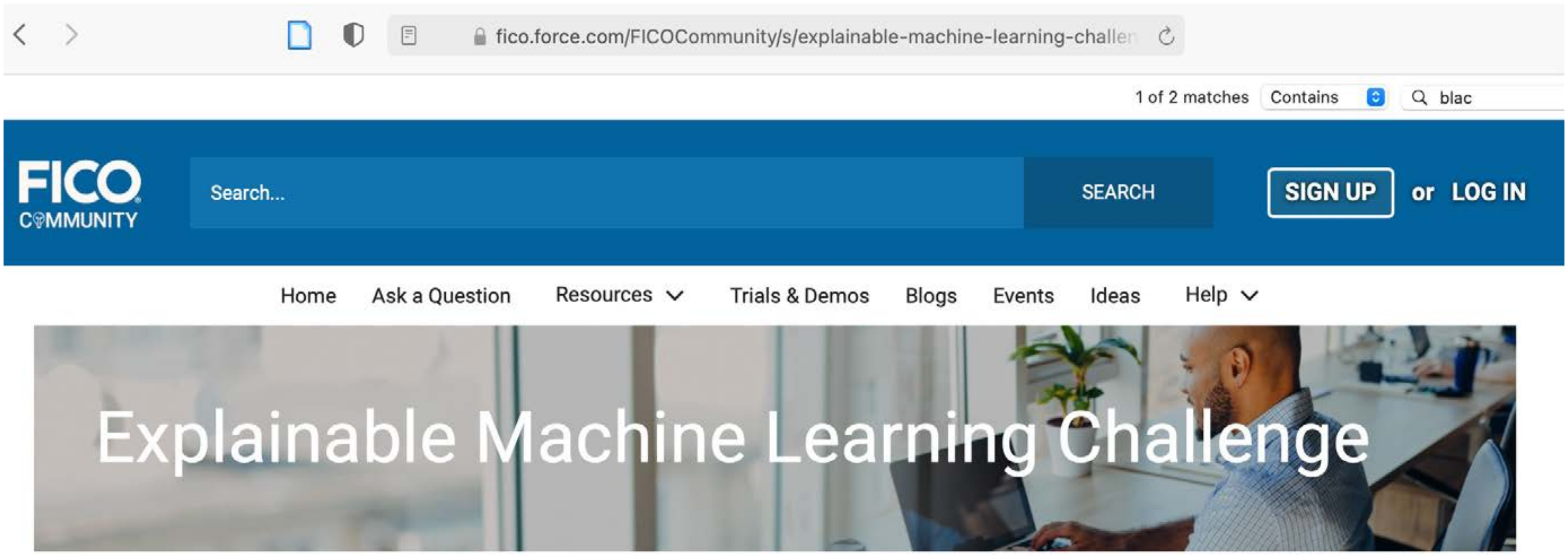
Interpretable Classification Models for Recidivism Prediction

Jiaming Zeng†, Berk Ustun†, Cynthia Rudin

†These authors contributed equally to this work.

Summary. 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 case might have an

Could interpretable models *really* be as accurate as black box models?



Home Equity Line of Credit (HELOC) Dataset

This competition focuses on an anonymized dataset of Home Equity Line of Credit (HELOC) applications made by real homeowners. A HELOC is a line of credit typically offered by a bank as a percentage of home equity (the difference between the current market value of a home and its purchase price). The customers in this dataset have requested a credit line in the range of $5,000 - $150,000. The fundamental task is to use the information about the applicant in their credit report to predict whether they will repay their HELOC account within 2 years. This prediction is then used to decide whether the homeowner qualifies for a line of credit and, if so, how much credit should be extended.

About the data

- ~10K loan applicants
- Factors:
  - External Risk Estimate
  - Months Since Oldest Trade Open
  - Months Since Most Recent Trade Open
  - Average Months In File
  - Number of Satisfactory Trades
  - Number Trades 60+ Ever
  - Number Trades 90+ Ever
  - Number of Total Trades
  - Number Trades Open In Last 12 Months
  - Percent Trades Never Delinquent
  - Months Since Most Recent Delinquency
  - Max Delinquency / Public Records Last 12 Months
  - Max Delinquency Ever
  - Percent Installment Trades
  - Net Fraction of Installment Burden
  - Number of Installment Trades with Balance
  - Months Since Most Recent Inquiry excluding 7 days
  - Number of Inquiries in Last 6 Months
  - Number of Inquiries in Last 6 Months excluding 7 days.
  - Net Fraction Revolving Burden. (Revolving balance divided by credit limit.)
  - Number Revolving Trades with Balance
  - Number Bank/Natl Trades with high utilization ratio
  - Percent of Trades with Balance

