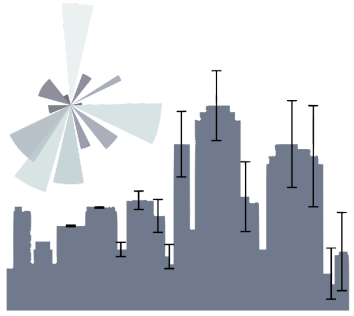
principles of Urban Science 7



machine learning | data ethics

dr.federica bianco | *fbb.space* | *fedhere* | *fedhere*

this slide deck: https://slides.com/federicabianco/pus2020_7

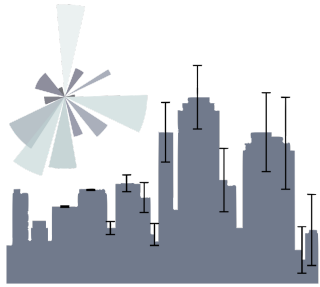
what is machine learning

machine learning best practices

issues in data ethics

epistemic transparency

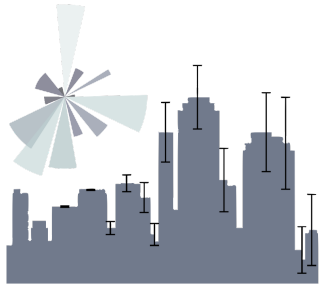
where does the bias enter models



1

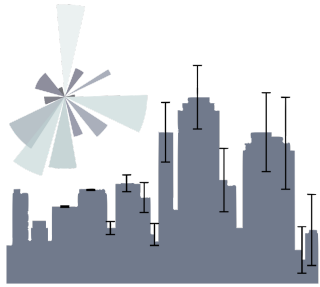
what is machine learning?

what is a model?



the best way to think about it *in the ML context:*

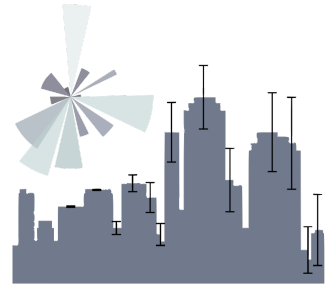
a model is a low
dimensional representation
of a higher dimensionality
datase



what is machine learning?

[Machine Learning is the] field of study that gives computers the ability to learn without being explicitly programmed.

Arthur Samuel, 1959



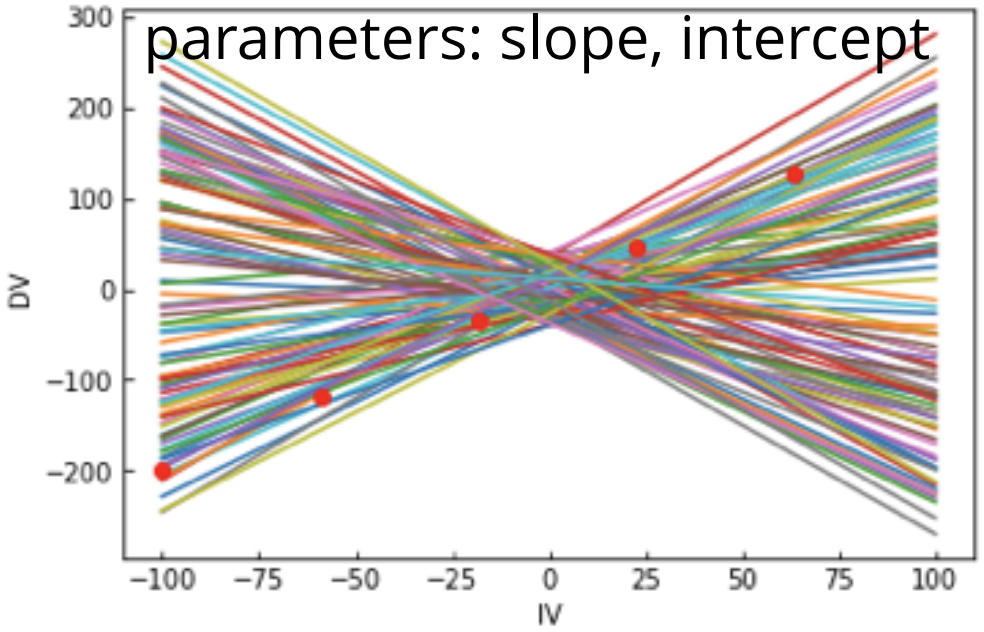
what is machine learning?

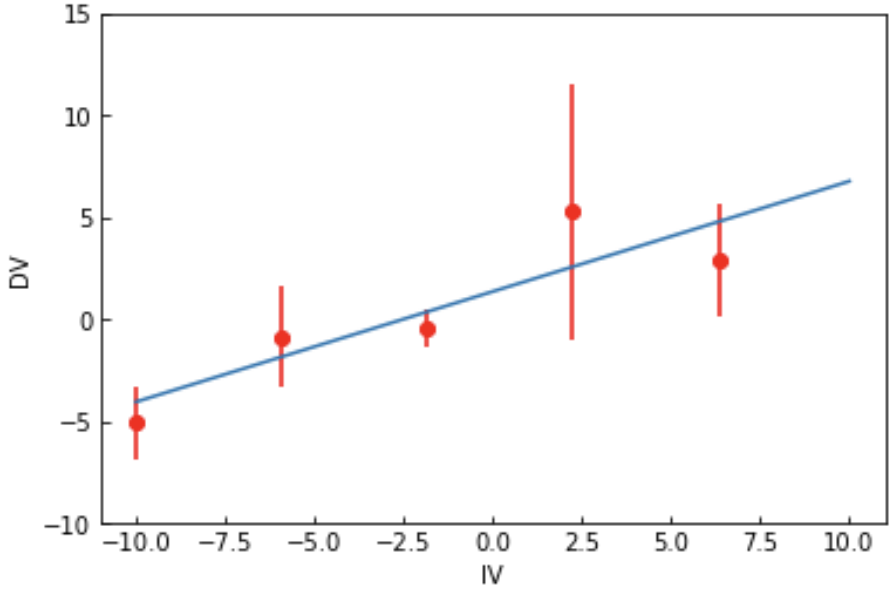
[Machine Learning is the] field of study that gives computers the ability to learn without being explicitly programmed.

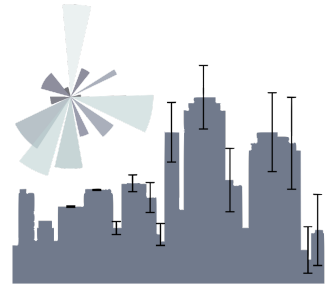
Arthur Samuel, 1959

model

parameters: slope, intercept







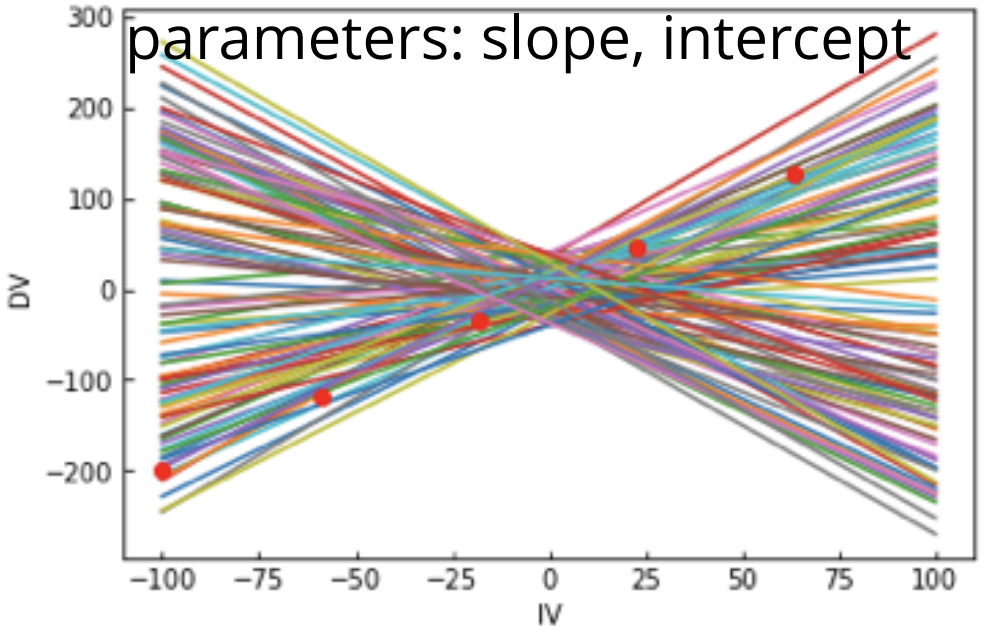
what is machine learning?

[Machine Learning is the] field of study that gives computers the ability to learn without being explicitly programmed.

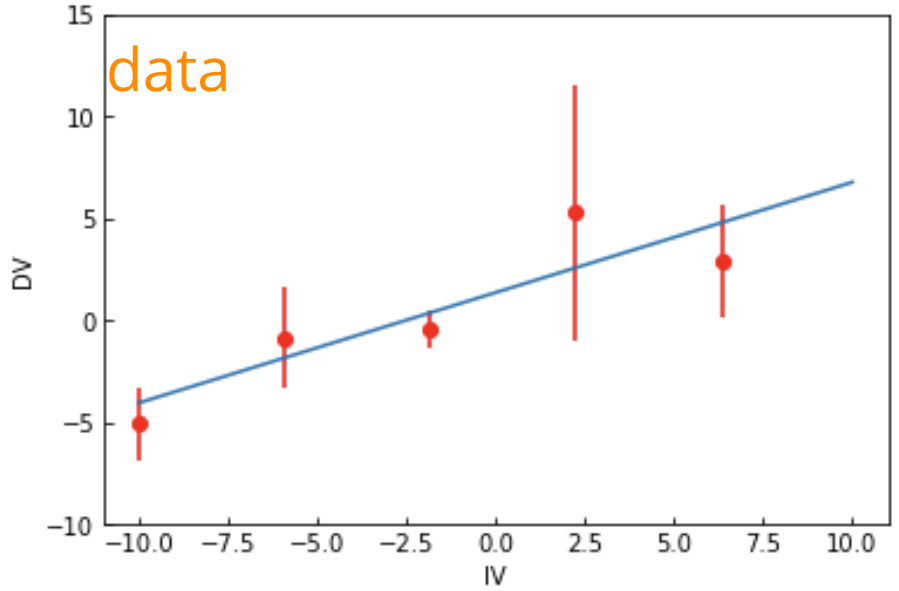
Arthur Samuel, 1959

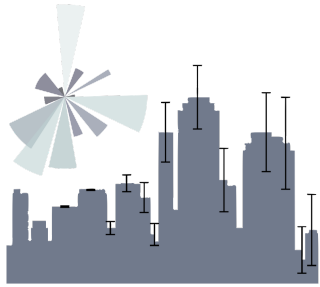
model

parameters: slope, intercept



data



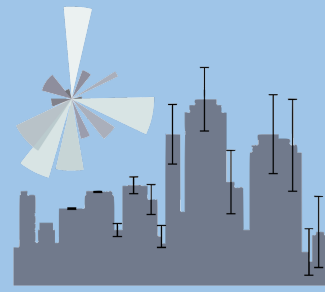


what is machine learning?

