National Human Services Data Consortium

2018 Spring Conference

Pittsburgh, PA

April 18-19, 2018

Analyzing Systemic Racial Disparities With Statistical Learning Models and HMIS Data

Clayton Aldern

Pierce County, Washington | caldern@co.pierce.wa.us

New Frontiers in Data

New Systems, Partners, and Technologies

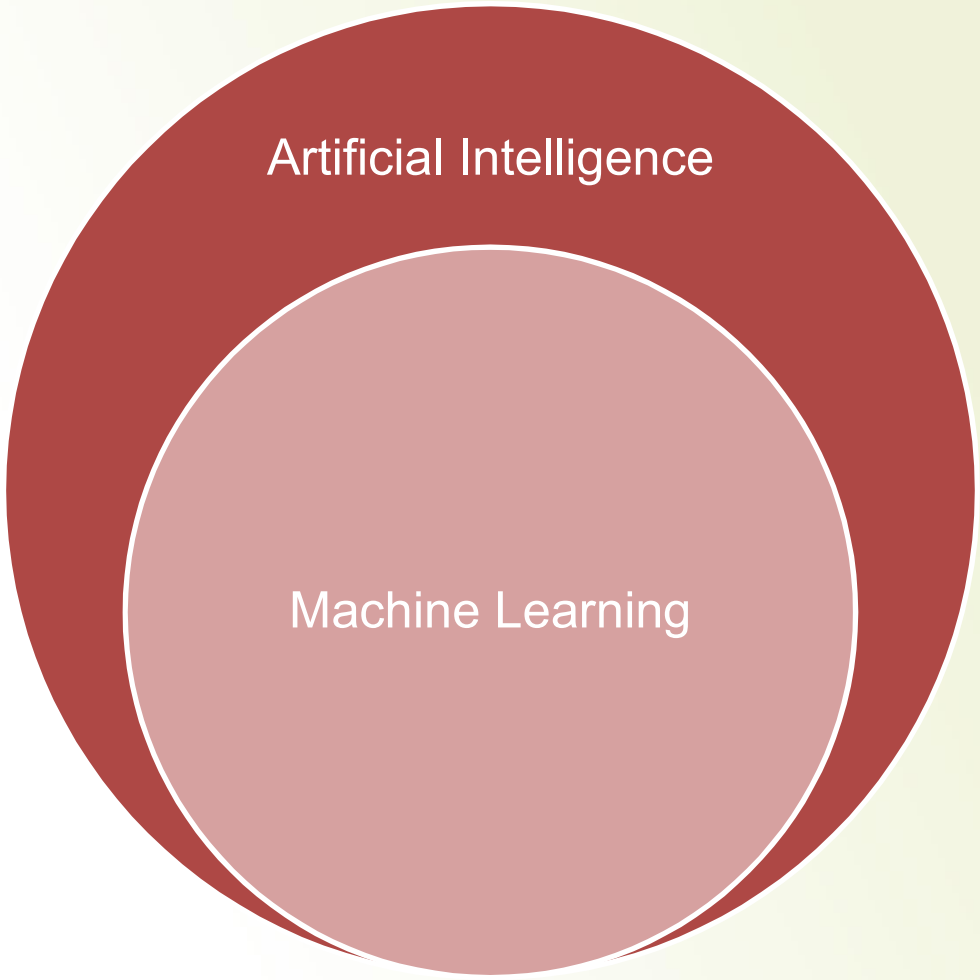
Sponsored by the National Human Services Data Consortium

Alternate title:
Machine learning isn't special, but it can be useful

I. What is a statistical learning model?
II. Case study
III. Tools

Analyzing Systemic Racial Disparities With Statistical Learning Models and HMIS Data

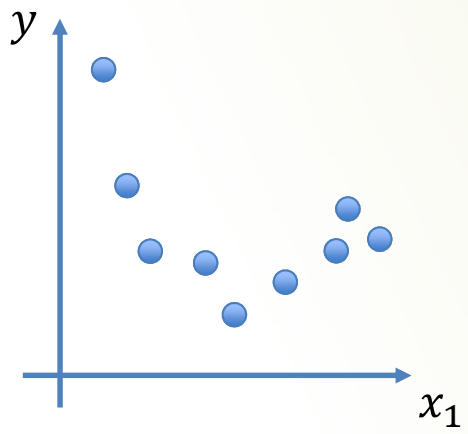
Clayton Aldern



We say a machine "learns" if it improves its performance *P* on a task *T*, given some experience *E*

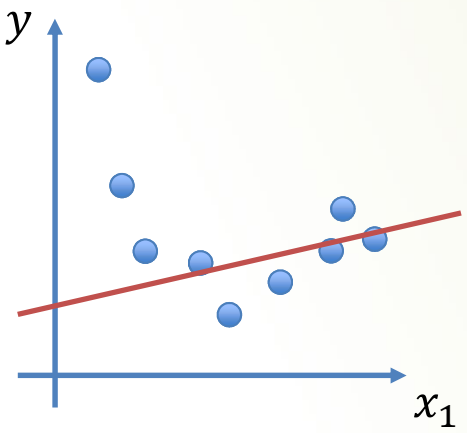
Analyzing Systemic Racial Disparities With Statistical Learning Models and HMIS Data

Clayton Aldern



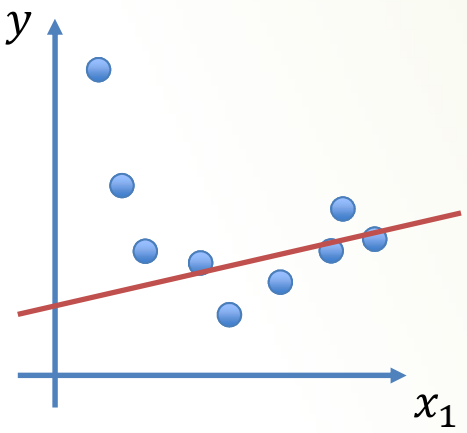
Analyzing Systemic Racial Disparities With Statistical Learning Models and HMIS Data

