A Path to Simpler Models Starts with Noise

Lesia Semenova, Harry Chen, Ronald Parr, Cynthia Rudin




Duke DEPARTMENT of COMPUTER SCIENCE

Machine Learning, 11, 63–91 (1993)

Very Simple Classification Rules Perform Well on Most Commonly Used Datasets

ROBERT C. HOLTE HOLTE@CSI.UOTTAWA.CA
Computer Science Department, University of Ottawa, Ottawa, Canada K1N 6N5

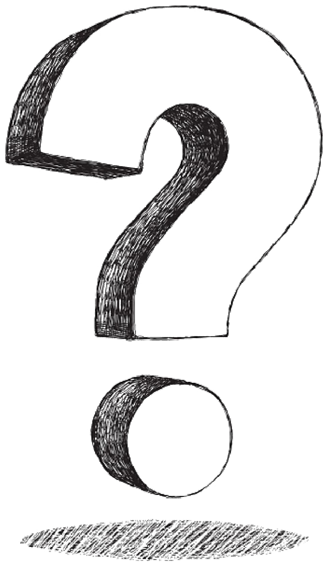
Larger
Rashomon ratios

Correlate with

Existence of simpler
(or interpretable)
models

$$\textbf{Rashomon ratio} = \frac{\text{Size of }\{\text{models that perform equally well}\}}{\text{Size of the hypothesis space}}$$