ML: any model with parameters learnt from the data

https://miro.medium.com/max/960/1*imhEKEpzX24CC_LIIureBw.gif

what is machine learning?



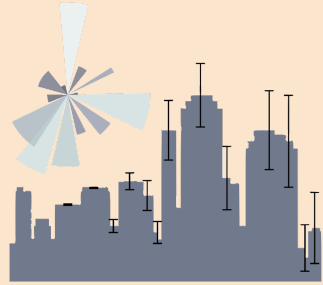
Machine Learning models are parametrized representation of "reality" where the parameters are learned from finite sets of realizations of that reality

(note: learning by instance, e.g. nearest neighbours, may not comply to this definition)

Machine Learning is the disciplines that conceptualizes, studies, and applies those models.

Key Concept

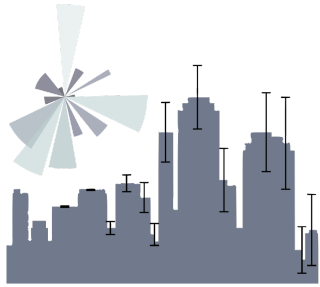
General ML points

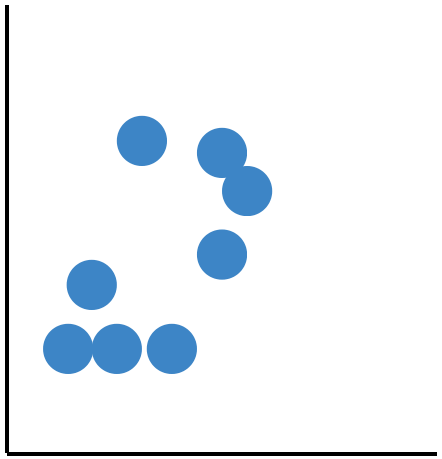


used to:

- understand structure of feature space
- classify based on examples,
- regression (classification with infinitely small classes)
- understand which features are important in prediction (to get close to causality)

unsupervised vs supervised learning

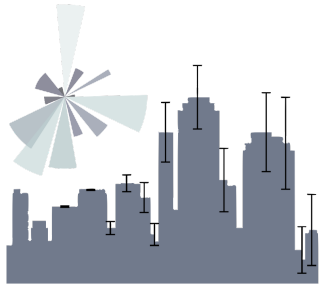




Clustering

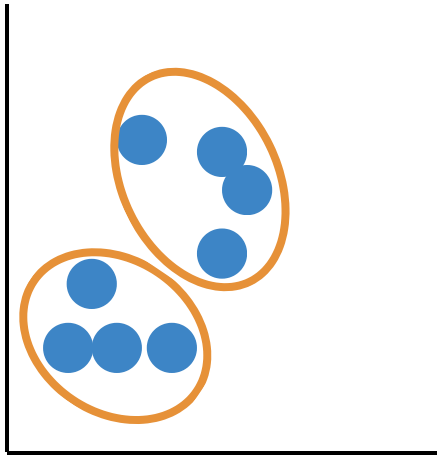
partitioning the feature space so that the existing data is grouped (according to some target function!)

unsupervised vs supervised learning



Unsupervised learning

- understanding structure
- anomaly detection
- dimensionality reduction

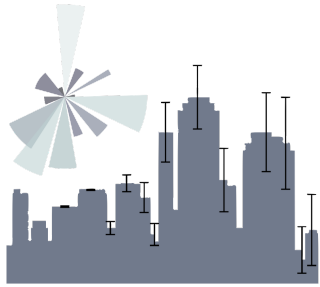


Clustering

partitioning the feature space so that the existing data is grouped (according to some target function!)

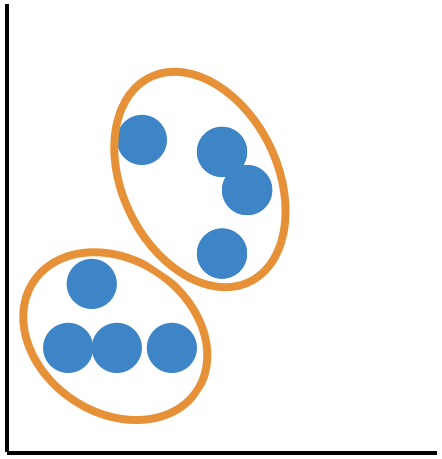
All features are observed for all datapoints

unsupervised vs supervised learning



Unsupervised learning

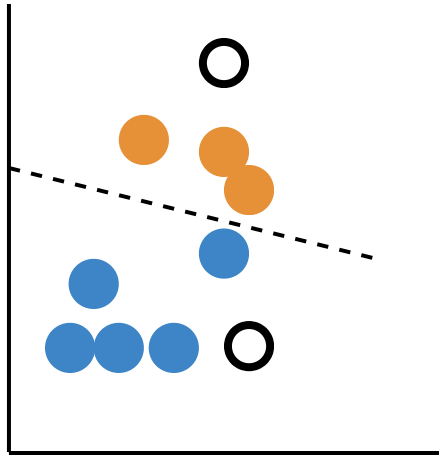
- understanding structure
- anomaly detection
- dimensionality reduction



Clustering

partitioning the feature space so that the existing data is grouped (according to some target function!)

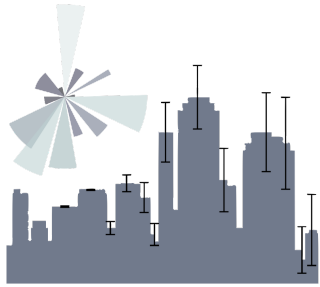
All features are observed for all datapoints



Classifying & regression

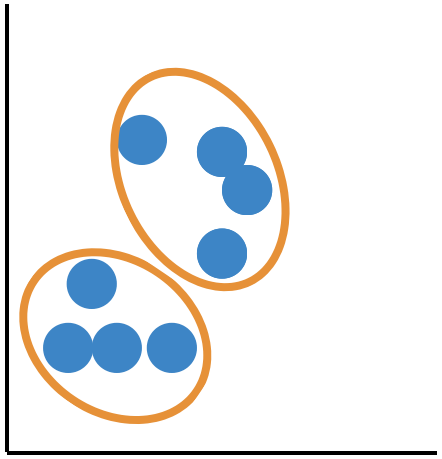
finding functions of the variables that allow to predict unobserved properties of new observations

unsupervised vs supervised learning



Unsupervised learning

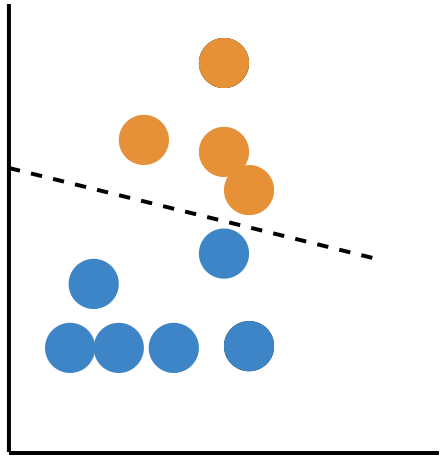
- understanding structure
- anomaly detection
- dimensionality reduction



Clustering

partitioning the feature space so that the existing data is grouped (according to some target function!)

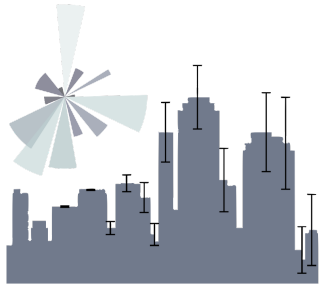
All features are observed for all datapoints



Classifying & regression

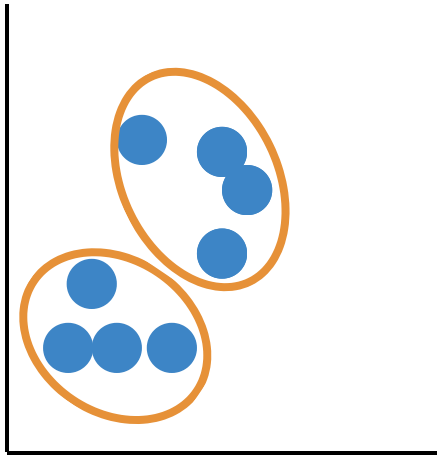
finding functions of the variables that allow to predict unobserved properties of new observations

unsupervised vs supervised learning



Unsupervised learning

- understanding structure
- anomaly detection
- dimensionality reduction



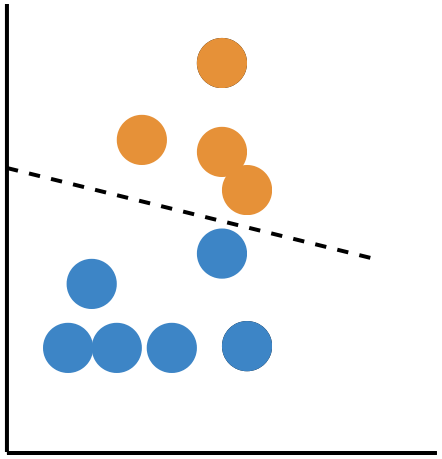
Clustering

partitioning the feature space so that the existing data is grouped (according to some target function!)