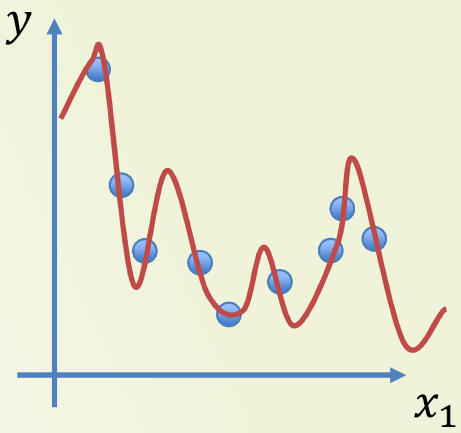
Clayton Aldern



Analyzing Systemic Racial Disparities With Statistical Learning Models and HMIS Data

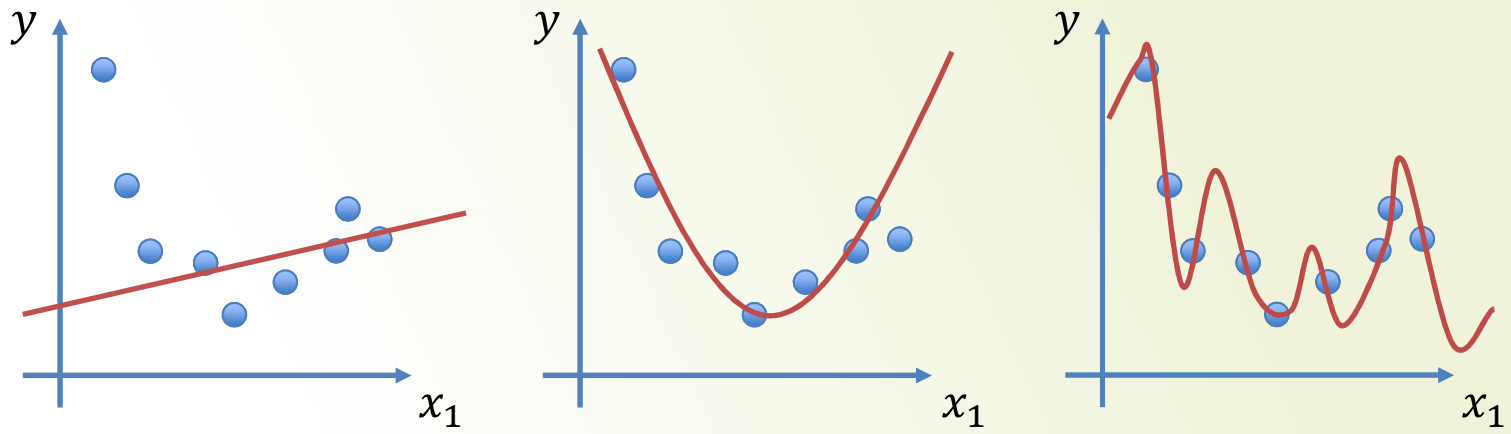
Clayton Aldern





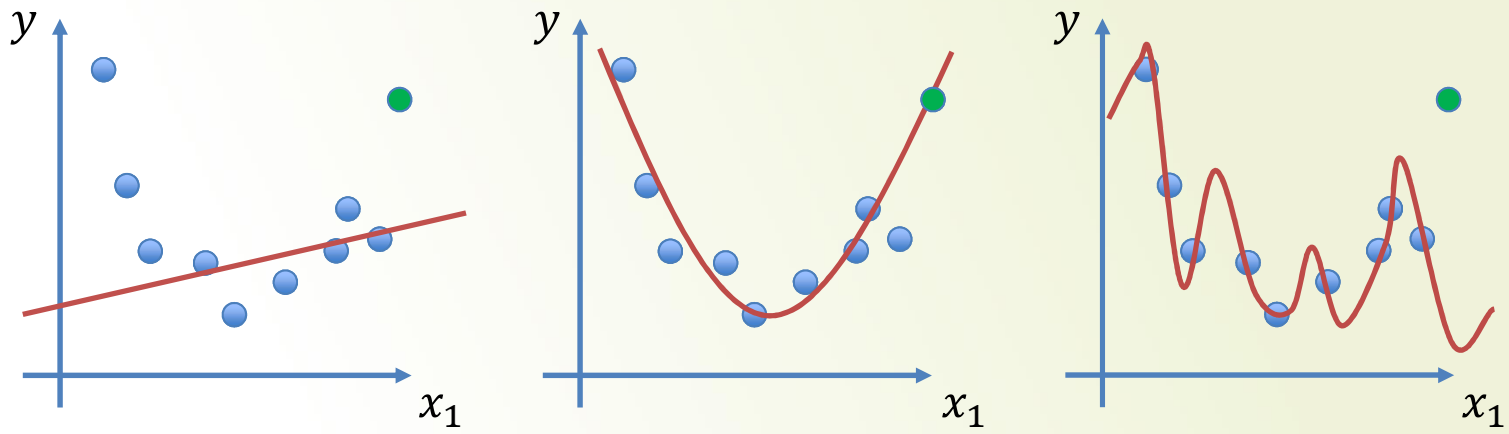
Analyzing Systemic Racial Disparities With Statistical Learning Models and HMIS Data

