Machine Learning:
A Primer

Achim Koh, MA in Liberal Studies, The Graduate Center

akoh@gradcenter.cuny.edu

@achimkoh

Machine Learning

Making computers learn from data

"Reverse engineering" data
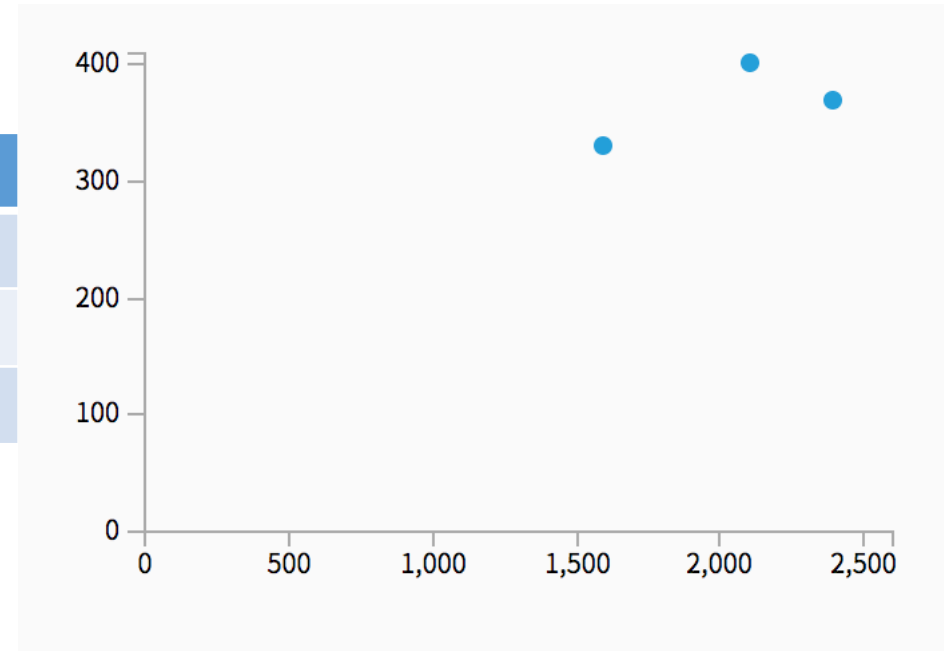
Goals of ML

- pattern detection
- prediction
- decision making

Learning from data?

- The house example

Area (sq ft) (x)	Price (y)
2,104	399,900
1,600	329,900
2,400	369,000

- A line that is at minimum aggregate distance from all data points
- https://jalammar.github.io/visual-interactive-guide-basics-neural-networks/

@achimkoh

Some Lingo

- Vectors
- Dimensions / Features
- Optimization
- Training
- Overfitting
- Accuracy, precision, recall
- Errors: false positive / true negative

@achimkoh

Confusion Matrix

		Prediction	
		Yes	No
Actual truth	Yes	TP	False Negative (error)
	No	False Positive (error)	TN