All features are observed for all datapoints

Supervised learning

- classification
- prediction
- feature selection

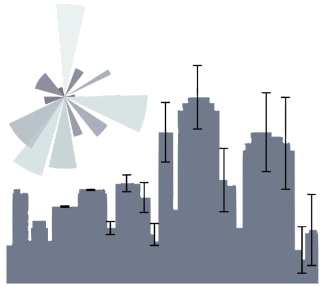


Classifying & regression

finding functions of the variables that allow to predict unobserved properties of new observations

Some features are not observed for some data points we want to predict them.

unsupervised vs supervised learning



Unsupervised learning

All features are observed for all datapoints

and we are looking for structure in the feature space

Supervised learning

Some features are not observed for some data points we want to predict them.

The datapoints for which the target feature is observed are said to be "*labeled*"

also...

Semi-supervised learning

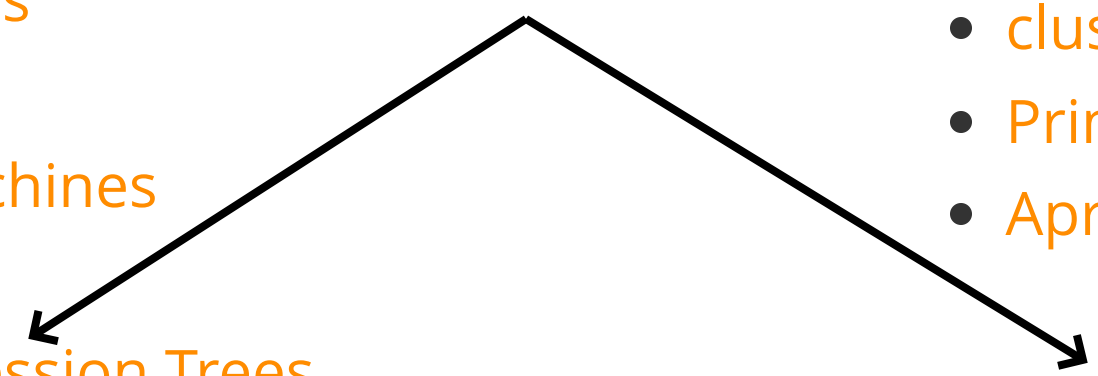
A small amount of labeled data is available. Data is cluster and clusters inherit labels

Active learning

The code can interact with the user to update labels.

what is machine learning?

- k-Nearest Neighbors
- Regression
- Support Vector Machines
- Neural networks
- Classification/Regression Trees



- clustering
- Principle Component Analsysis
- Apriori (association rule)

supervised learning

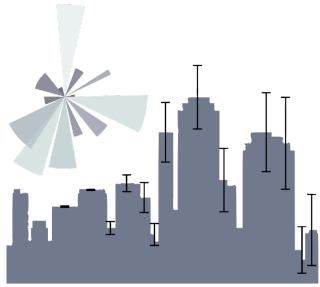
extract features and create models that allow prediction where the correct answer is known for a subset of the data

unsupervised learning

identify features and create models that allow to understand structure in the data

train, test, and validate

MLtsa: validating a model



How do we measure if a model is good?

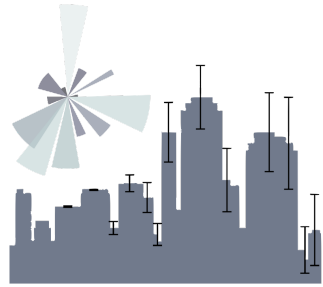
- Accuracy
- Precision
- Recall
- ROC
- AOC

We will talk more about this later...

but for now focus on

regression performance metrics

MLtsa: validating a model



How do we measure if a model is good? $\epsilon_i = y_i - f(t_i)$

- Accuracy
- Precision
- Recall
- ROC
- AOC

We will talk more about this later...

but for now focus on

regression performance metrics

$$AE = \sum_i |\epsilon_i|$$ **Absolute error**

$$SE = \sum_i \epsilon_i^2$$ **Squared error**

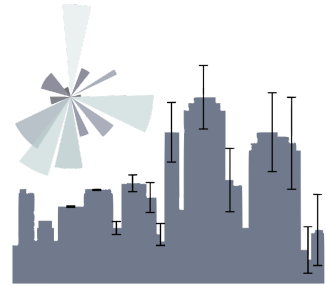
$$MSE = \frac{1}{N} SE$$ **Mean squared error**

$$RMSE = \sqrt{MSE}$$ **Root mean squared error**

$$rMSE = \frac{MSE}{\sigma^2}$$ **Relative mean squared error**

$$R^2 = 1 - rMSE$$ **R squared**

MLtsa: validating a model



do you recognize these??

How do we measure if a model is good? $\epsilon_i = y_i - f(t_i)$

Accuracy

Precision

Recall

ROC

AOC

$$AE = \sum_i |\epsilon_i|$$ **Absolute error**

$$SE = \sum_i \epsilon_i^2$$ **Squared error**

$$MSE = \frac{1}{N} SE$$ **Mean squared error**

$$RMSE = \sqrt{MSE}$$ **Root mean squared error**

$$rMSE = \frac{MSE}{\sigma^2}$$ **Relative mean squared error**

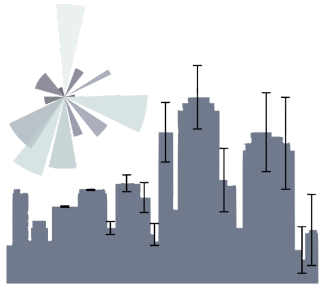
$$R^2 = 1 - rMSE$$ **R squared**

We will talk more about this later...

but for now focus on

regression performance metrics

MLtsa: validating a model



How do we measure if a model is good? $\epsilon_i = y_i - f(t_i)$

- Accuracy
- Precision
- Recall
- ROC
- AOC

We will talk more about this later...

but for now focus on

regression performance metrics

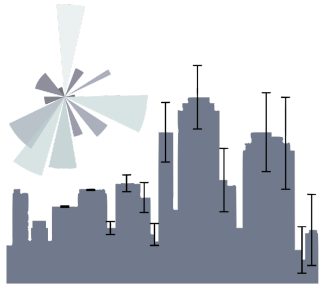
$$R^2 = 1 - rMSE$$

Split the sample in test and training sets

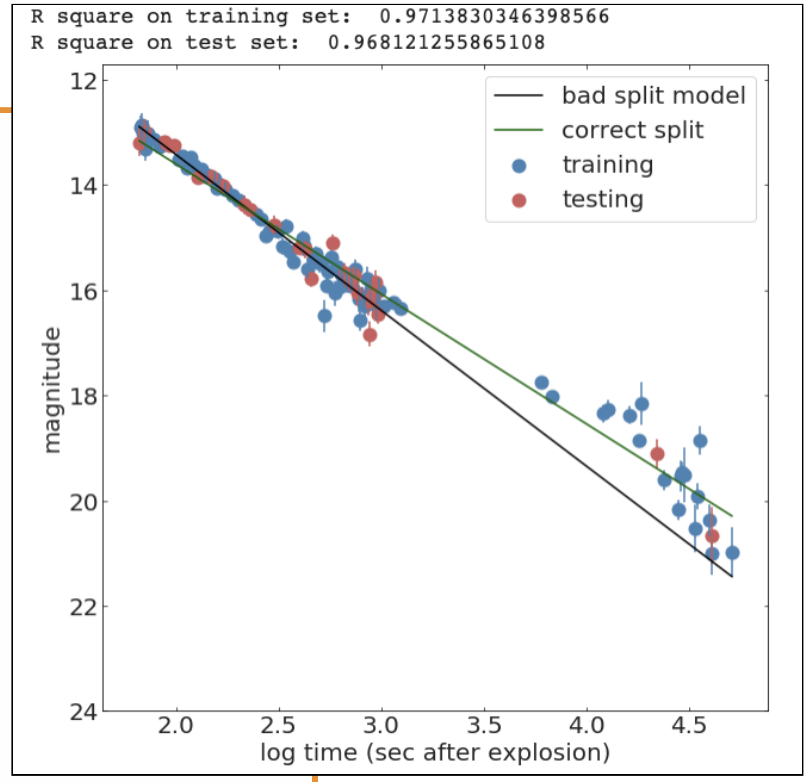
Train on the training set

Test (measure accuracy) on the test set

MLtsa: validating a model



```
from sklearn.model_selection import train_test_split

def line(x, intercept, slope):
    return slope * x + intercept

def chi2(args, x, y, s):
    a, b = args
    return sum((y - line(x, a, b))**2 / s)

x_train, x_test, y_train, y_test, s_train, s_test = train_test_split(
     x, y, s, test_size=0.25, random_state=42)

initialGuess = (10, 1)

chi2Solution_goodsplit = minimize(chi2, initialGuess,
        args=(x_train, y_train, s_train))

print("best fit parameters from the minimization of the chi squared: " +
       "slope {:.2f}, intercept {:.2f}".format(*chi2Solution_goodsplit.x))

print("R square on training set: ", Rsquare(chi2Solution_goodsplit.x, x_train, y_train))
print("R square on test set: ", Rsquare(chi2Solution_goodsplit.x, x_test, y_test))
```



In ML models need to be "validated":

1. split the data into a training and a test set (typical split 70/30).
2. learn the model parameters by "training" the model on the training set
3. "test" the model on the test set: measure the accuracy of the prediction (e.g. as the distance between the prediction and the test data).

The performance on the model is the performance achieved on the test set.