Best black box accuracy
(boosted decision trees) 73%

Best black box AUC
(2-layer neural network) .80

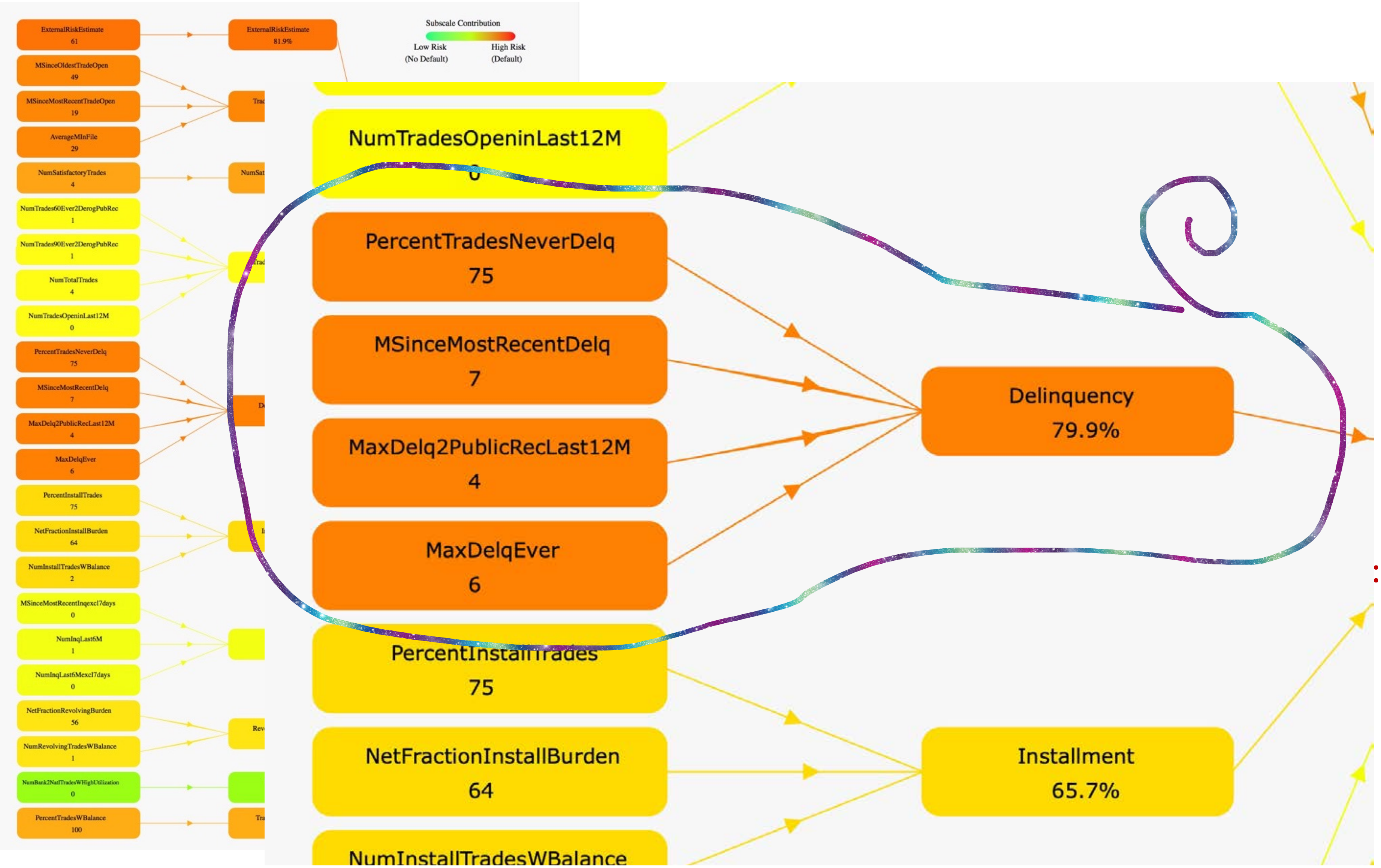
Best black box accuracy
(boosted decision trees) 73%

Best black box AUC
(2-layer neural network) .80

IBM model (First Prize): 6 questions
Accuracy = 71.8%
AUC = .62

Our entry (won FICO Recognition Prize):
Two-layer additive risk model
10 subscales + one final scoring model

Accuracy = 73.8%
AUC = .806



Subscale Contribution
Low Risk
(No Default)
High Risk
(Default)
ExternalRiskEstimate
61
ExternalRiskEstimate
81.9%
MSinceOldestTradeOpen
49
MSinceMostRecentTradeOpen
19
AverageMInFile
29
NumSatisfactoryTrades
4
NumTrades60Ever2DerogPubRec
1
NumTrades90Ever2DerogPubRec
1
NumTotalTrades
4
NumTradesOpeninLast12M
0
PercentTradesNeverDelq
75
MSinceMostRecentDelq
7
MaxDelq2PublicRecLast12M
4
MaxDelqEver
6
PercentInstallTrades
75
NetFractionInstallBurden
64
NumInstallTradesWBalance
2
MSinceMostRecentInqexcl7days
0
NumInqLast6M
1
NumInqLast6Mexcl7days
0
NetFractionRevolvingBurden
56
NumRevolvingTradesWBalance
1
NumBank2NatlTradesWHighUtilization
0
PercentTradesWBalance
100
NumTradesOpeninLast12M
0
PercentTradesNeverDelq
75
MSinceMostRecentDelq
7
MaxDelq2PublicRecLast12M
4
MaxDelqEver
6
Delinquency
79.9%
PercentInstallTrades
75
NetFractionInstallBurden
64
Installment
65.7%
NumInstallTradesWBalance

Delinquency Subscale

Intervals	Points
0-59	+1.567
59-84	+1.012
84-89	+0.601
89-96	+0.366
96-Inf	-0.147
-7	0
-8	0
-9	0

PercentTradesNeverDelq

Intervals	Points
0-8	-0.058
9-17	-0.058
18-32	-0.22
33-47	-0.392
48-Inf	-0.482
-7	+0.198
-8	+0.137
-9	0