@achimkoh

Supervised Learning

- Classification

Is price y for house x a 'good' price or not?
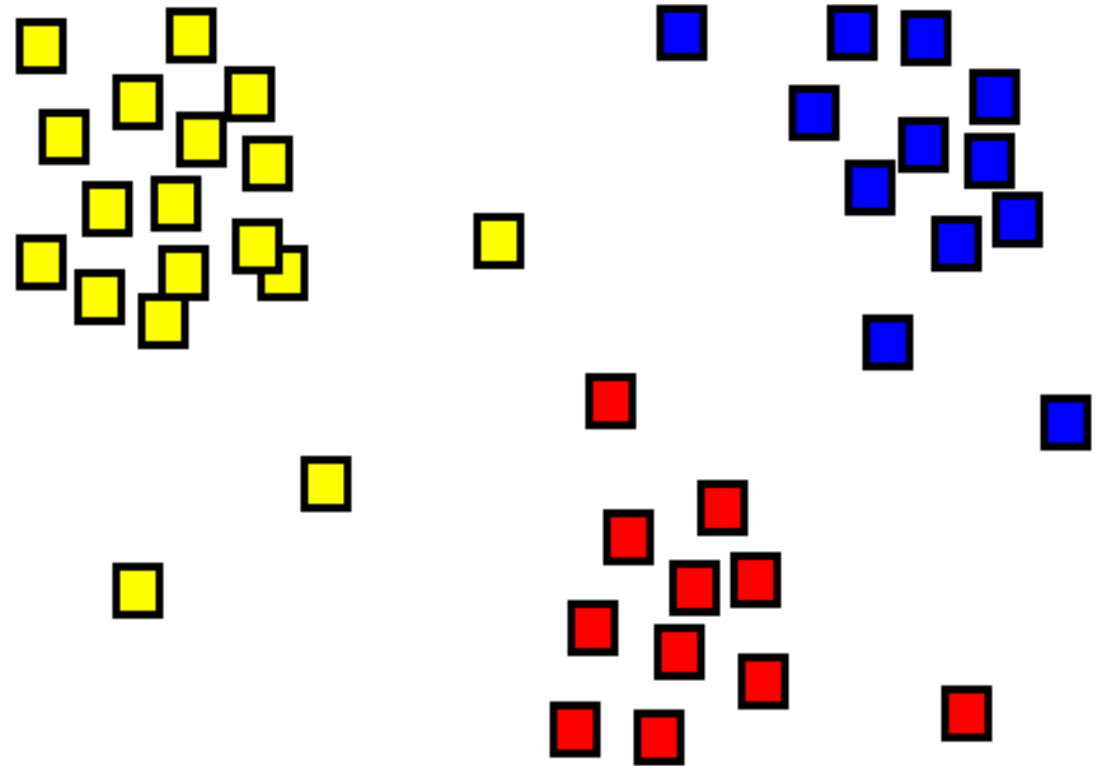
- Regression

Based on data, what would be the predicted price for a house not in dataset?
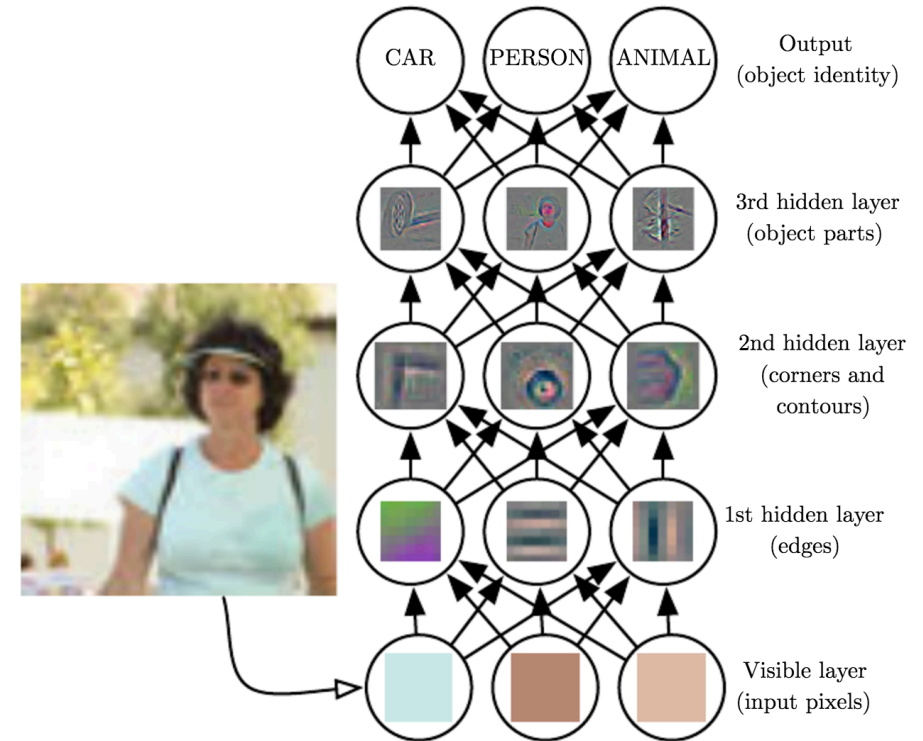
Unsupervised Learning

- Clustering

Which objects are similar?