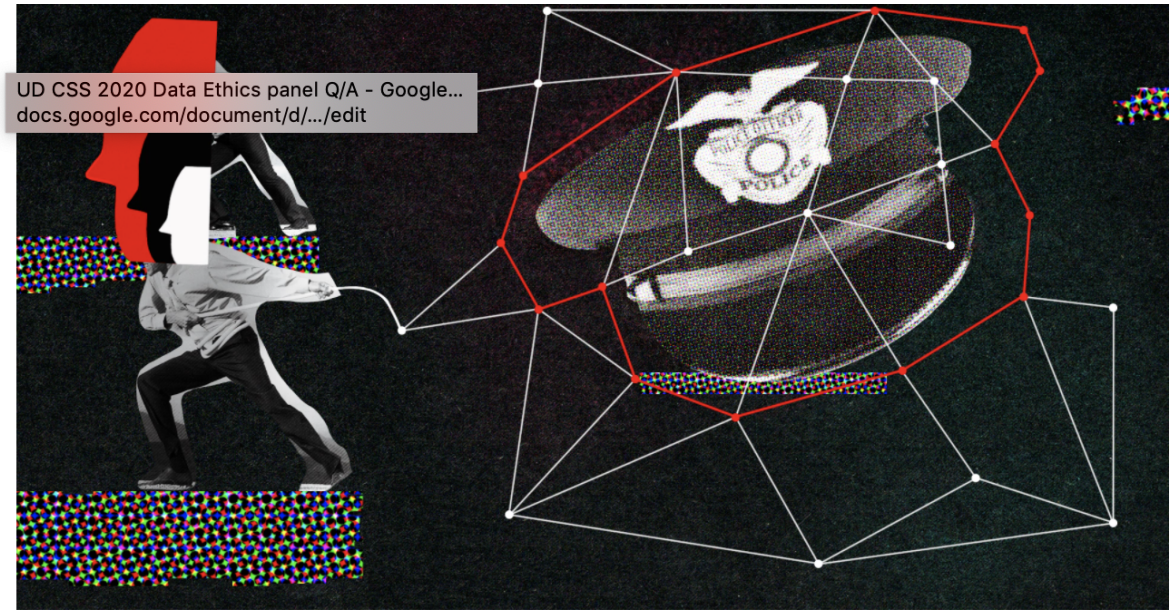
a significance performance degradation on the test compared to training set indicates that the model is "overtrained" and does not generalize well.

An upgrade on this workflow is to create a training, a test, and a validation test. Iterate between training and test to achieve optimal performance, then measure accuracy on the validation set.This is because you can use the test set performance to tune the model hyperparameters (model selection) but then you would report a performance that is tuned on the test set.

data ethics

intended and unintended

Pixabay



FRANZISKA BARCZYK

The Ideology of Racism: Misusing Science to Justify Racial Discrimination

https://www.un.org/en/chronicle/article/ideology-racism-misusing-science-justify-racial-discrimination

Artificial intelligence

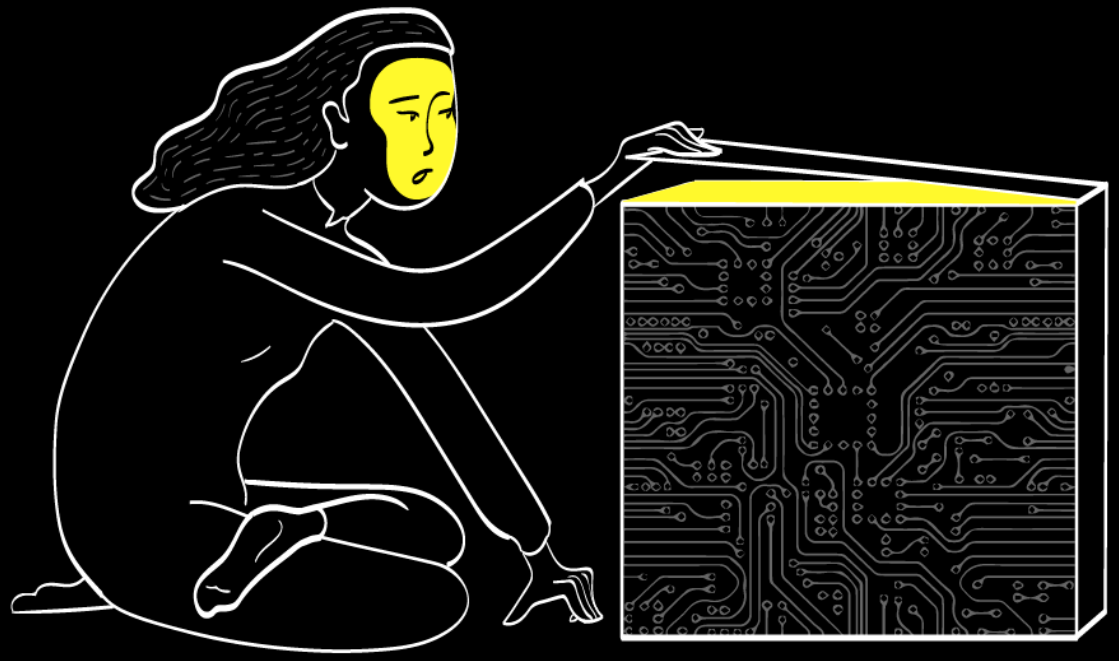
Predictive policing algorithms are racist. They need to be dismantled.

Lack of transparency and biased training data mean these tools are not fit for purpose. If we can't fix them, we should ditch them.

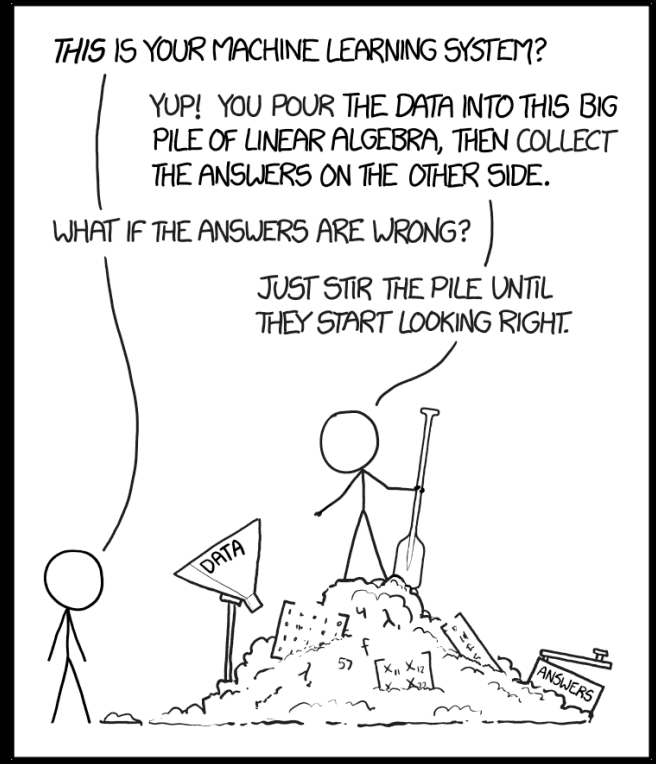
https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/