Clayton Aldern



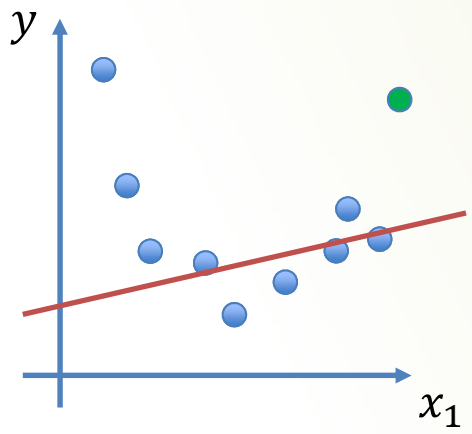
Analyzing Systemic Racial Disparities With Statistical Learning Models and HMIS Data

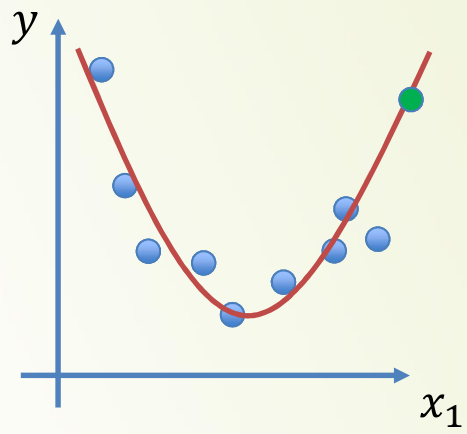
Clayton Aldern

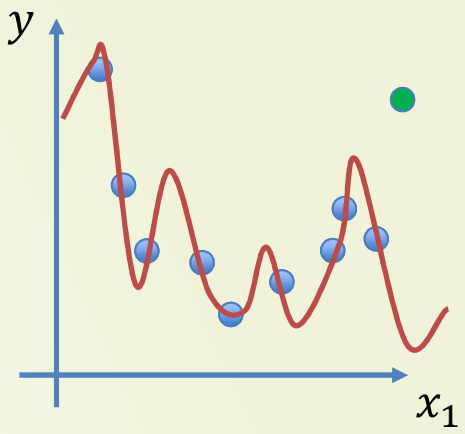


Analyzing Systemic Racial Disparities With Statistical Learning Models and HMIS Data

Clayton Aldern







"underfitting"

"overfitting"

What are the goals of model selection?

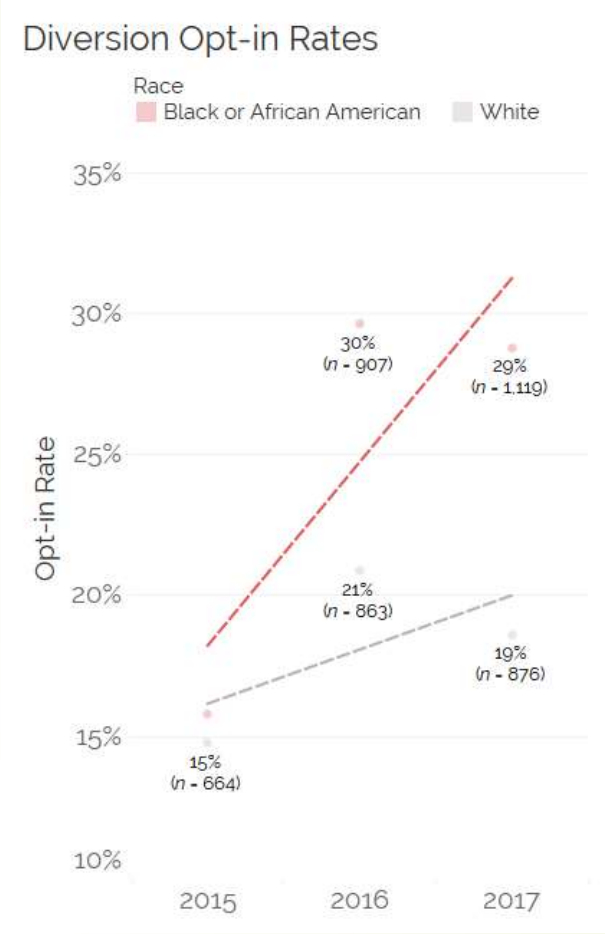
I. Construct a good predictor. (Values of model coefficients are irrelevant.)
II. Give causal interpretations of the factors (and determine which variables are "important").

(Bickel and Li, 2006)