Semenova et al. On the existence of simpler machine learning models. FaccT, 2022



Leads to →

Larger
Rashomon ratios

Correlate with ↔

Existence of simpler
(or interpretable)
models

Noise
in data generation

Leads to →

Larger
Rashomon ratios

↔ **Correlate with**

Existence of simpler
(or interpretable)
models

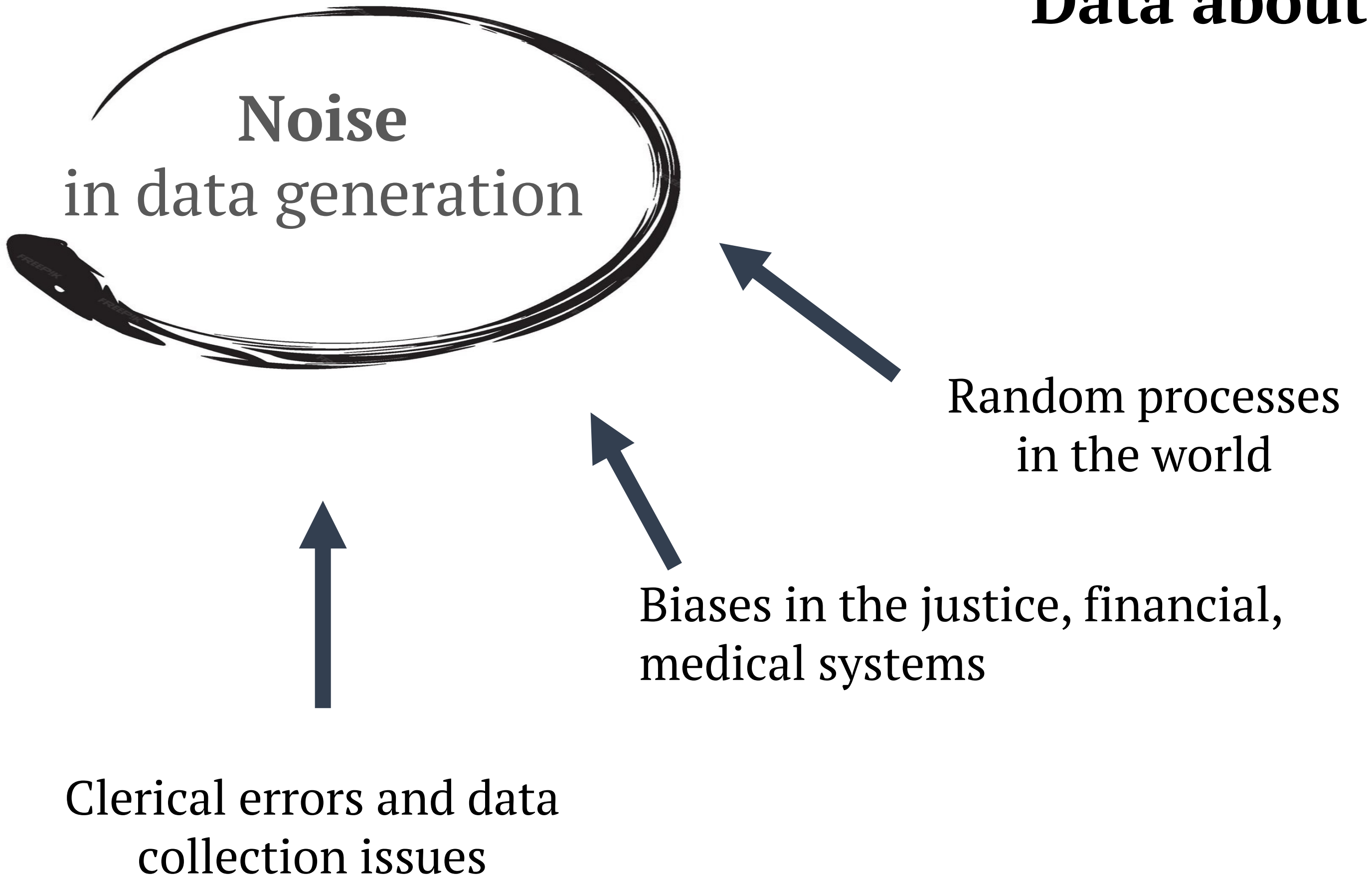
Noise
in data generation

Larger
Rashomon ratios

Existence of simpler
(or interpretable)
models

decisions made by analysts
to compensate for the noise
(think cross-validation and
regularization)

Data about humans are noisy!



Noise in high-stakes decision domains

leads to ***technical justification***

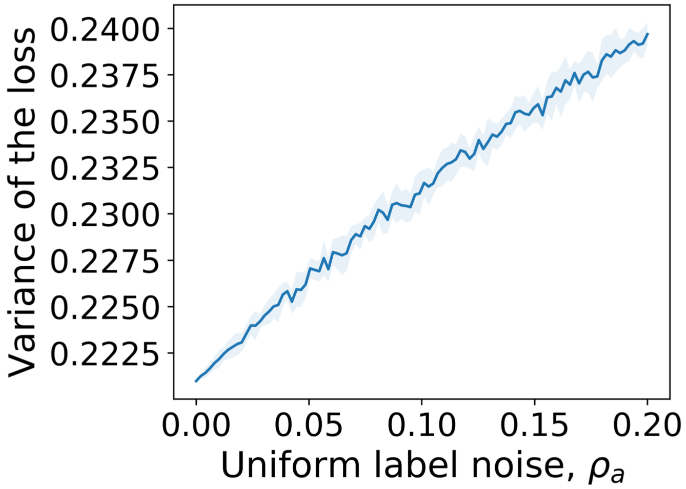
for demanding

simpler (interpretable) models

- COMPAS recidivism risk score prediction dataset
- Space of decision trees

noise

- COMPAS recidivism risk score prediction dataset
- Space of decision trees

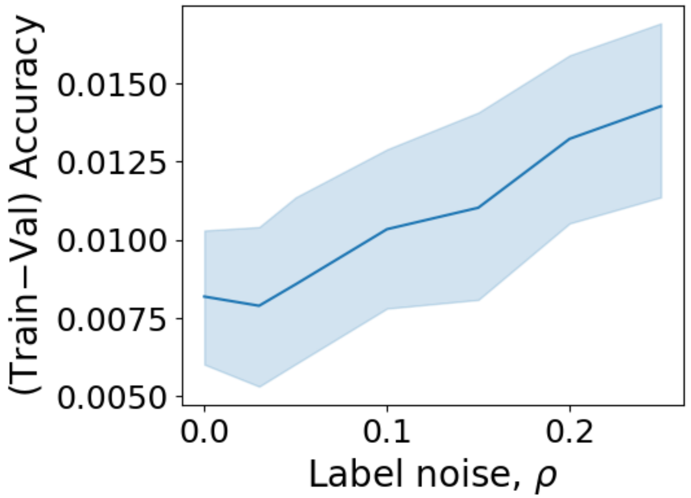


↗ variance of the loss

noise

Uniform random label noise (Theorem 2)
Labels flip with probability p(x) (Theorem 12)
Margin noise (Theorem 15)