two dangerous data-ethics myths

Data Science is a black box



Models are neutral, data is biased

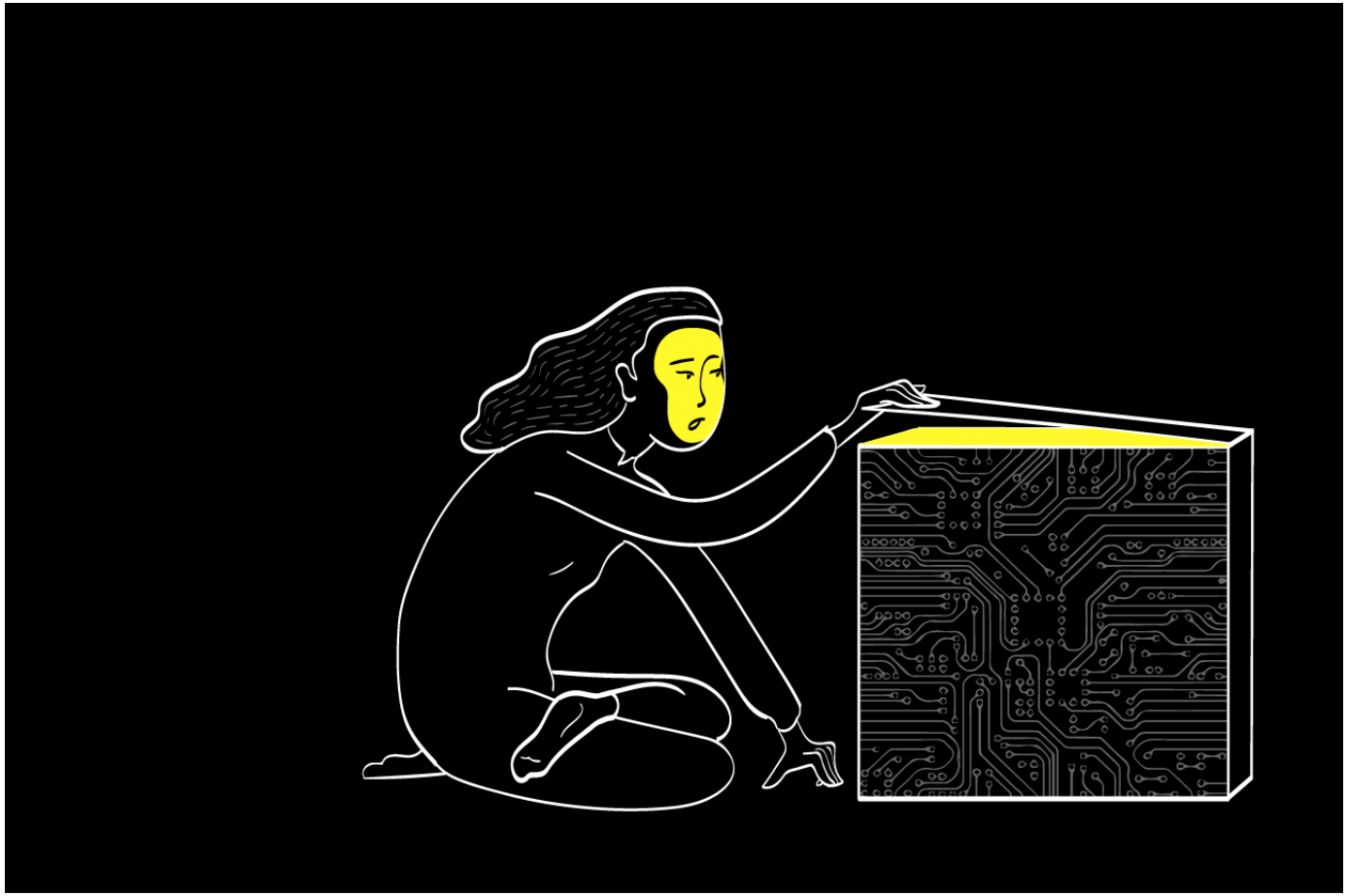


machine learning models are Data Science is a black box

Epistemic transparency

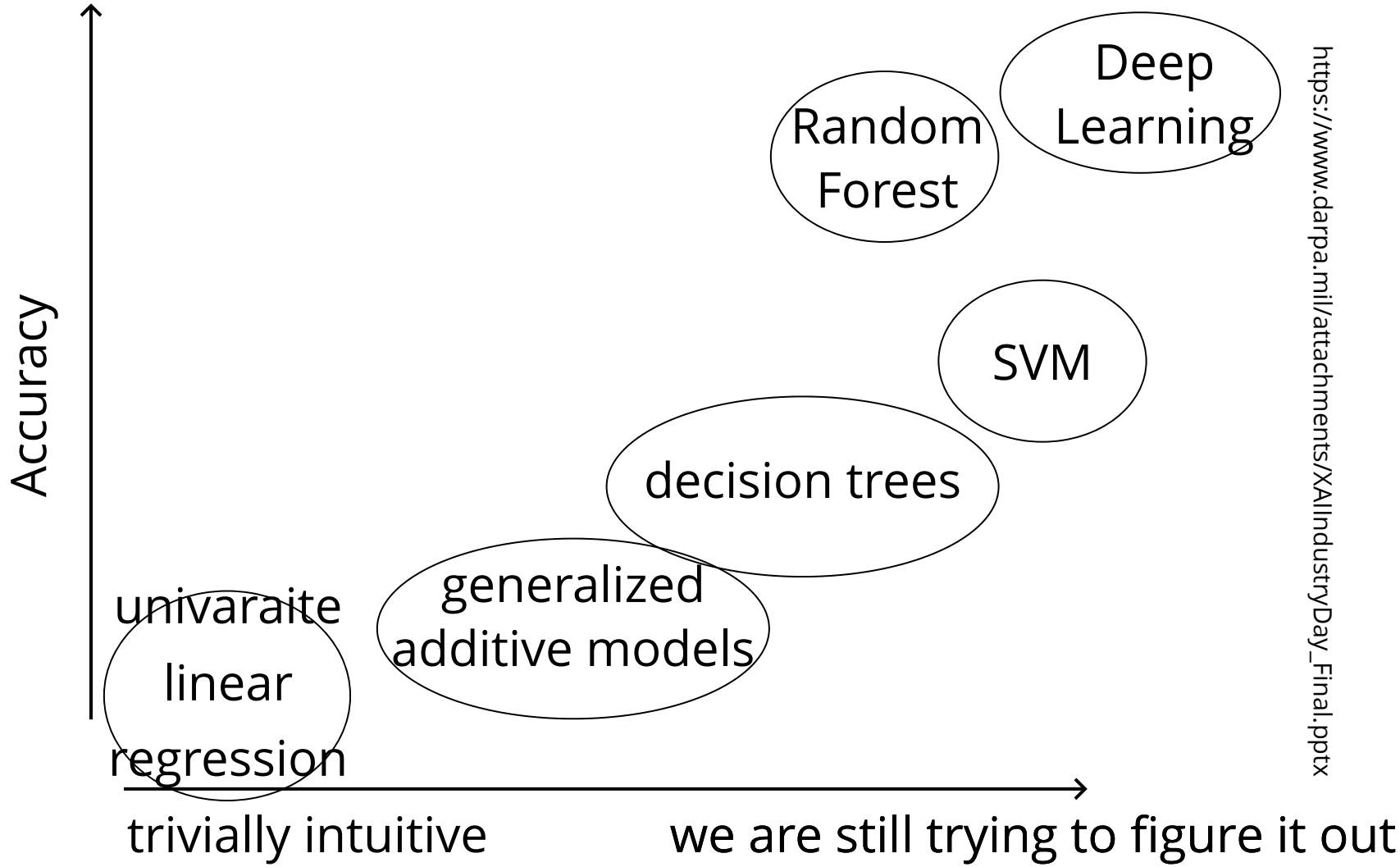
Accountability: who is responsible if an algorithm does harm

Right to explanation: the scope of a general "right to explanation" is a matter of ongoing debate