Motivating observations

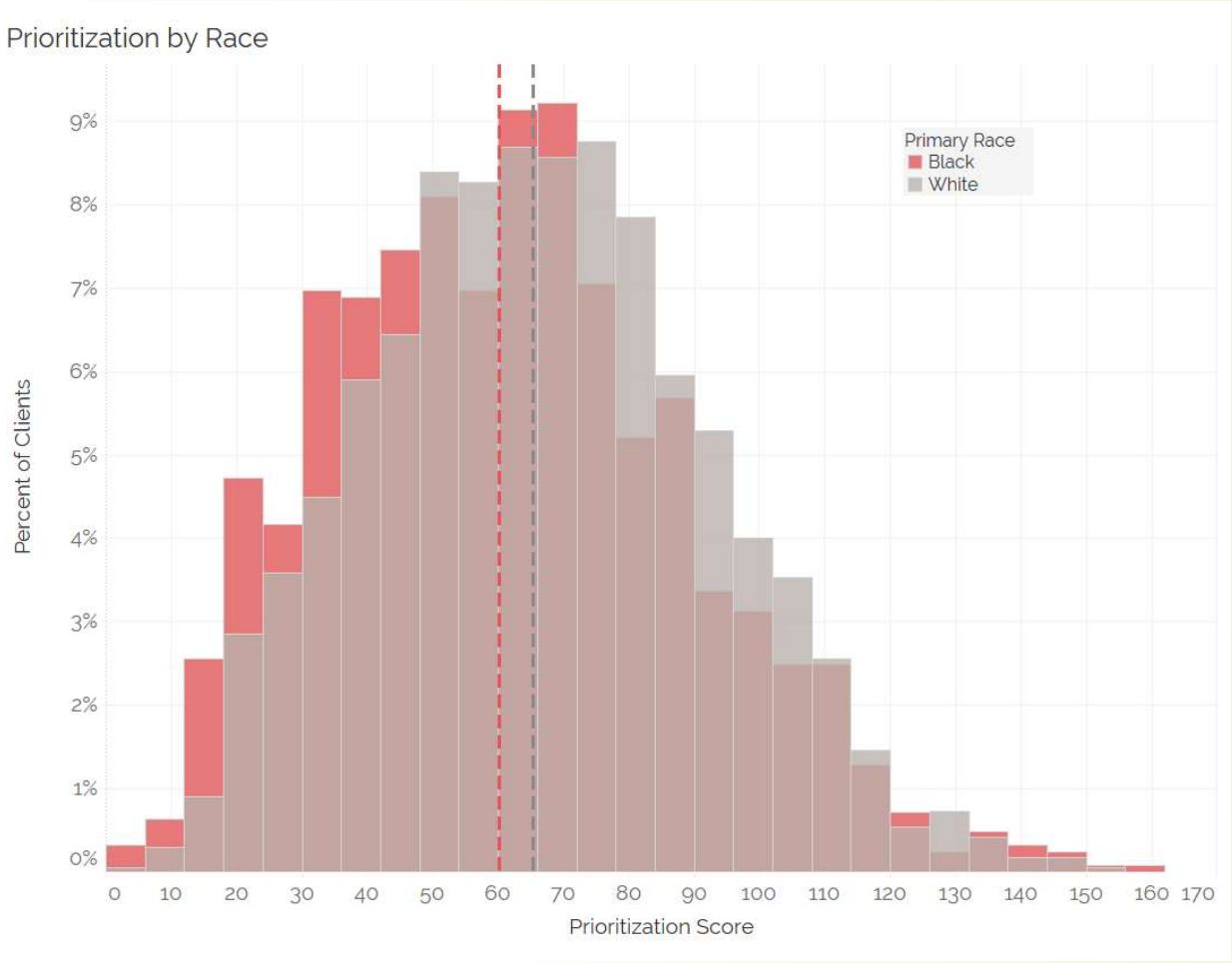
Analyzing Systemic Racial Disparities With Statistical Learning Models and HMIS Data

Clayton Aldern



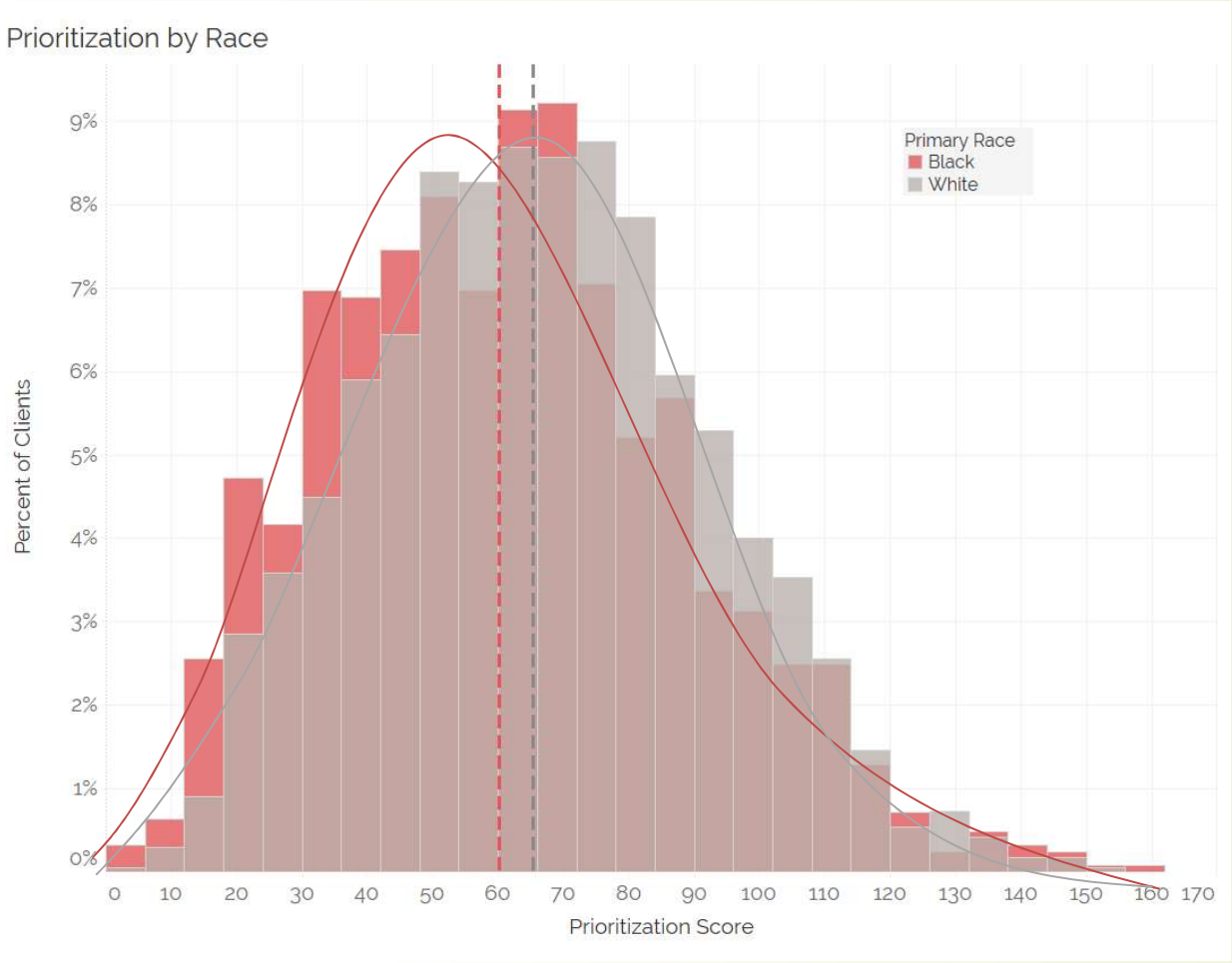
Analyzing Systemic Racial Disparities With Statistical Learning Models and HMIS Data

Clayton Aldern



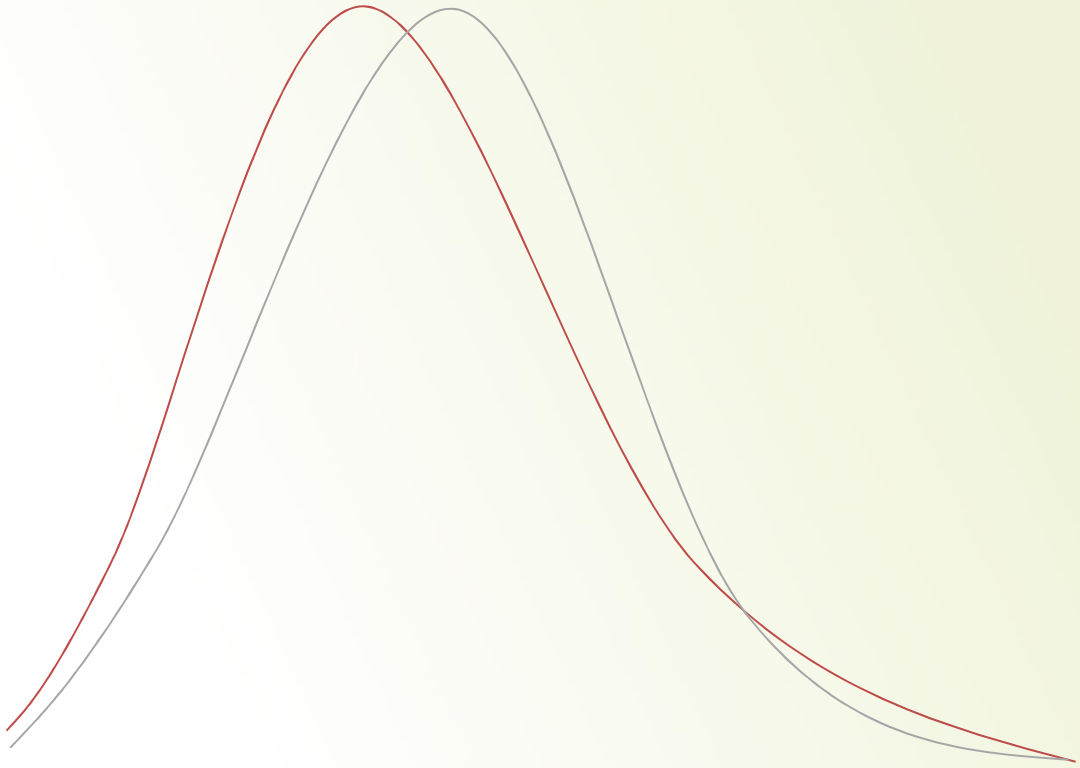
Analyzing Systemic Racial Disparities With Statistical Learning Models and HMIS Data

Clayton Aldern



Analyzing Systemic Racial Disparities With Statistical Learning Models and HMIS Data

Clayton Aldern



Which assessment questions and responses are significantly correlated with race?

Analyzing Systemic Racial Disparities
With Statistical Learning Models and HMIS Data
Clayton Aldern

Which assessment questions and responses are significantly correlated with race?

Can we predict someone's race, given only their assessment responses?

Analyzing Systemic Racial Disparities With Statistical Learning Models and HMIS Data

Clayton Aldern

Race	Age	Gender	…	Medical conditions in household
White	54	Male	…	4+
Black	32	Female	…	2
Black	33	Male	…	1
…	…	…	…	…
White	35	Female	…	3

Analyzing Systemic Racial Disparities With Statistical Learning Models and HMIS Data

Clayton Aldern

Race	Age	Gender	…	medcon 1	medcon 2	medcon 3	medcon 4+
1	54	1	…	0	0	0	1
0	32	0	…	0	1	0	0
0	33	1	…	1	0	0	0
…	…	…	…	…	…	…	…
1	35	0	…	0	0	1	0

Analyzing Systemic Racial Disparities With Statistical Learning Models and HMIS Data

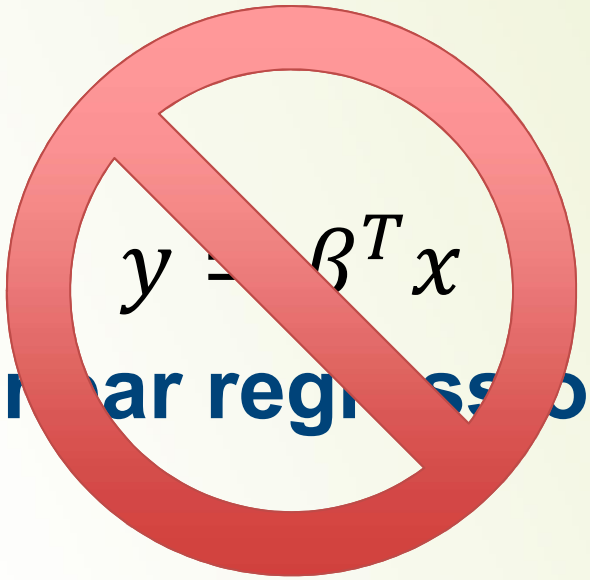
Clayton Aldern

y	x_1	x_2	...	x_{n-3}	x_{n-2}	x_{n-1}	x_n
1	54	1	...	0	0	0	1
0	32	0	...	0	1	0	0
0	33	1	...	1	0	0	0
...	...	...	...	...	...	...	...
1	35	0	...	0	0	1	0

$$y = \beta^T x$$

(linear regression)

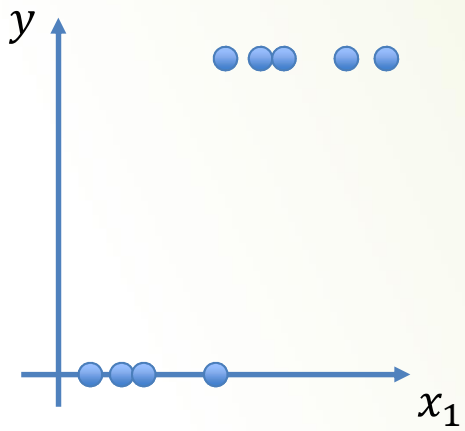
Analyzing Systemic Racial Disparities
With Statistical Learning Models and HMIS Data
Clayton Aldern



(linear regression)

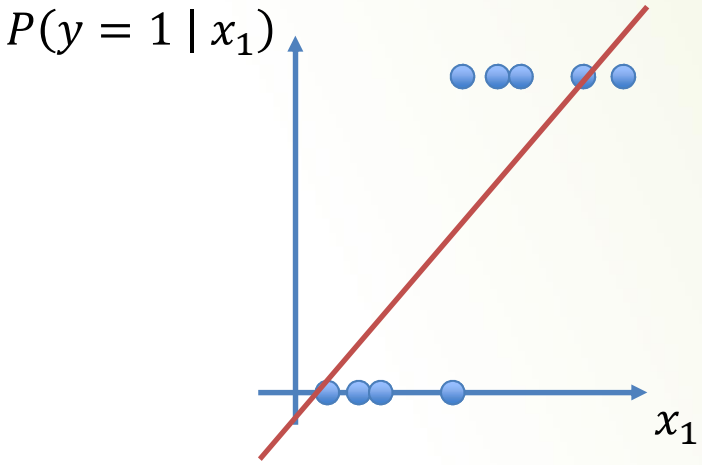
Analyzing Systemic Racial Disparities With Statistical Learning Models and HMIS Data

Clayton Aldern



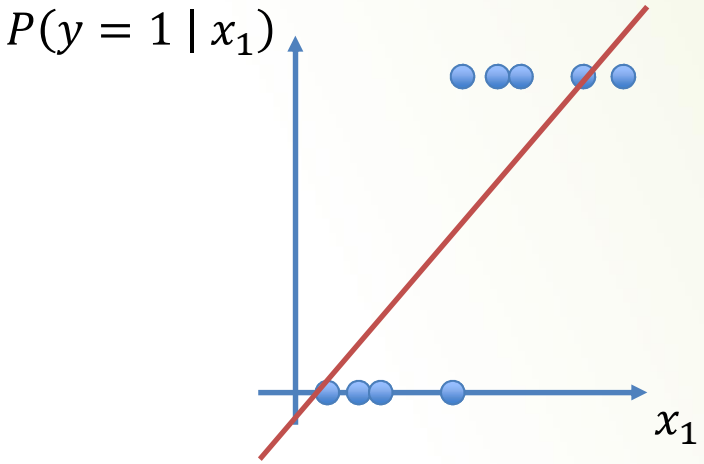
Analyzing Systemic Racial Disparities With Statistical Learning Models and HMIS Data

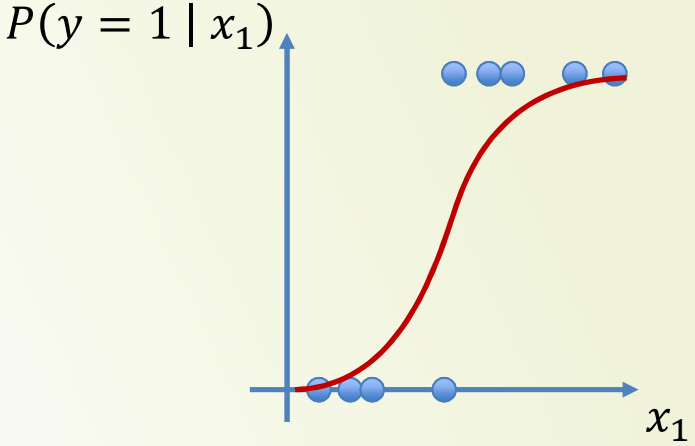
Clayton Aldern



Analyzing Systemic Racial Disparities With Statistical Learning Models and HMIS Data