MSinceMostRecentDelq

Intervals	Points
0-3	+0.806
4-5	+0.806
6	+0.408
7-8	-0.147
9-Inf	-0.147
-7	0
-8	0
-9	0

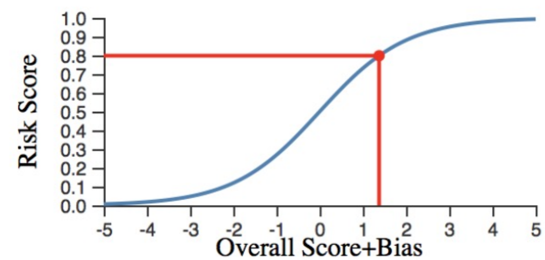
MaxDelq2PublicRecLast12M

Intervals	Points
0-2	-0.017
3	-0.147
4-5	-0.147
6	-0.147
7-Inf	-0.147
-7	0
-8	0
-9	0

MaxDelqEver

Overall Score	1.613
Bias	-0.237
Associated Risk (for subscale Delinquency)	**79.8%**

Activation Function



Best black box accuracy
(boosted decision trees) 73%

Best black box AUC
(2-layer neural network) .80

IBM model (First Prize): 6 questions
Accuracy = 71.8%
AUC = .62

Our entry (won FICO Recognition Prize):
Two-layer additive risk model
10 subscales + one final scoring model

Accuracy = 73.8%
AUC = .806

Even on challenging benchmark datasets,

interpretable models' accuracy = black box accuracy.

Journal of the Royal Statistical Society, 2017

Interpretable Classification Models for Recidivism Prediction

Jiaming Zeng†, Berk Ustun†, Cynthia Rudin

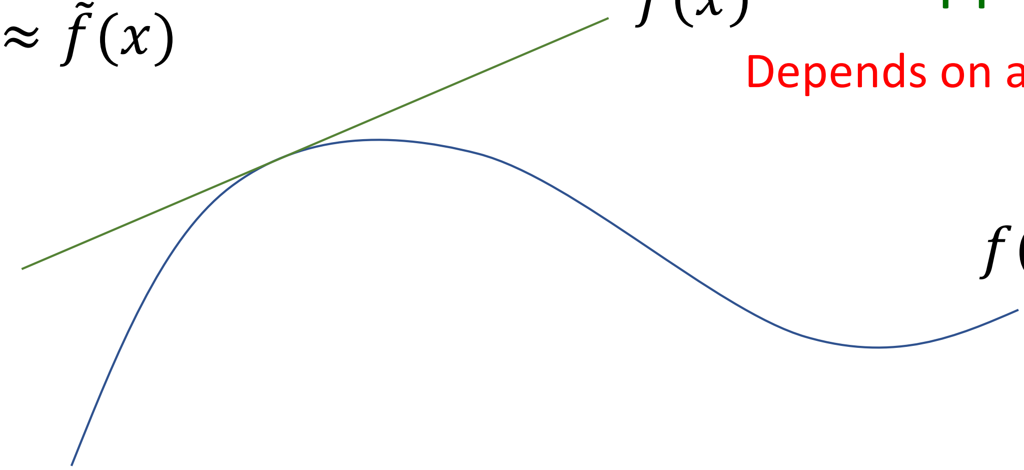
†These authors contributed equally to this work.

Summary. 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 case might have an

Interpretable Models ≠ Explanations of Black Box Models

Approximations are not "explanations"!

$f(x) \approx \tilde{f}(x)$



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.

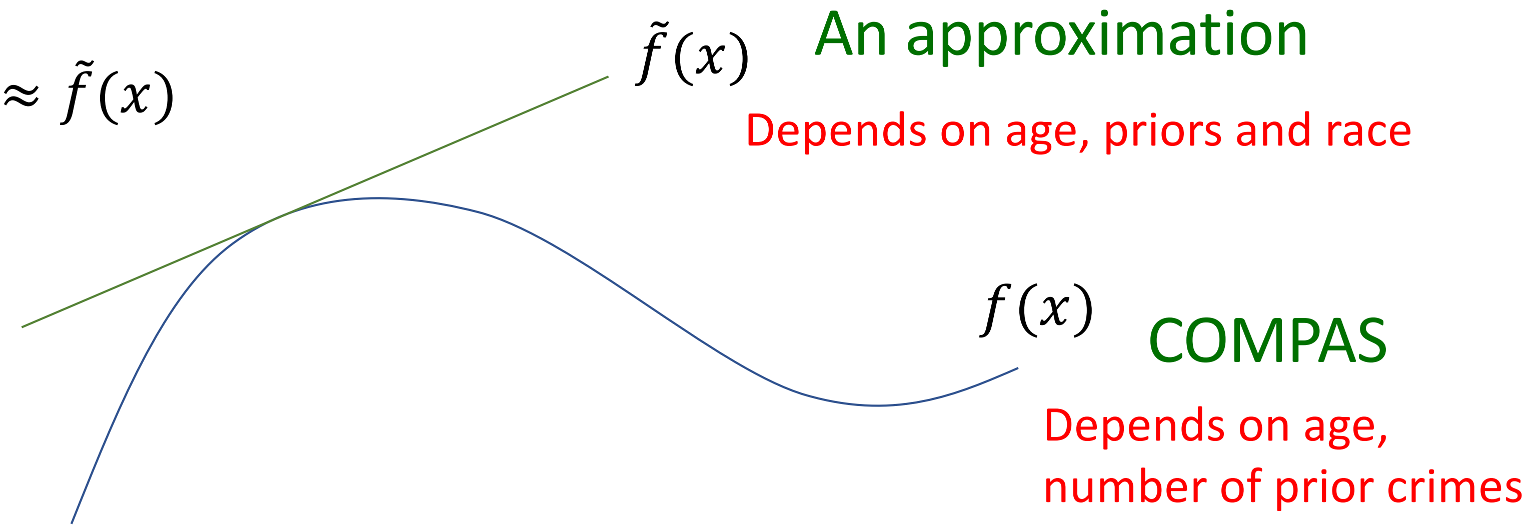
by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica

May 23, 2016

Interpretable Models ≠ Explanations of Black Box Models

- **Approximations are not "explanations"**!

$$f(x) \approx \tilde{f}(x)$$

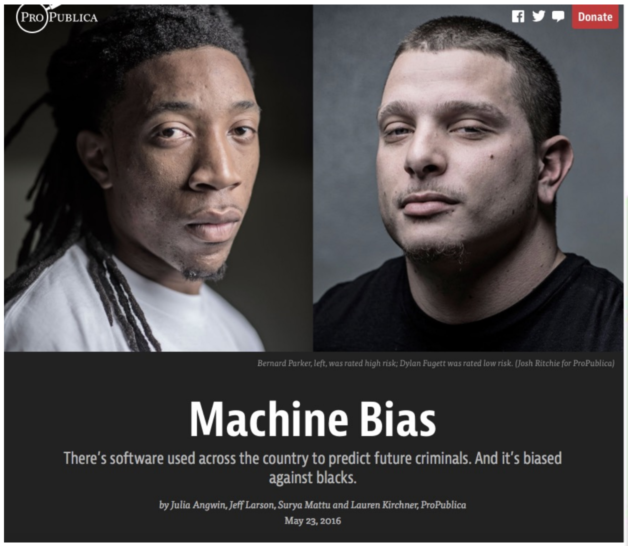


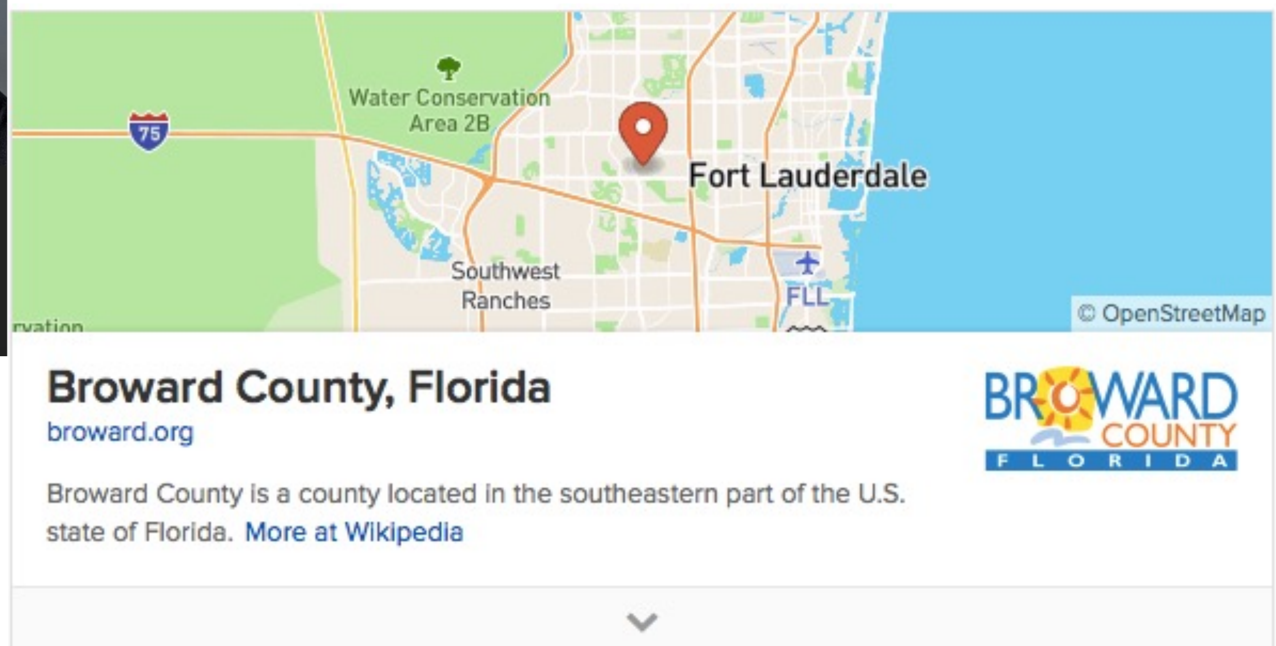
What ProPublica Did

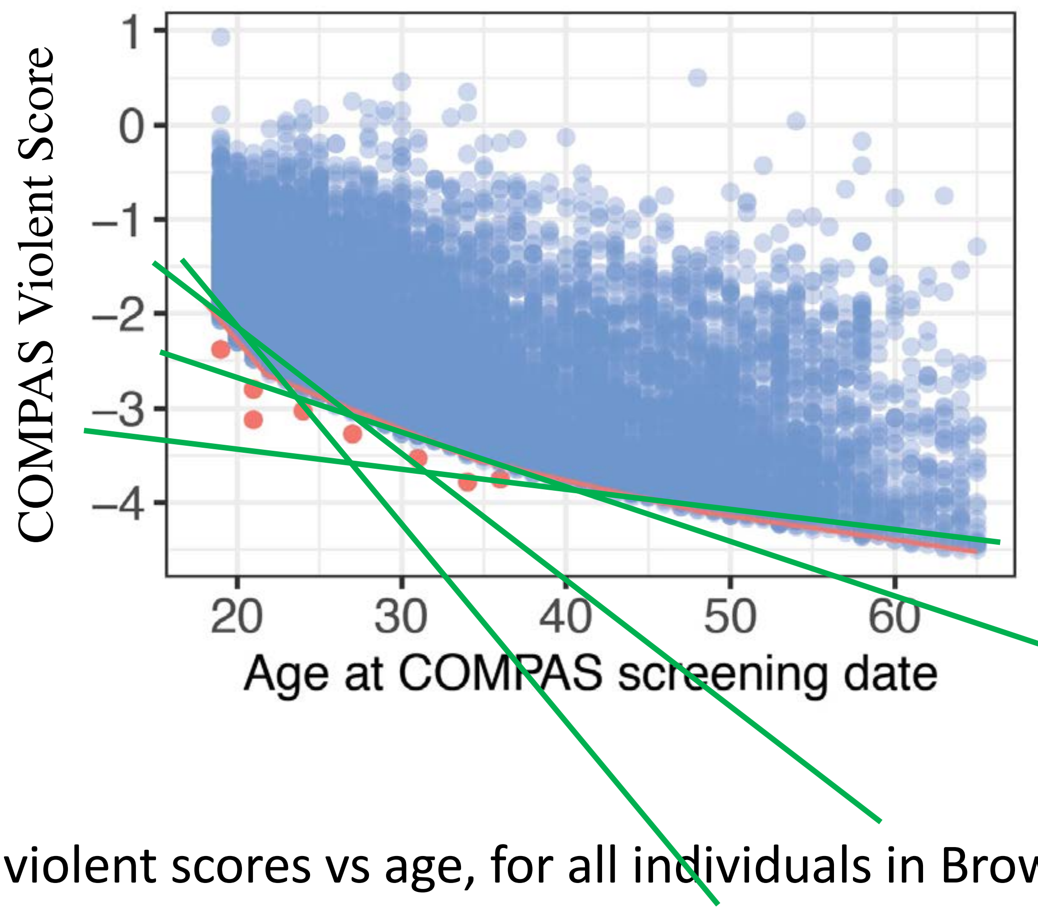
- They showed that FPR and FNR of COMPAS varied by race.
- They suggested maybe this might not be a good comparison, we should include age and number of priors and reexamine.
- After including age and number of priors, still found a linear approximation to COMPAS with a low pvalue for the race covariate.