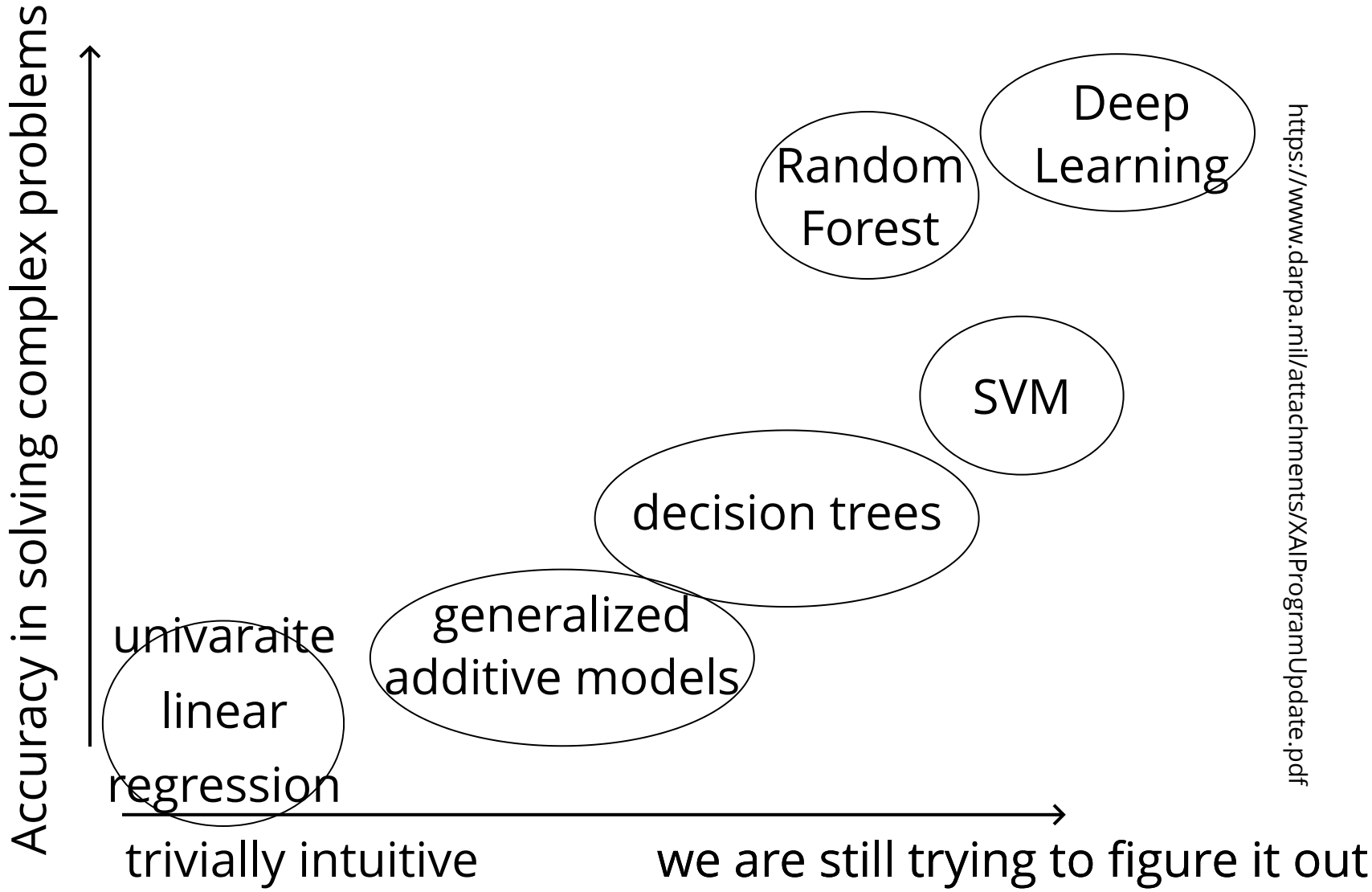


tration by Hanne Morstad

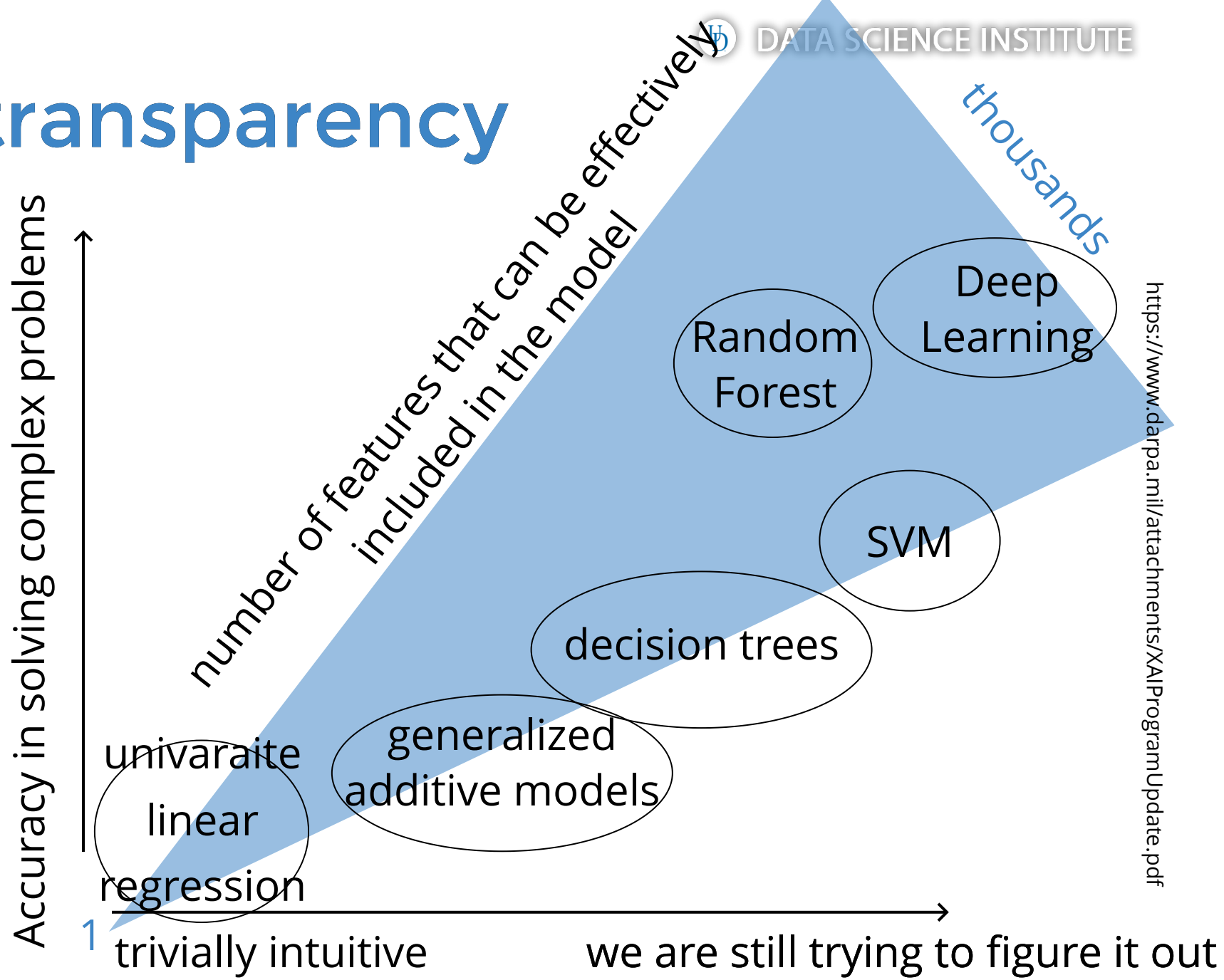
algorithmic transparency



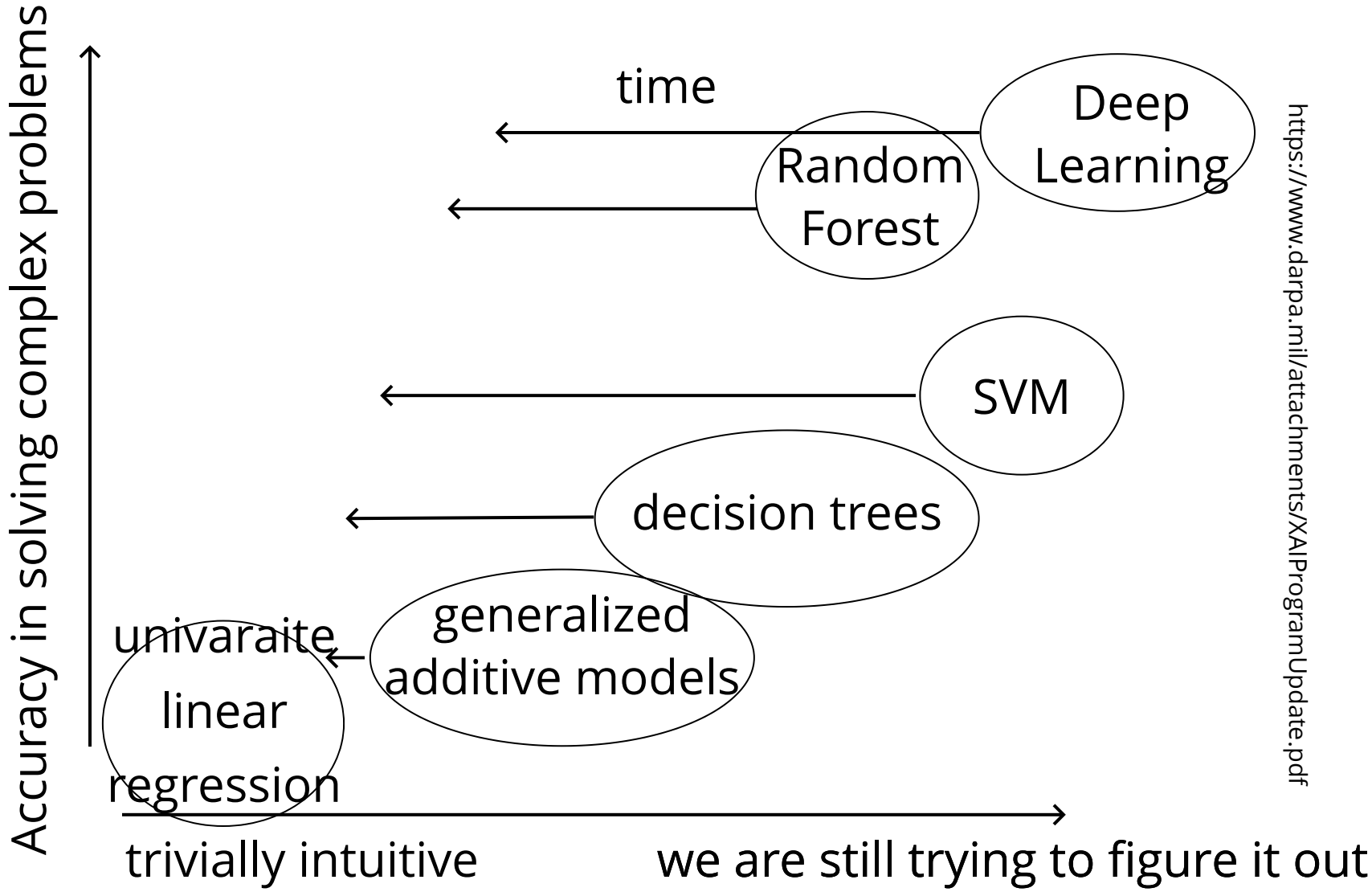
algorithmic transparency



algorithmic transparency



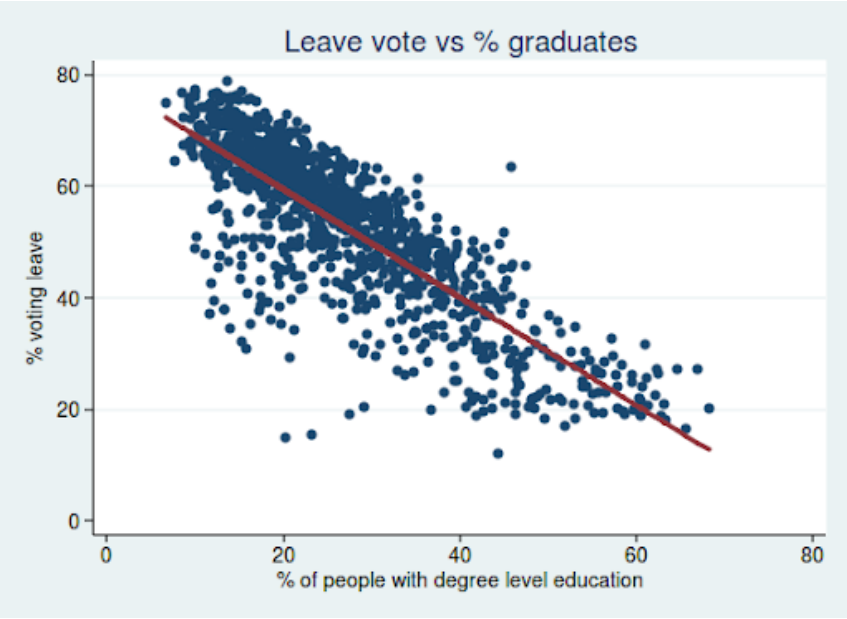
algorithmic transparency



algorithmic transparency

1

Machine learning: any method that learns parameters from the data



http://www.statsguy.co.uk/brexit-voting-and-education/

2

The transparency of an algorithm is proportional to its complexity *and* the complexity of the data space

3

The transparency of an algorithm is limited by our own ability and preparedness to interpret it