That's Not All

- Neural Networks / Deep Learning

- Reinforcement Learning

- …
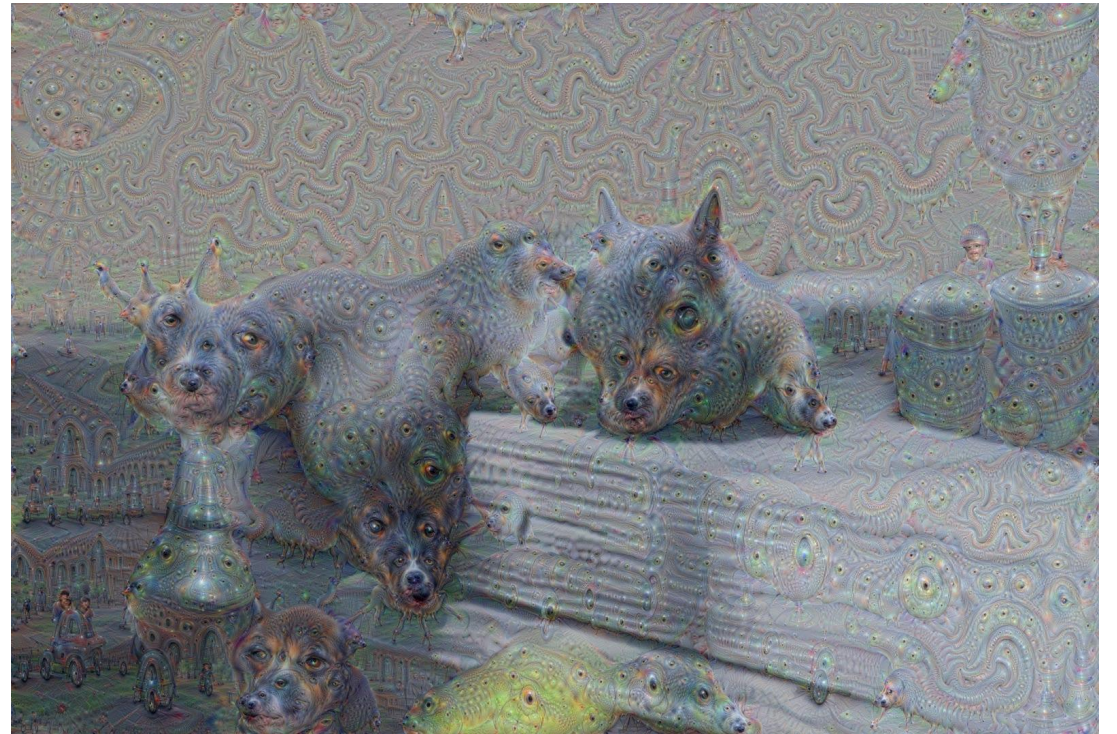

http://www.deeplearningbook.org/

@achimkoh

Applications of Machine Learning

- natural language processing, speech recognition

- computer vision

