DIFFERENTIAL REPLICATION AS A TOOL FOR MACHINE LEARNING ACCOUNTABILITY IN PRACTICE.

IRENE UNCETA.

Instituto de Ciencia de Datos e Inteligencia Artificial
December 22, 2021

esade
RAMON LLULL UNIVERSITY

ME.

ME.

Bilbao / Barcelona

ME.

Bilbao / Barcelona

Bachelor in Physics
MSc Computational Science
Industrial PhD in Mathematics and Computer Science

ME.

Bilbao / Barcelona

Bachelor in Physics
MSc Computational Science
Industrial PhD in Mathematics and Computer Science

Sinnergiak Social Innovation UPV/EHU
KSNET Knowledge Sharing Network
Tech for Innovation

ME.

Bilbao / Barcelona

Bachelor in Physics
MSc Computational Science
Industrial PhD in Mathematics and Computer Science

Sinnergiak Social Innovation UPV/EHU
KSNET Knowledge Sharing Network
Tech for Innovation

Decidata

ME.

Bilbao / Barcelona

Bachelor in Physics
MSc Computational Science
Industrial PhD in Mathematics and Computer Science

Sinnergiak Social Innovation UPV/EHU
KSNET Knowledge Sharing Network
Tech for Innovation

Decidata

BBVA Data & Analytics

ME.

Bilbao / Barcelona

Bachelor in Physics
MSc Computational Science
Industrial PhD in Mathematics and Computer Science

Sinnergiak Social Innovation UPV/EHU
KSNET Knowledge Sharing Network
Tech for Innovation

Decidata

BBVA Data & Analytics

ESADE

INTRODUCTION.

Decisions based on machine learning have a **substantial impact** on our everyday lives.

CHEN, C., SEFF, A., KORNHAUSER, A., AND XIAO, J. Deepdriving: Learning affordance for direct perception in autonomous driving. In *Proceedings of the IEEE International Conference on Computer Vision* (Santiago, Chile, 2015).

GARG, A., ADHIKARI, N., MCDONALD, H., ROSAS ARELLANO, M., DEVEREAUX, P., BEYENE, J., SAN, J., AND HAYNES, R. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA* 293, 10 (2005).

KOU, Y., LU, C.-T., SIRWONWATTANA, S., AND HUANG, Y. P. Survey of fraud detection techniques. In *Proceedings of the IEEE International Conference on Networking, Sensing and Control* (Taipei, Taiwan, 2004)

LISBOA, P., IFEACHOR, E., AND SZCZEPANIAK, P. *Artificial Neural Networks in Biomedicine.* Springer Science & Business Media, Berlin, Germany (2000)

SRIVASTAVA, A., KUNDU, A., SURAL, S., AND MAJUMDAR, A. Credit card fraud detection using hidden markov models. *IEEE Transactions on Dependable and Secure Computing* 5, 1 (2008)

INTRODUCTION.

Decisions based on machine learning have a **substantial impact** on our everyday lives.

However, deploying machine learning in practice **remains a challenge** for most companies.

BAROCAS, S.,AND BOYD, D. