PART I

An Oversimplified Overview
A Hypothetical Situation

Alice          Bob

Math
Science

Sum: 18     Max: 9
Sum: 17     Max: 10

✓          ✓
A Hypothetical Situation

Alice

Bob
A Hypothetical Situation

Alice

Bob

Caucasian

African American
A Hypothetical Situation

Alice

Bob

Caucasian

African American

✓

✓
A Hypothetical Situation

Alice

Bob

Math

Science

Caucasian

African American
A Hypothetical Situation

Alice

Math 9
Science 9

Bob

Math 10
Science 7

Caucasian

African American
A Hypothetical Situation

Alice

Math: 9
Science: 9
Sum: 18

Bob

Math: 10
Science: 7
Sum: 17
A Hypothetical Situation

Alice

Bob

Math
9
9

Science

9

Sum: 18
Max: 9

Sum: 17
Max: 10
A Hypothetical Situation

Alice

9
9

Math
Science

Sum: 18 ✓
Max: 9

Bob

10
7

Sum: 17
Max: 10 ✓
A Hypothetical Situation

Alice

9
9

Math
Science

Sum: 18 ✓
Max: 9

Bob

Caucasian

9
10

Math
Science

Sum: 17
Max: 10 ✓
A Hypothetical Situation

Alice

Math
9
Science
9

Sum: 18 ✓
Max: 9

Bob

Math
10
Science
7

Sum: 17
Max: 10 ✓

African American
A Hypothetical Situation: Extrapolation

Decision Making System using Sum formula:
Reinforces Bias against African Americans
A Hypothetical Situation: Extrapolation

Decision Making System using Sum formula:

Reinforces Bias against African Americans
A Hypothetical Situation: Extrapolation

Decision Making System using Sum formula:
Reinforces Bias against African Americans
A Hypothetical Situation: Extrapolation

Decision Making System using Sum formula:

Reinforces Bias against African Americans
A Hypothetical Situation: Extrapolation

Decision Making System using **Sum** formula: Reinforces **Bias** against African Americans
PART II
The Real Deal
Applications
Setting: Decision Making

Applications

Admission Committee
Setting: Decision Making

Applications

Historical Data

Admission Committee
But what is Linear Regression!?
But what is Linear Regression!?
But what is Linear Regression!?

Alice

9
9

Math
Science

Bob

10
7

3 \times \text{Math} + \text{Science}

Linear Regression
Learn Best **Weighted Sum** Formula
Data

- National Education Longitudinal Study of 1988
Data

- National Education Longitudinal Study of 1988
- Students entering 8th grade
Data

- National Education Longitudinal Study of 1988
- Students entering 8th grade
- Successful student: College GPA $> 3.25$
- National Education Longitudinal Study of 1988
- Students entering 8th grade
- Successful student: College GPA > 3.25
- Caucasian: 4173 (48.62% successful)
  African American: 442 (30.88% successful)
  Hispanics: 506 (39.14% successful)
● National Education Longitudinal Study of 1988
● Students entering 8th grade
● Successful student: College GPA > 3.25
● Caucasian: 4173 (48.62% successful)
  African American: 442 (30.88% successful)
  Hispanics: 506 (39.14% successful)

Affirmative Action is practiced in College Admission in the US since 1970s
Fitting to a Line

→ 9th grade Math vs. College success
Fitting to a Line

→ 9th grade Math vs. College success

→ "You are good at math
   You will do well"
Fitting to a Line

9th grade Math vs. College success

"You are good at math. You will do well"
Fitting to a Line

9th grade Math vs. College success

"You are good at math You will do well"
Fitting to a Line

9th grade Math vs. College success

"You are good at math. You will do well”

What happens if we admit top 10% students?
Fitting to a Line

→ 9th grade Math vs. College success

→ "You are good at math. You will do well"

→ What happens if we admit top 10% students?