  - We don't think COMPAS is linear
- Concluded that COMPAS depends on race.
  - Bad idea

Rudin, Wang, and Coker. The *Age* of Secrecy and Unfairness in Recidivism Prediction. Harvard Data Science Review, 2020

A peek inside COMPAS?







COMPAS violent scores vs age, for all individuals in Broward County FL.

Rudin, Wang, and Coker. The *Age* of Secrecy and Unfairness in Recidivism Prediction. Harvard Data Science Review, 2020

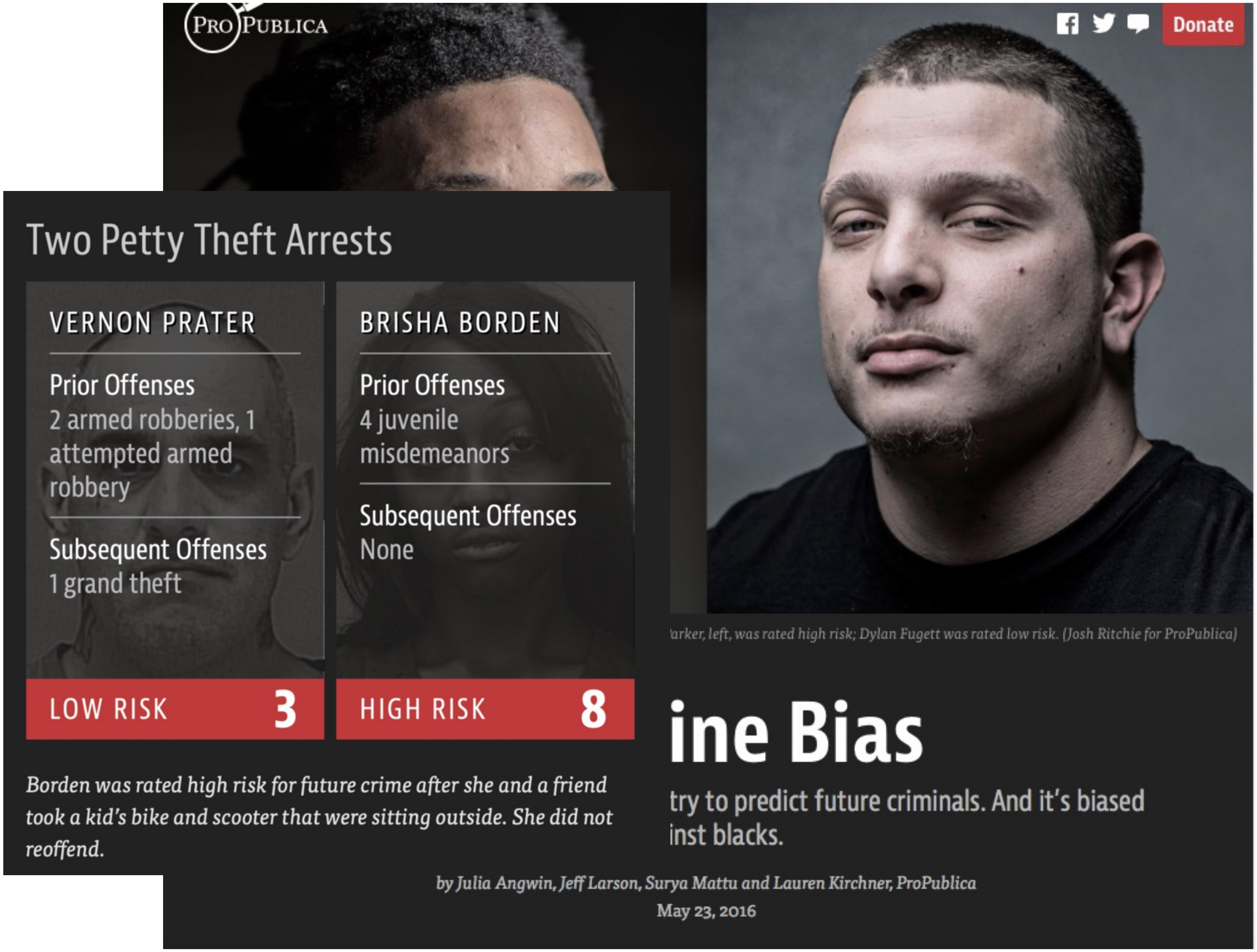
A peek inside COMPAS?