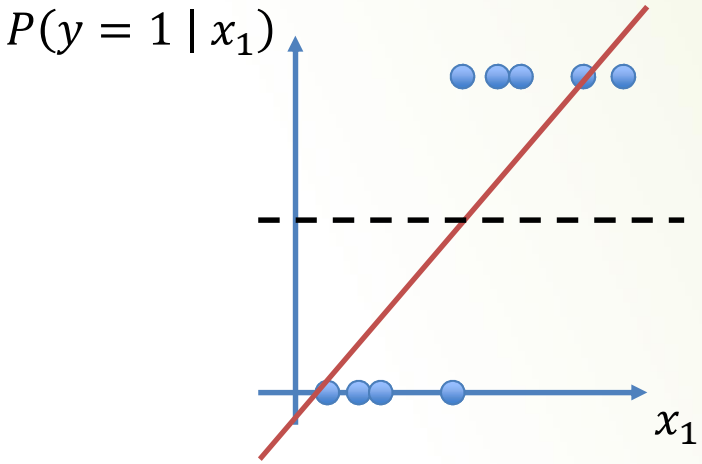
Clayton Aldern

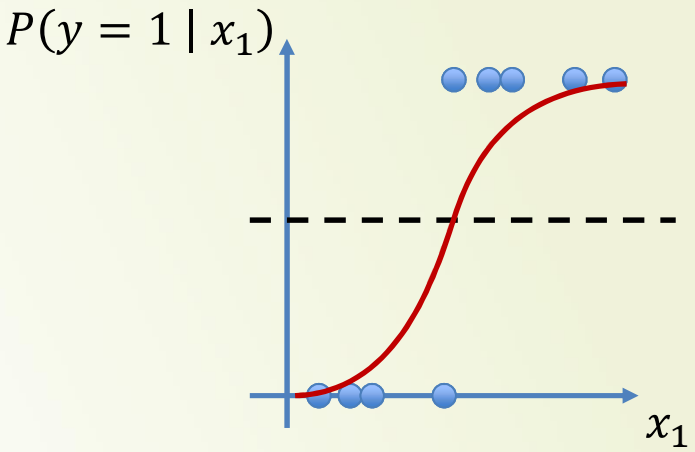




Analyzing Systemic Racial Disparities With Statistical Learning Models and HMIS Data

Clayton Aldern





Analyzing Systemic Racial Disparities
With Statistical Learning Models and HMIS Data
Clayton Aldern

$$P(y = 1 \mid x) = h_\beta(x) = \frac{1}{1 + e^{(-\beta^T X)}}$$

$$\log\left(\frac{P(y = 1 \mid x)}{1 - P(y = 1 \mid x)}\right) = \beta^T X$$

(logistic regression)

Analyzing Systemic Racial Disparities
With Statistical Learning Models and HMIS Data
Clayton Aldern

$$P(y = 1 \mid x) = h_\beta(x) = \frac{1}{1 + e^{(-\beta^T X)}}$$

$$\log\left(\frac{P_{White}(x)}{P_{Black}(x)}\right) = \beta^T X$$

(logistic regression)

Analyzing Systemic Racial Disparities
With Statistical Learning Models and HMIS Data
Clayton Aldern

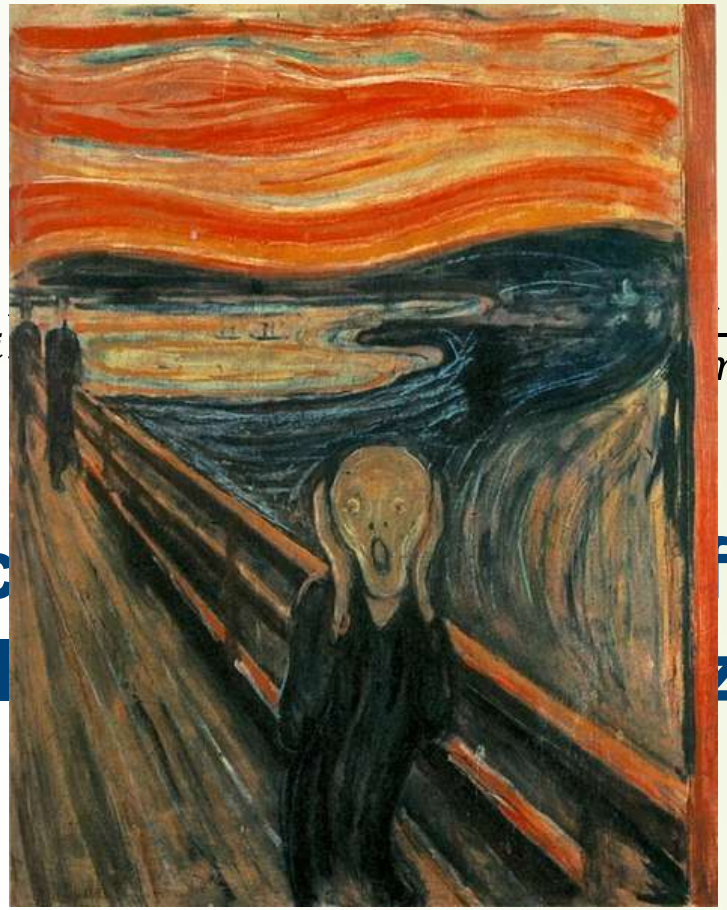
$$\min_{\beta}\left\{-\frac{1}{m}\sum_{i=1}^{m}\left[y_i\log\left(h_\beta(x_i)\right)+(1-y_i)\log\left(1-h_\beta(x_i)\right)\right]+\frac{\lambda}{2m}\sum_{j=1}^{n}\left(\alpha\left|\beta_j\right|+(1-\alpha)\beta_j^2\right)\right\}$$

(logistic regression cost function with elastic net regularization)

Analyzing Systemic Racial Disparities
With Statistical Learning Models and HMIS Data

Clayton Aldern

$$\min_{\beta}\left\{-\frac{1}{m}\sum_{i=1}^{m}\left[y_i \log\left(h_\beta(x_i\right)\right.\right. \qquad \frac{}{n}\sum_{j=1}^{n}\left(\alpha\left|\beta_j\right| + (1-\alpha)\beta_j^2\right)\right\}$$



(logistic function
with el zation)