→ Error is not Uniform Across Race groups
Number of Influential Features

→ 9\textsuperscript{th} and 11\textsuperscript{th} grades, Standardized tests, Extra Curriculars, ...
Number of Influential Features

- 9th and 11th grades,
- Standardized tests,
- Extra Curriculars, ...
Number of Influential Features

9th and 11th grades, Standardized tests, Extra Curriculars, ...
Number of Influential Features

→ 9th and 11th grades, Standardized tests, Extra Curriculars, ...

→ Caucasians have few influential features
Number of Influential Features

9 th and 11 th grades, Standardized tests, Extra Curriculars, ...

Caucasians have few influential features

The regressor formula of minority groups is complex
$\ell_p$-norm Rankings

$(x, y)$ is a point in plane

$\text{th norm of } (x, y) \text{ is }$

$(|x|^p + |y|^p)^{1/p}$

→ African Americans prefer $\infty$-norm over $1$-norm

→ Hispanics "prefer" $0$-norm
$l_p$-norm Rankings

→ $(x, y)$ is a point in plane
   $p^{th}$ norm of $(x, y)$ is
   $(|x|^p + |y|^p)^{1/p}$

→ 1-norm is $|x| + |y|$
   2-norm is standard distance
   $\infty$-norm is $\max(|x|, |y|)$
\( \ell_p \)-norm Rankings

\[ (x, y) \text{ is a point in plane} \]
\[ p^{th} \text{ norm of } (x, y) \text{ is} \]
\[ (|x|^p + |y|^p)^{1/p} \]

\[ \rightarrow \]
\[ 1\text{-norm is } |x| + |y| \]
\[ 2\text{-norm is standard distance} \]
\[ \infty\text{-norm is max}(|x|, |y|) \]
$\ell_p$-norm Rankings

→ $(x, y)$ is a point in plane

$p^{th}$ norm of $(x, y)$ is

\[ (|x|^p + |y|^p)^{1/p} \]

→ 1-norm is $|x| + |y|$

2-norm is standard distance

$\infty$-norm is $\max(|x|, |y|)$
\( \ell_p \)-norm Rankings

\( (x, y) \) is a point in plane

\( p^{th} \) norm of \((x, y)\) is

\[ (|x|^p + |y|^p)^{1/p} \]

\( 1 \)-norm is \(|x| + |y|\)

2-norm is standard distance

\( \infty \)-norm is \( \max(|x|, |y|) \)

→ African Americans prefer

\( \infty \)-norm over 1-norm
\( \ell_p \)-norm Rankings

\( (x, y) \) is a point in plane

\( p^{th} \) norm of \((x, y)\) is

\[ (|x|^p + |y|^p)^{1/p} \]

\[ \rightarrow \]

1-norm is \(|x| + |y|\)

2-norm is standard distance

\( \infty \)-norm is \( \max(|x|, |y|) \)

\[ \rightarrow \]

African Americans prefer \( \infty \)-norm over 1-norm

\[ \rightarrow \]

Hispanics "prefer" 1-norm
Polynomial Regression

→ From Geometric Insights to Algebraic Tools
Polynomial Regression

→ From Geometric Insights to Algebraic Tools

→ Training/Test: 50/50
Polynomial Regression

→ From Geometric Insights to Algebraic Tools

→ Training/Test: 50/50
Polynomial Regression

→ From **Geometric Insights** to **Algebraic Tools**

→ **Training/Test : 50/50**
Polynomial Regression

→ From Geometric Insights to Algebraic Tools

→ Training/Test: 50/50

→ Poly Regression captures more complex relationships
Polynomial Regression

→ From Geometric Insights to Algebraic Tools

→ Training/Test: 50/50

→ Poly Regression captures more complex relationships

→ A proof of concept
   Not a solution
Each minority group must have its own predictor
(formalized in Kleinberg-Ludwig-Mullainathan-Rambachan ’18)
Each minority group must have its own predictor
(formalized in Kleinberg-Ludwig-Mullainathan-Rambachan ’18)

Predictor must be sensitive to the geometry of its group
Each minority group must have its own predictor
(formalized in Kleinberg-Ludwig-Mullainathan-Rambachan ’18)

Predictor must be sensitive to the geometry of its group

Linear Regression cannot capture
the complexity of minority group data
THANK YOU!