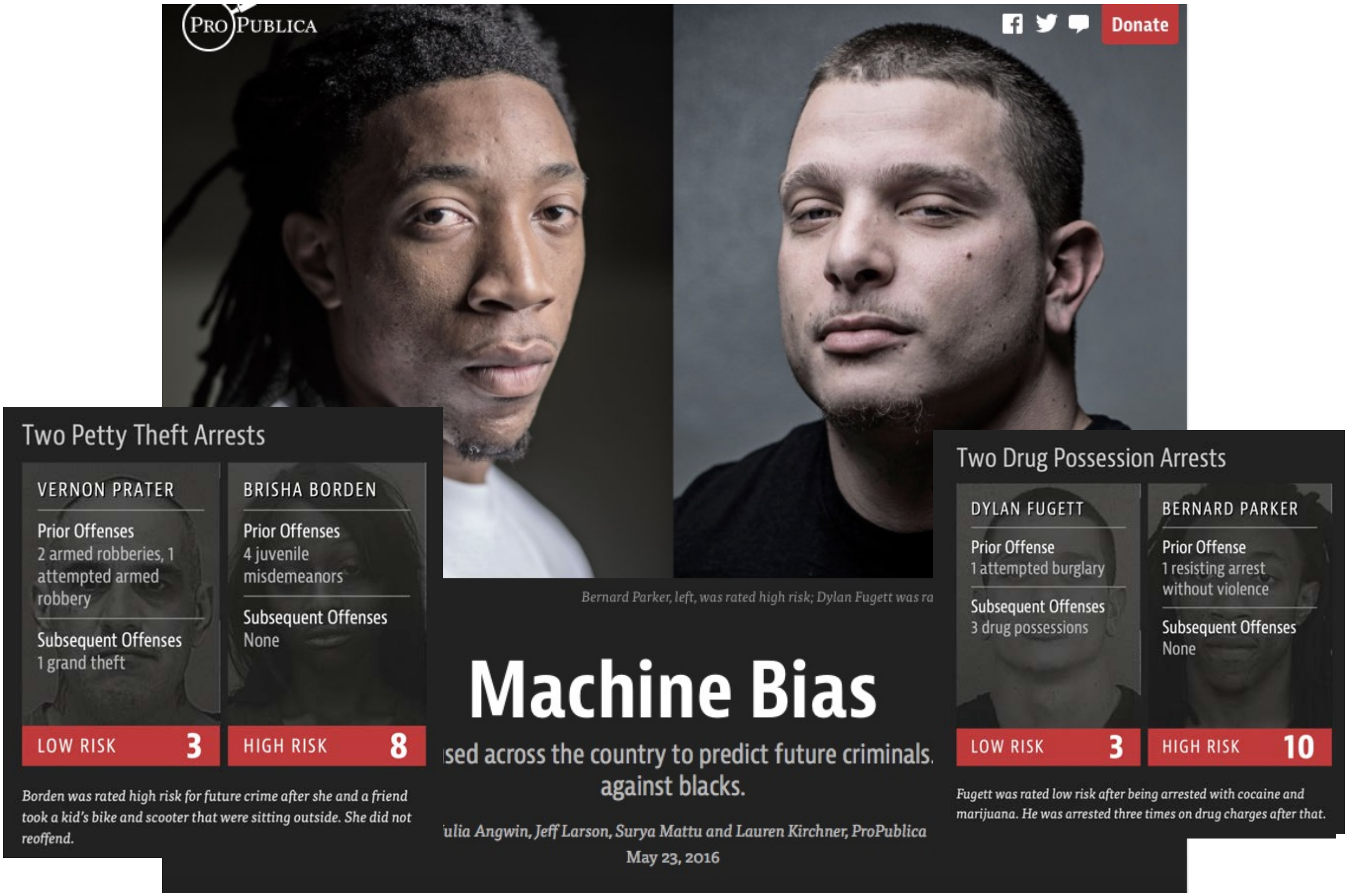
Does COMPAS $- f_{\text{age}}$ depend on race?