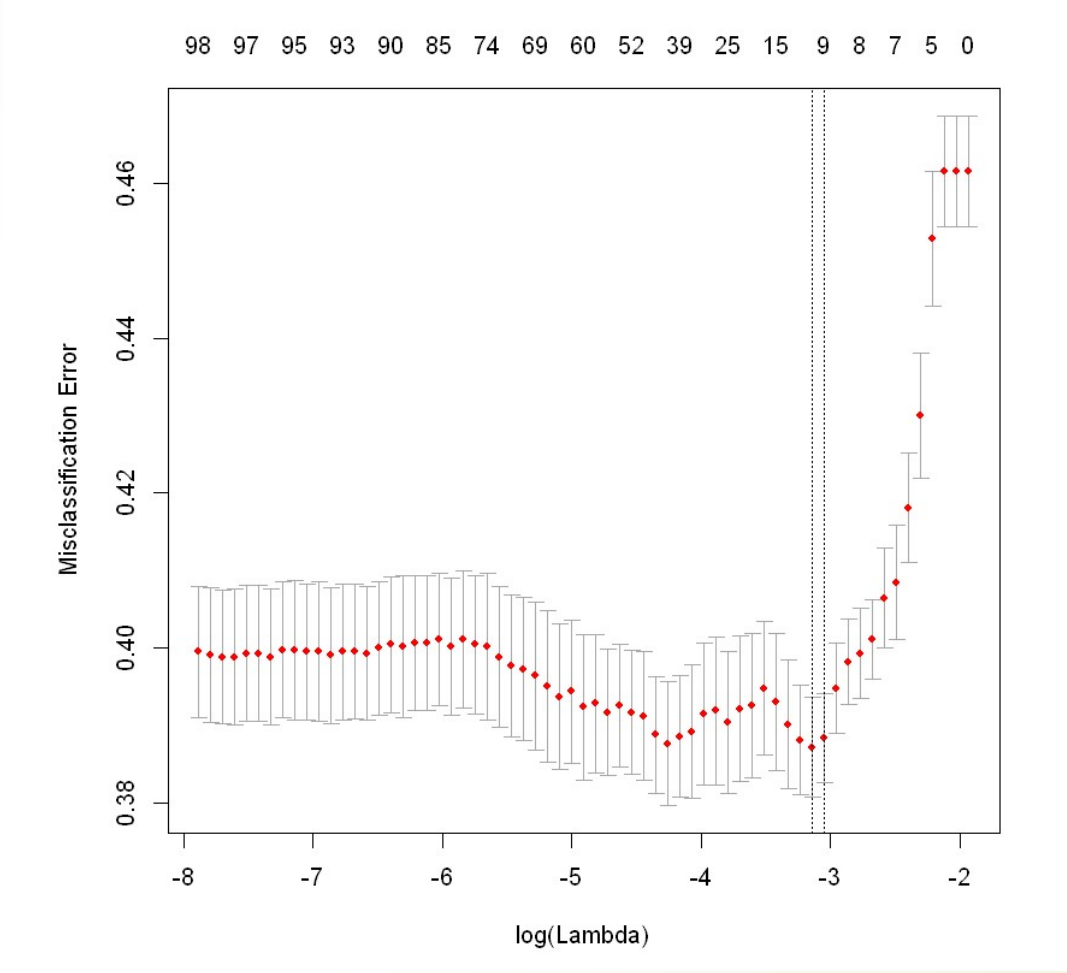
error complexity

$$\min_{\beta}\left\{-\frac{1}{m}\sum_{i=1}^{m}\left[y_i\log\left(h_\beta(x_i)\right)+(1-y_i)\log\left(1-h_\beta(x_i)\right)\right]+\frac{\lambda}{2m}\sum_{j=1}^{n}\left(\alpha\left|\beta_j\right|+(1-\alpha)\beta_j^2\right)\right\}$$

(logistic regression cost function with elastic net regularization)

Analyzing Systemic Racial Disparities With Statistical Learning Models and HMIS Data

Clayton Aldern



Analyzing Systemic Racial Disparities With Statistical Learning Models and HMIS Data

Clayton Aldern

name	*coefficient*	*odds ratio*
(Intercept)	-0.34	0.71
Age	0.01	1.01
Feelings of safety: I don't often feel safe.	-0.02	0.98
Feelings of safety: I usually feel safe.	0.12	1.13
Household AMI <10%	0.15	1.16
Household AMI 30–50%	-0.14	0.87
Disabled: No	-0.10	0.91
Disabled: Yes	0.08	1.09
Medical conditions in household: 0	-0.22	0.80
Medical conditions in household: 4+	0.28	1.32

What is the definition of vulnerability?

From p-values to predictors

- Books
  - *An Introduction to Statistical Learning with Applications in R*
  - free Machine Learning courses online (Ng)
  - *The Elements of Statistical Learning*
  - *Deep Learning* (deeplearningbook.org)
- Computing languages
  - R, Python, Octave, etc.
- Packages
  - R: glmnet; Python: scikit-learn

From p-values to predictors

- Books
  - *An Introduction to Statistical Learning with Applications in R*
  - free Machine Learning courses online (Ng)
  - *The Elements of Statistical Learning*
  - *Deep Learning* (deeplearningbook.org)
- Computing languages
  - R, Python, Octave, etc.
- Packages
  - R: glmnet; Python: scikit-learn

a little
amount of math
a lot

National Human Services Data Consortium

2018 Spring Conference

Pittsburgh, PA

April 18-19, 2018

Analyzing Systemic Racial Disparities With Statistical Learning Models and HMIS Data

Clayton Aldern

Pierce County, Washington | caldern@co.pierce.wa.us

Thank you!

New Frontiers in Data

New Systems, Partners, and Technologies