- dynamic pricing

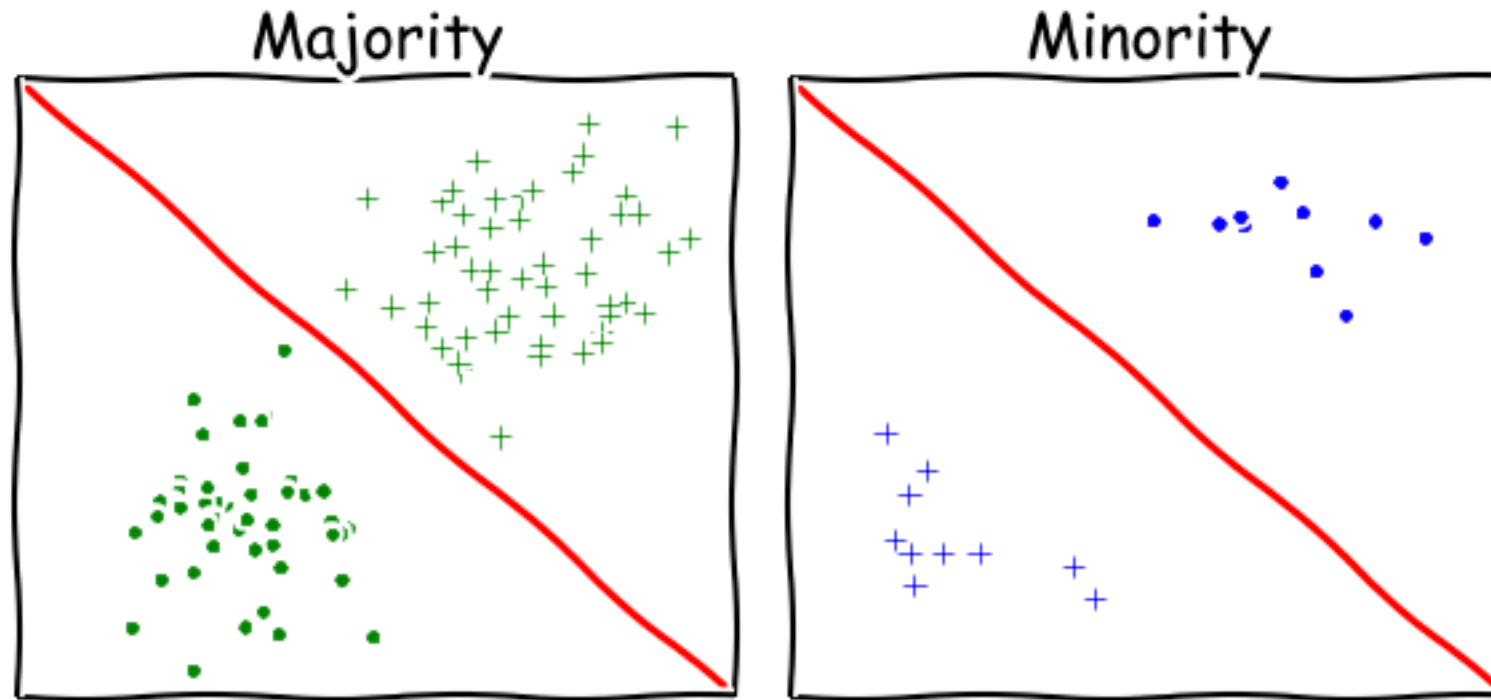- recommendation, content curation

- robotics, self-driving cars

- art https://aiexperiments.withgoogle.com/

Paul Klee, *Strange Garden*

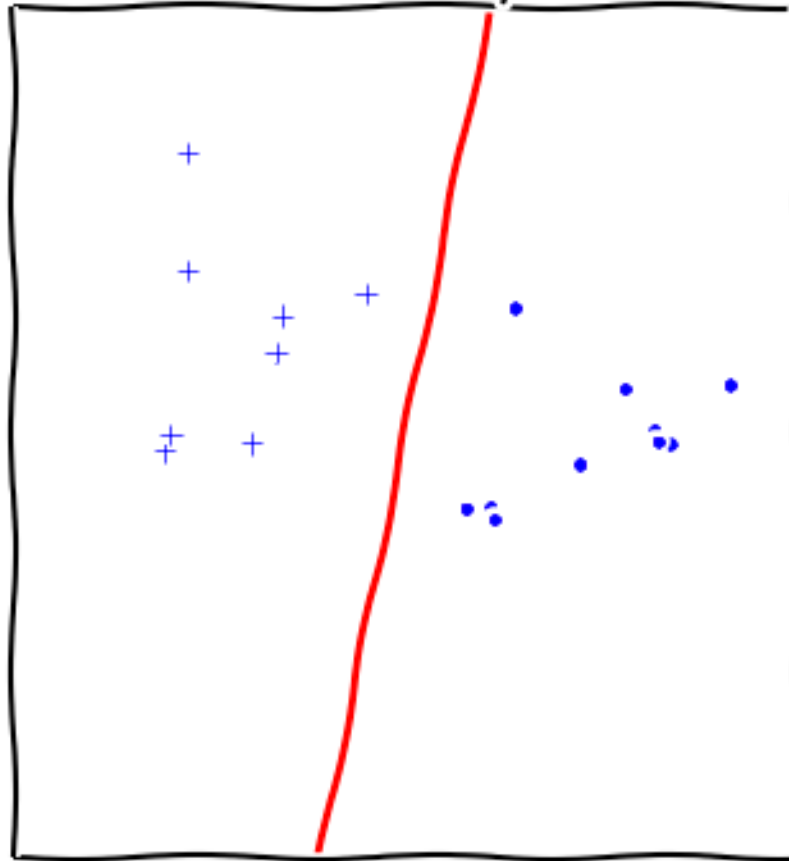**Image generated by a Convolutional Network**