- COMPAS recidivism risk score prediction dataset
- Space of decision trees



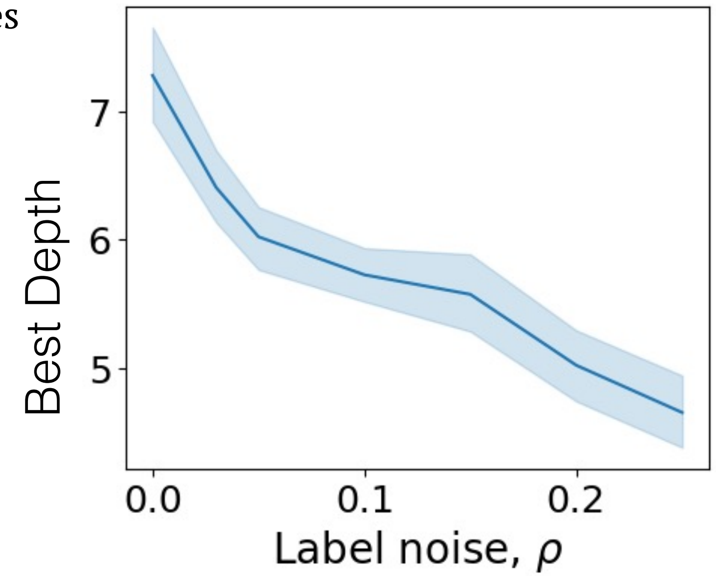
generalization

variance of the loss

noise

Generalization bounds for continuous and discrete hypothesis spaces

- COMPAS recidivism risk score prediction dataset
- Space of decision trees



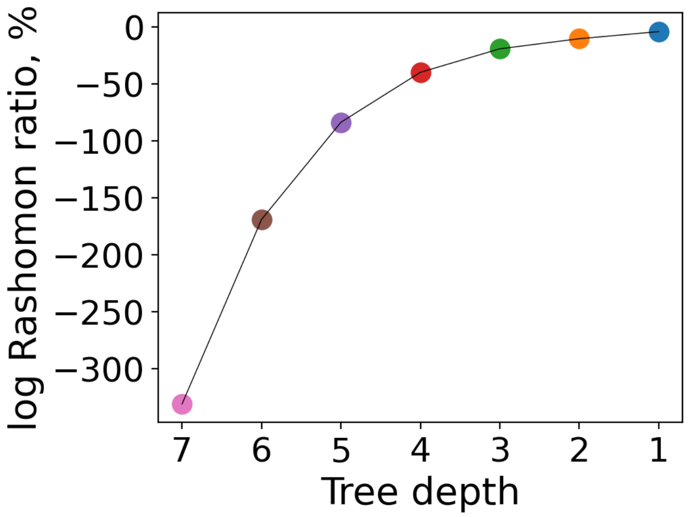
↘ complexity of the space

↘ generalization

↗ variance of the loss

noise

- COMPAS recidivism risk score prediction dataset
- Space of decision trees



↗ the Rashomon ratio

↘ complexity of the space

↘ generalization

↗ variance of the loss

noise

Rashomon ratio is larger for decision trees of smaller depth (Proposition 6)
Rashomon ratio increases with noise for ridge regression (Theorem 7)

Path

- noise
- ↗ variance of the loss
- ↘ generalization
- ↘ complexity of the space
- ↗ the Rashomon ratio

$\nearrow$ **the Rashomon ratio**

Our results *explain* **why** on *noisier* datasets **simpler models often tend to perform as well as black boxes**

noise

A Path to Simpler Models Starts With Noise

Lesia Semenova **Harry Chen** **Ronald Parr** **Cynthia Rudin**
Department of Computer Science, Duke University
{lesia.semenova,harry.chen084,ronald.parr,cynthia.rudin}@duke.edu

https://arxiv.org/pdf/2310.19726.pdf