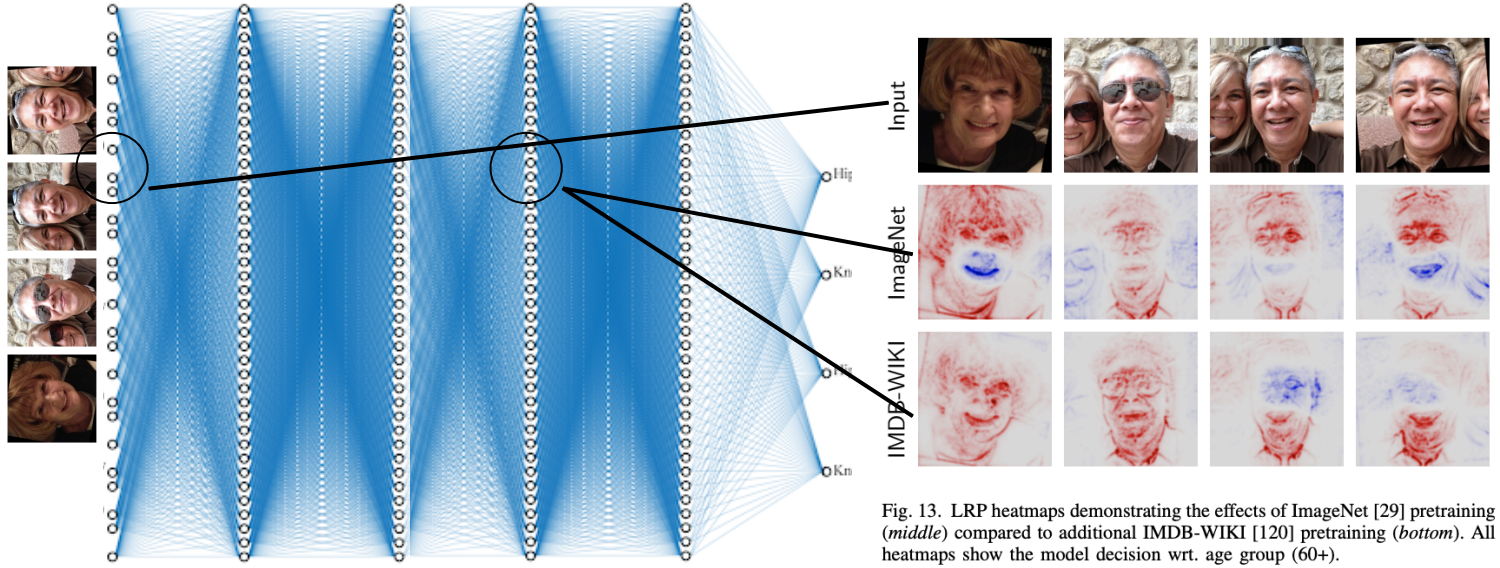


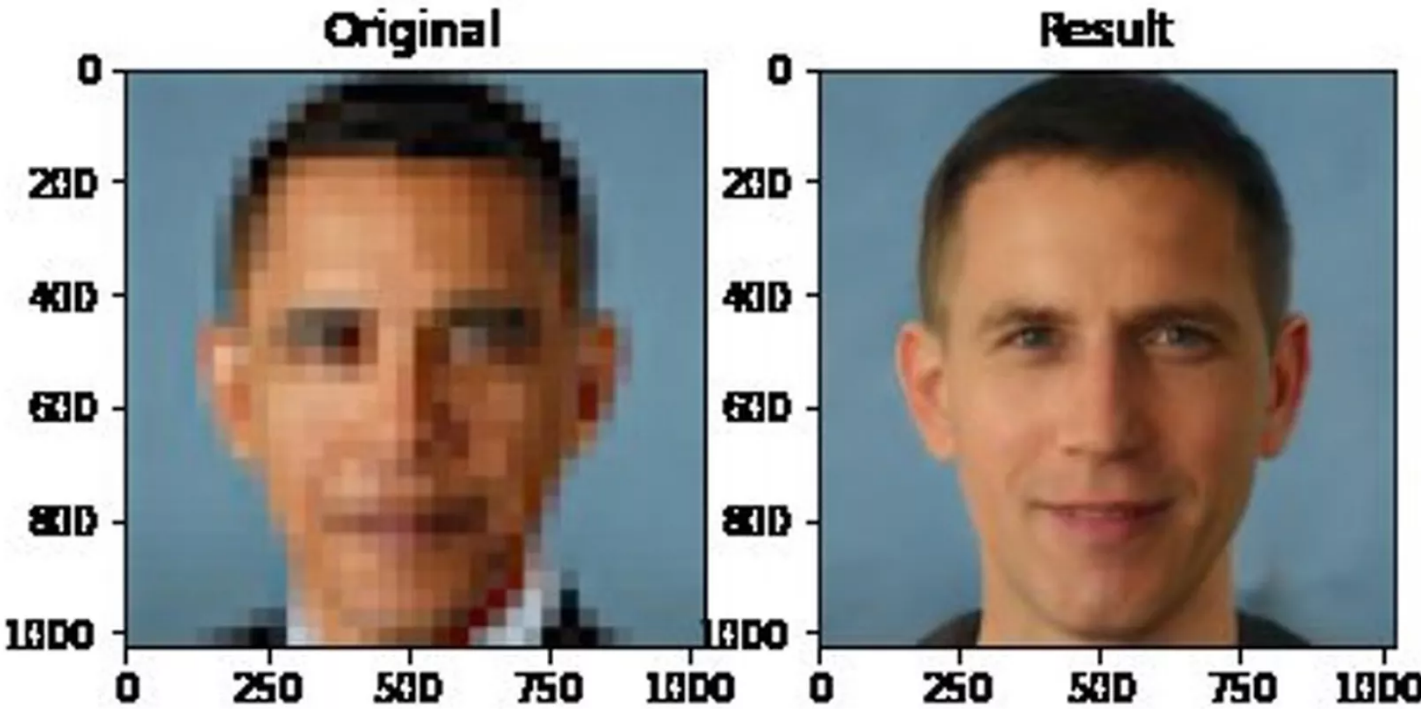
Fig. 13. LRP heatmaps demonstrating the effects of ImageNet [29] pretraining (*middle*) compared to additional IMDB-WIKI [120] pretraining (*bottom*). All heatmaps show the model decision wrt. age group (60+).

Toward Interpretable Machine Learning, Samek+2003

models are neutral, the bias is in the data

Why does this AI model whitens Obama face?

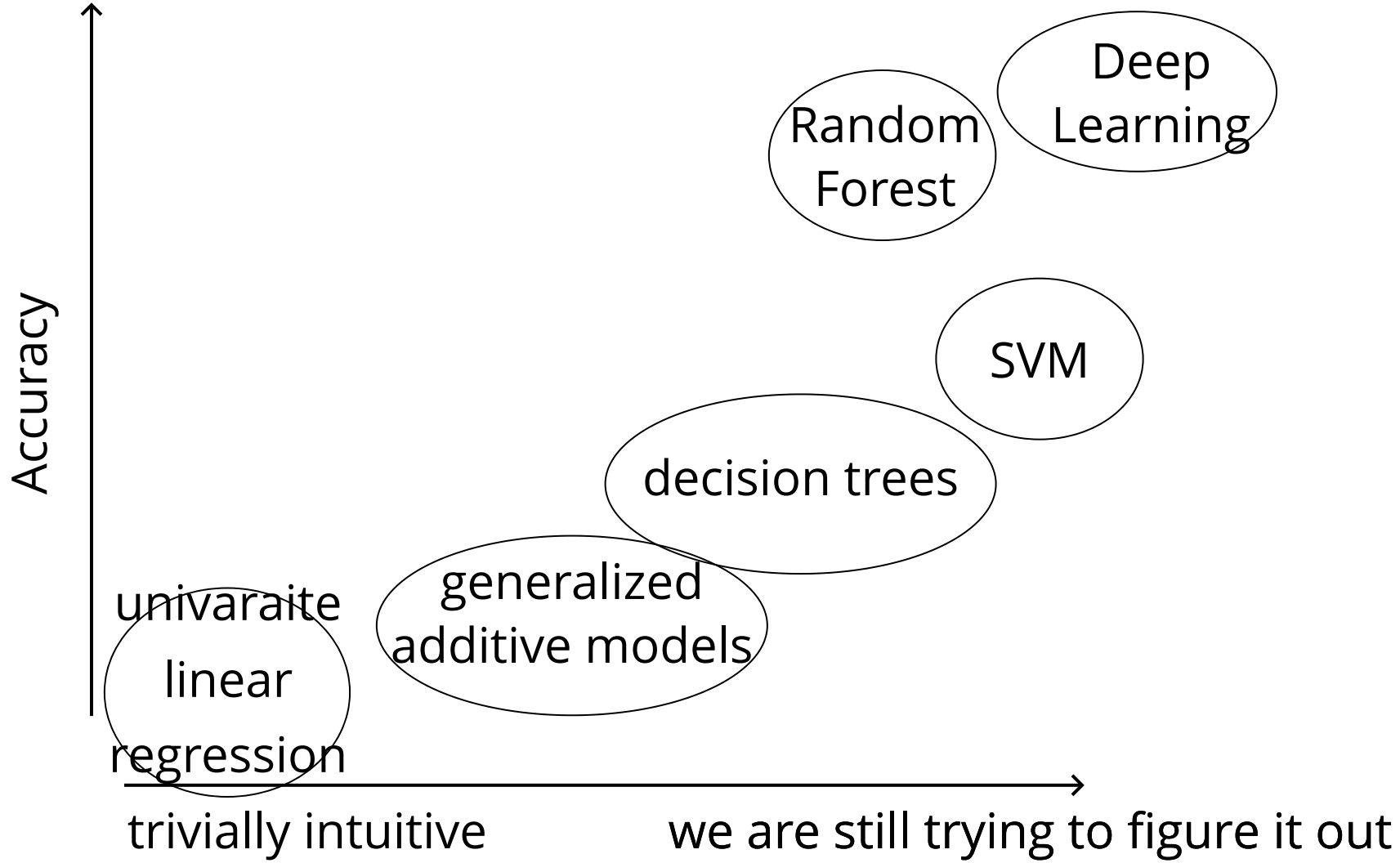
Simple answer: the data is biased. The algorithm is fed more images of white people



https://www.theverge.com/21298762/face-depixelizer-ai-machine-learning-tool-pulse-stylegan-obama-bias

where is the bias?

Decide which model is appropriate (depends on data and question)



where is the bias?

Decide what your target function is

Machine learning models are functions that "learn" their parameters from the data

They "learn" by minimizing or maximize some quantity.

What should you minimize?

https://miro.medium.com/max/960/1*imhEKEpzX24CC_LIlureBw.gif

https://towardsdatascience.com/machine-learning-fundamentals-via-linear-regression-41a5d11f5220

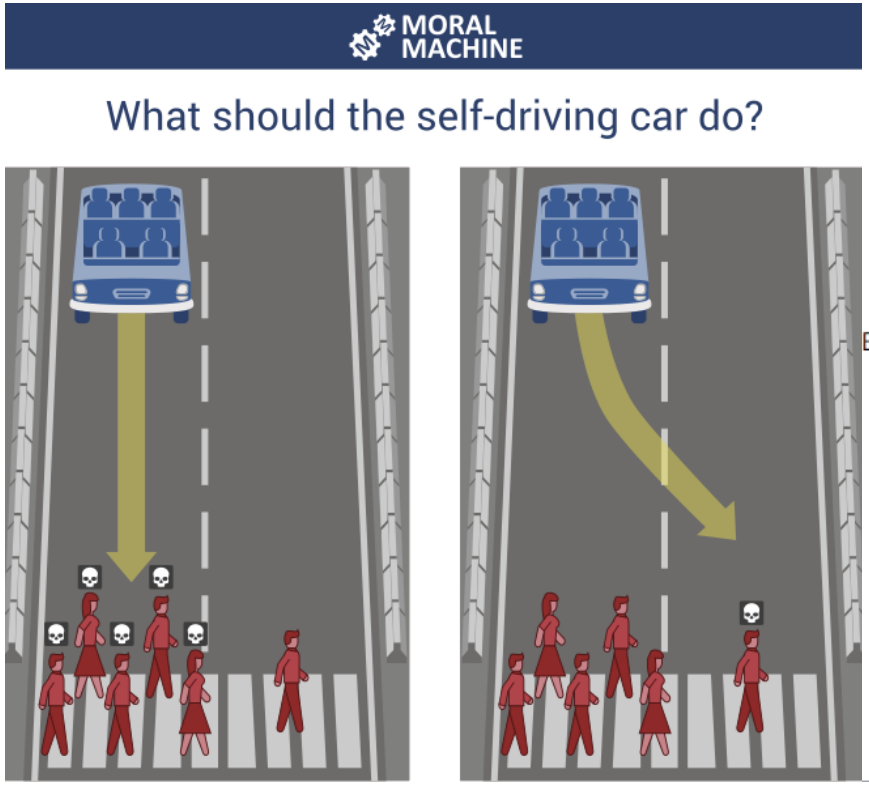
where is the bias?