Heterogenic data

Majority

Minority

Moritz Hardt

Complex data
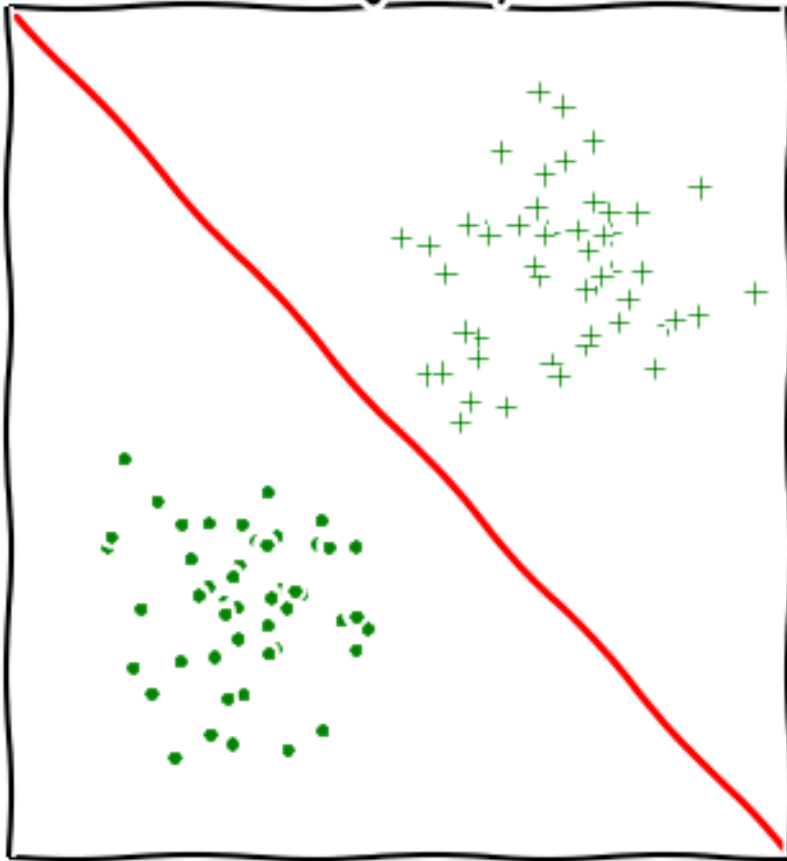
Majority Minority Population :-(