It doesn't seem to.

(We ran machine learning methods *with and without race* to see if they need race to predict COMPAS well. They performed similarly.)

Rudin, Wang, and Coker. The *Age* of Secrecy and Unfairness in Recidivism Prediction. Harvard Data Science Review, 2020


PRO PUBLICA
Donate
Two Petty Theft Arrests
VERNON PRATER
Prior Offenses
2 armed robberies, 1 attempted armed robbery
Subsequent Offenses
1 grand theft
LOW RISK 3
BRISHA BORDEN
Prior Offenses
4 juvenile misdemeanors
Subsequent Offenses
None
HIGH RISK 8
Borden was rated high risk for future crime after she and a friend took a kid's bike and scooter that were sitting outside. She did not reoffend.
arker, left, was rated high risk; Dylan Fugett was rated low risk. (Josh Ritchie for ProPublica)
ine Bias
try to predict future criminals. And it's biased
inst blacks.
by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica
May 23, 2016



ProPublica
Donate
Two Petty Theft Arrests
VERNON PRATER
Prior Offenses
2 armed robberies, 1 attempted armed robbery
Subsequent Offenses
1 grand theft
LOW RISK 3
BRISHA BORDEN
Prior Offenses
4 juvenile misdemeanors
Subsequent Offenses
None
HIGH RISK 8
Borden was rated high risk for future crime after she and a friend took a kid's bike and scooter that were sitting outside. She did not reoffend.
Bernard Parker, left, was rated high risk; Dylan Fugett was ra
Machine Bias
sed across the country to predict future criminals.
against blacks.
ulia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica
May 23, 2016
Two Drug Possession Arrests
DYLAN FUGETT
Prior Offense
1 attempted burglary
Subsequent Offenses
3 drug possessions
LOW RISK 3
BERNARD PARKER
Prior Offense
1 resisting arrest without violence
Subsequent Offenses
None
HIGH RISK 10
Fugett was rated low risk after being arrested with cocaine and marijuana. He was arrested three times on drug charges after that.

137 factors entered by hand for each survey

1% error rate → 75% chance of at least one typo on a survey

This is a serious disadvantage to complicated or proprietary models.

In Florida….?

The New York Times

OP-ED CONTRIBUTOR

When a Computer Program Keeps You in Jail

By Rebecca Wexler

June 13, 2017

232

Name	COMPAS Violent Decile	# Arrests	# Charges	Selected Prior Charges	Selected Subseq. Charges
Shirley Darby	1	2	4	Aggravated Battery (F,1), Child Abuse (F,1), Resist Officer w/Violence (F,1)	
Joseph Salera	1	8	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)	
Bart Sandell	1	9	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)
Miguel Wilkins	1	11	22	Aggrav Battery w/Deadly Weapon (F,1), Driving Under The Influence (M,2), Carrying Concealed Firearm (F,1)	
Jonathan Gabbard	1	7	28	Robbery / Deadly Weapon (F,11), Poss Firearm Commission Felony (F,7)	
Brandon Jackel	1	22	40	Resist/obstruct Officer W/viol (F,3), Battery on Law Enforc Officer (F,2), Attempted Robbery Deadly Weapon (F,1), Robbery 1 / Deadly Weapon (F,1)	
Fernando Galarza	2	2	6	Murder in the First Degree (F,1), Aggrav Battery w/Deadly Weapon (F,1), Carrying Concealed Firearm (F,1)	

Continued on next page

Name	COMPAS Violent Decile	# Arrests	# Charges	Selected Prior Charges	Selected Subseq. Charges
Nathan Keller	2	8	17	Aggravated Assault (F,5), Aggravated Assault W/dead Weap (F,2), Shoot/throw Into Vehicle (F,2), Battery Upon Detainee (F,1)	
Zachary Campanelli	2	11	21	Armed Trafficking In Cocaine (F,1), Poss Weapon Commission Felony (F,1), Carrying Concealed Firearm (F,1)	
Aaron Coleburn	2	16	25	Attempt Murder in the First Degree (F,1), Carrying Concealed Firearm (F,1), Felon in Pos of Firearm or Amm (F,1)	
Bruce Poblano	2	22	39	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)
Phillip Sperry	3	11	16	Aggravated Assault W/dead Weap (F,1), Burglary Damage Property>$1000 (F,1), Burglary Unoccupied Dwelling (F,1)	
Dylan Azzi	3	11	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)
Russell Michaels	3	9	23	Solicit to Commit Armed Robbery (F,1), Armed False Imprisonment (F,1), Home Invasion Robbery (F,1)	Driving While License Revoked (F,3)
Bradley Haddock	3	15	25	Attempt Sexual Batt / Vict 12+ (F,1), Resist/obstruct Officer W/viol (F,1), Poss Firearm W/alter/remov Id# (F,1)	
Randy Walkman	3	24	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)
Carol Hartman	4	5	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)

Possibly typos in the COMPAS documentation from Northpointe?

COMPAS Documentation

Violent Recidivism Risk Score

$$= (\text{age} * -\text{w}) + (\text{age-at-first-arrest} * -\text{w}) + (\text{history of violence} * \text{w})$$
$$+ \quad (\text{vocation education} * \text{w}) + (\text{history of noncompliance} * \text{w})$$

Corrected version?

Violent Recidivism Risk Score

$$= (f(\text{age}) * -\text{w}) + (g(\text{age-at-first-arrest}) * -\text{w}) + (\text{history of violence} * \text{w})$$
$$+ \quad (\text{vocation education} * \text{w}) + (\text{history of noncompliance} * \text{w})\ ,$$

where *f* and *g* are proprietary transformations of age, such as linear splines?

HDSR

Search

HOME JOURNAL CATEGORIES TOPICS MEDIA FEATURES PODCAST SUBMISSIONS ABOUT MA

Issue 2.1, Winter 2020 ··· 2 more **Published on** Mar 31, 2020 **DOI** 10.1162/99608f92.6ed64b30

The Age of Secrecy and Unfairness in Recidivism Prediction

by Cynthia Rudin, Caroline Wang, and Beau Coker

Published on Mar 31, 2020

A **Commentary** on this Pub

“Nothing Is More Opaque than Absolute Transparency”: The Use of Prior History to Guide Sentencing

by Shawn D. Bushway

Published on Mar 31, 2020 · hdsr.mitpress.mit.edu

A **Commentary** on this Pub

Transparency and Simplicity in Criminal Risk Assessment

by Alexandra Chouldechova

Published on Mar 31, 2020 · hdsr.mitpress.mit.edu

A **Commentary** on this Pub

The Role of Risk Assessment in the Criminal Justice System: Moving Beyond a Return to the Status Quo

by Sarah L. Desmarais

Published on Mar 31, 2020 · hdsr.mitpress.mit.edu

A **Commentary** on this Pub

Justice in Forensic Algorithms

by Brandon L. Garrett

Published on Mar 31, 2020 · hdsr.mitpress.mit.edu

A **Commentary** on this Pub

Setting the Record Straight: What the COMPAS Core Risk and Need Assessment Is and Is Not

by Eugenie Jackson and Christina Mendoza

Published on Mar 31, 2020 · hdsr.mitpress.mit.edu

A **Commentary** on this Pub

Transparency, Statistics, and Justice System Knowledge Is Essential for Science of Risk Assessment

by Greg Ridgeway

Published on Jan 31, 2020 · hdsr.mitpress.mit.edu

HDSR Search Dashb

HOME JOURNAL CATEGORIES TOPICS MEDIA FEATURES PODCAST SUBMISSIONS ABOUT MASTHEAI

Issue 2.1, Winter 2020 Commentary Published on Mar 31, 2020 DOI 10.1162/99608f92.1b3dadaa

Setting the Record Straight: What the COMPAS Core Risk and Need Assessment Is and Is Not

by Eugenie Jackson and Christina Mendoza

Published on Mar 31, 2020

assumption regarding the age dependence in risk scores. The authors have taken a clearly informal description of the VRRS score in the *Practitioner's Guide to COMPAS Core* (Northpointe, 2019) for a complete technical description of the VRRS model. This guide is written for practitioners and is not intended to be a technical document. Discussions of appropriate variable transformations are beyond its scope and would not further its goals; however, we note that the skewed age variable is an ideal candidate for a normalizing transformation (see Figure A3 in authors' Appendix)[12].

So there *is* a typo in the practitioners guide!

4. Transparency

Striking a balance between protecting the investments made in developing the risk assessments and allowing increased transparency has been a goal of Northpointe for some time. Northpointe and its parent company, equivant, are pursuing copyrights for the GRRS and VRRS. A feature that has been

Whoa!!

Summary

Scoring systems are good, typos are bad

(when you optimize them)

(which happen more often with complicated or black box models)

Interpretable Machine Learning Lab

1.	Any cEEG Pattern with Frequency 2 Hz	1 point	···
2.	Epileptiform Discharges	1 point	+ ···
3.	Patterns include [LPD, LRDA, BIPD]	1 point	+ ···
4.	Patterns Superimposed with Fast or Sharp Activity	1 point	+ ···
5.	Prior Seizure	1 point	+ ···
6.	Brief Rhythmic Discharges	2 points	+ ···
		SCORE	= ···

SCORE	0	1	2	3	4	5	6+
RISK	<5%	11.9%	26.9%	50.0%	73.1%	88.1%	95.3%

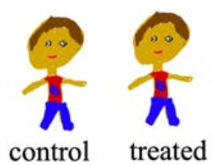
Scoring Systems
(healthcare, criminal justice)

Optimal Sparse
Decision Trees
(materials science)

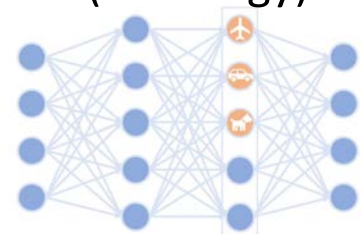
Data Visualization/
Dimension Reduction
(biology)

Thanks

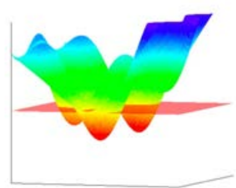
Interpretable Neural Networks for
Computer Vision
(radiology)



Almost Exact Matching for Causal Inference
(criminal justice)



Understanding the
Set of Good Models
and Importance of Variables



Neural Disentanglement