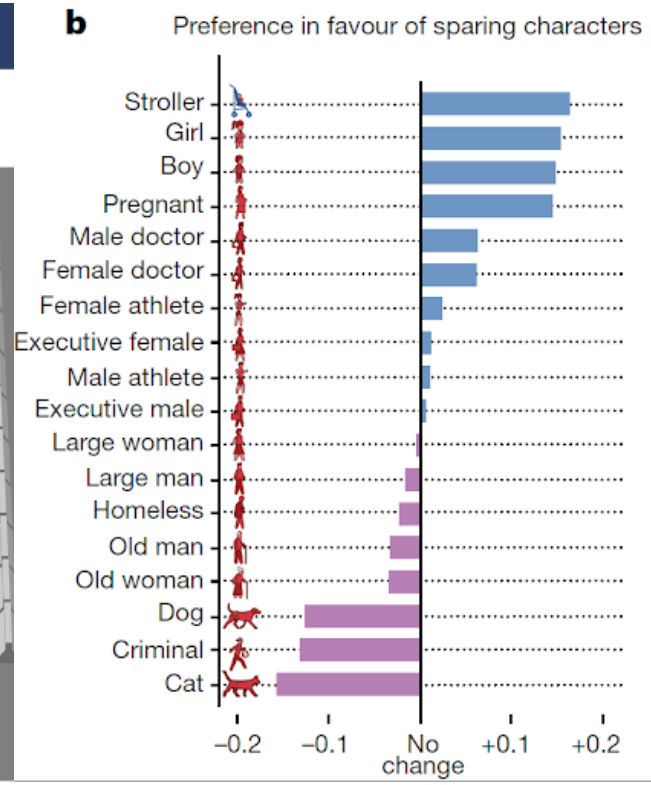
2 - cost function

self-driving cars

They "learn" by minimizing or maximize some quantity.

What should you minimize?





the hypothetical trolley problem suddenly is real

where is the bias?

2 - cost function

prosecutorial justice

Illustration: The In

https://theintercept.com/2020/07/12/risk-assessment-tools-bail-reform/

They "learn" by minimizing or maximize some quantity.

What should you minimize?

A BAIL REFORM TOOL INTENDED TO CURB MASS INCARCERATION HAS ONLY REPLICATED BIASES IN THE CRIMINAL JUSTICE SYSTEM

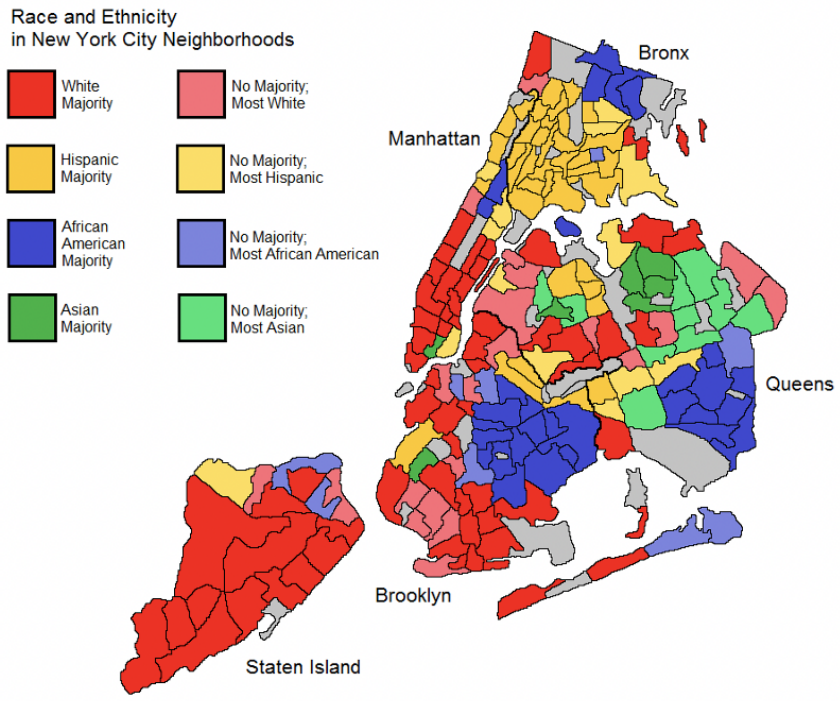
minimize number of people incarcerated injustly OR *maximize public safety*

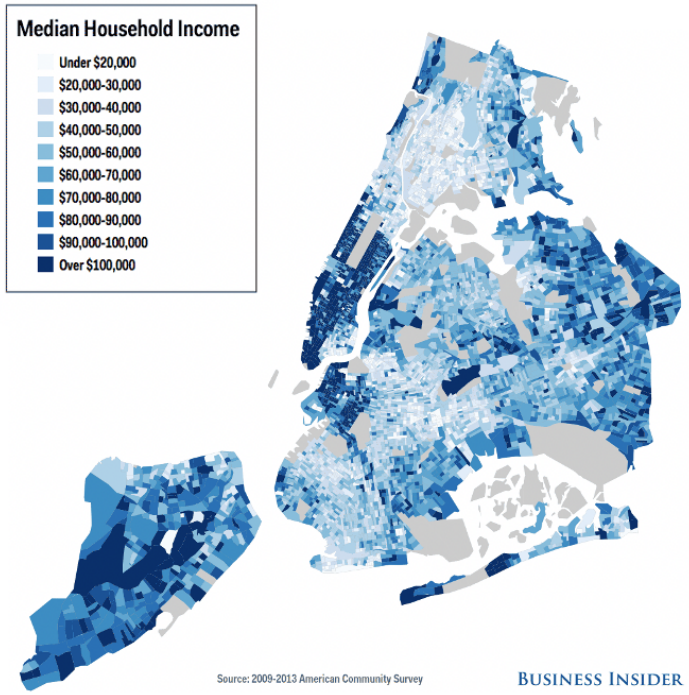
where is the bias?

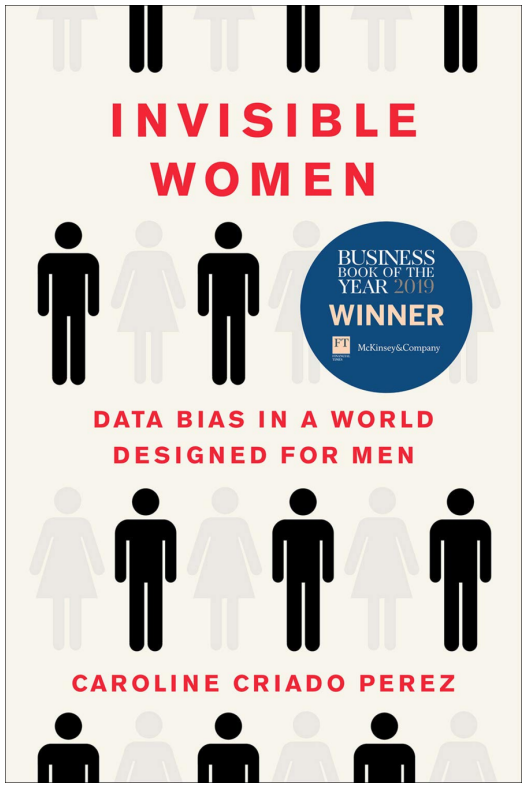
3 - data selection and preparation

Explore the data → discover some of the bias (trust me, there is more!) → remove the bias... (few try)

it's not easy
there's covariance
missing data







~~models are neutral, the bias is in the data~~

Should AI reflect

who we are

or should it reflect who we aspire to be?

~~models are neutral, the bias is in the data~~

The bias is in the data

Should AI reflect

who we are

or should it reflect who we aspire to be?

~~models are neutral, the bias is in the data~~

The bias is in the data

The bias is in the models and the decision we make

Should AI reflect

who we are

or should it reflect who we aspire to be?

~~models are neutral, the bias is in the data~~

The bias is in the data

The bias is in the models and the decision we make

The bias is in how we choose to optimize our model

Should AI reflect

who we are

or should it reflect who we aspire to be?

~~models are neutral, the bias is in the data~~

The bias is in the data

The bias is in the models and the decision we make

The bias is in how we choose to optimize our model

Should AI reflect

who we are

or should it reflect who we aspire to be?

The bias is society that provides the framework to validate our biased models

~~models are neutral, the bias is in the data~~

The bias is in the data

The bias is in the models and the decision we make

The bias is in how we choose to optimize our model

Should AI reflect

who we are

(*and enforce and grow our bias*)

or should it reflect who we aspire to be?

(*and who decides what that is?*)

The bias is society that provides the framework to validate our biased models

key concepts

MACHINE LEARNING

- Machine Learning models are parametrized representation of "reality" where the parameters are learned from finite sets of realizations of that reality
- Unsupervised learning: all variables observed for all data, looking for natural grouping of datapoints in the N-dim space
- Supervised learning: a target variable is known for (a subset of) the data and the goal is to predict it for new (the rest of the) data

DATA ETHICS

- epistemic transparency:not all models are the same
- there is a tradeoff between epistemic transparency and the ability to handle complex data
- The bias enter data science in (at least) data; model selection; target function and optimization choices; validation

references

Midterm project due!

12/20 (regular homework timeline, no other homework)

Write a project proposal for your final projefollowing

this template