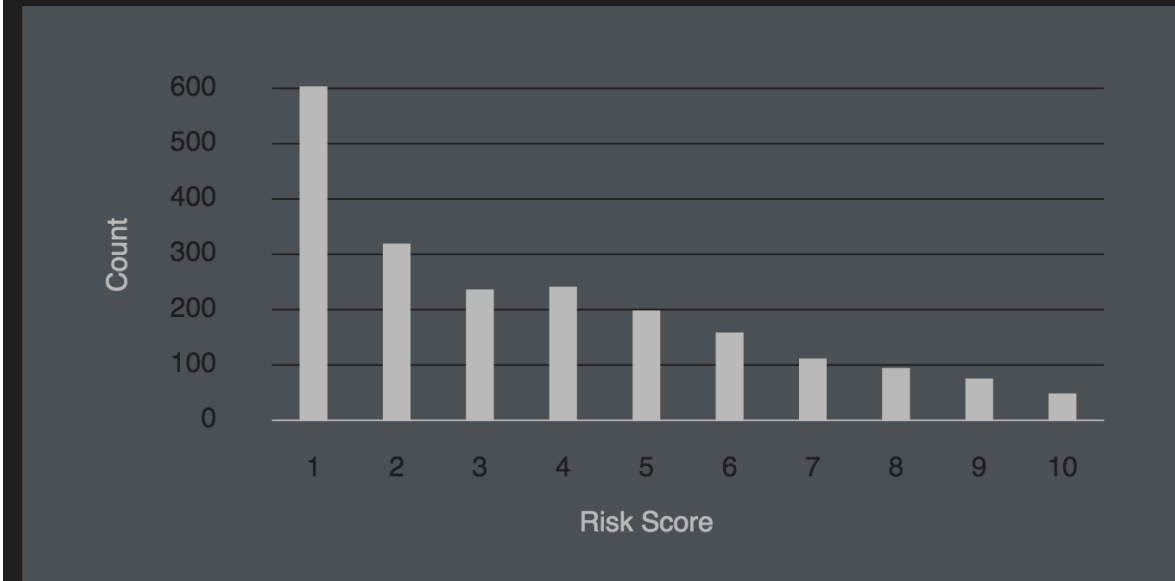Moritz Hardt
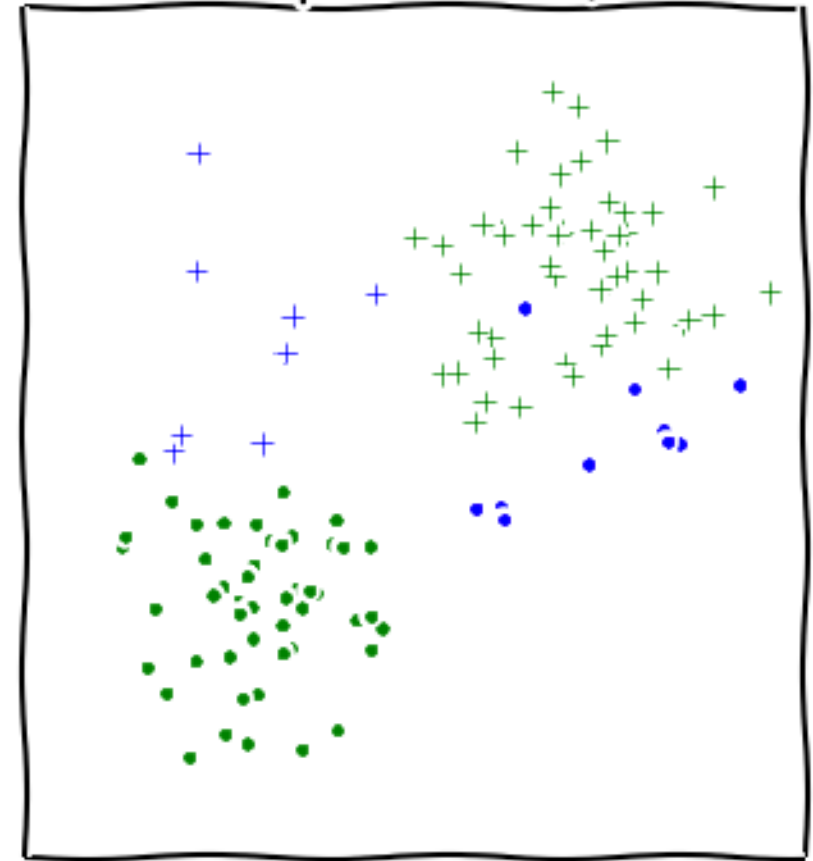
Machine Bias

Black Defendants' Risk Scores
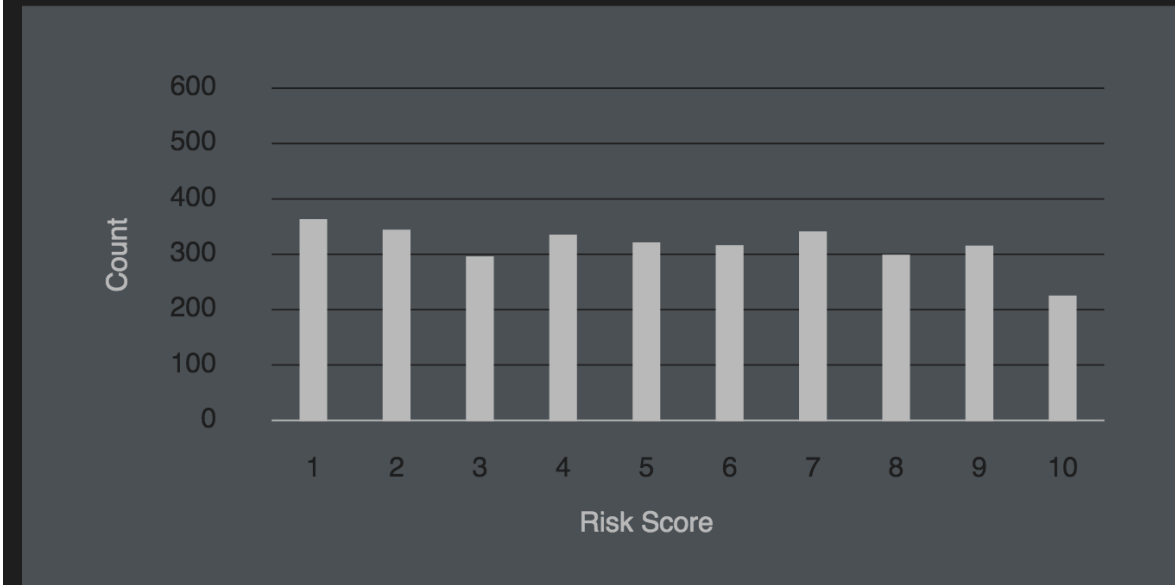
White Defendants' Risk Scores

These charts show that scores for white defendants were skewed toward lower-risk categories. Scores for black defendants were not. *(Source: ProPublica analysis of data from Broward County, Fla.)*

Machine Bias

Prediction Fails Differently for Black Defendants

	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

Overall, Northpointe's assessment tool correctly predicts recidivism 61 percent of the time. But blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend. It makes the opposite mistake among whites: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes. (Source: ProPublica analysis of data from Broward County, Fla.)

https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

@achimkoh

Confusion Matrix

		Prediction	
		Yes	No
Actual truth	Yes	TP	False Negative (error)
	No	False Positive (error)	TN

@achimkoh

Absence / Discrepancies in Data

- Missing Datasets (Onuoha)

https://github.com/MimiOnuoha/missing-datasets

Implications & Challenges

- Opacity (Burrell, 2016)
  - intentional corporate or state secrecy
  - technical illiteracy
  - a result of the characteristics of ML algorithms themselves

@achimkoh

Resources

Technical

- Lists of resources
  - https://github.com/ZuzooVn/machine-learning-for-software-engineers
  - https://medium.com/machine-learnings/a-humans-guide-to-machine-learning-e179f43b67a0#.55o46t3tv
- Specific field-oriented
  - The Programming Historian
  - ML4A
  - Machine Learning for Musicians and Artists

Nice primer reads

- The Great A.I. Awakening http://nyti.ms/2hE6XZ5
- Soon We Won't Program Computers. We'll Train Them Like Dogs http://www.wired.com/2016/05/the-end-of-code/

@achimkoh

Visually Rich Introductions

- [Machine Learning Is Fun!](#)
- r2d3 http://www.r2d3.us/visual-intro-to-machine-learning-part-1/
- Facebook https://code.facebook.com/pages/1902086376686983

@achimkoh

Critical discourse

- Scholarly Journals

- Journalism

- The Social Media Collective (SMC)'s Critical Algorithm Studies: a Reading List

- Industry

- Data&Society's links page

- Conferences & other events

@achimkoh

Thank you!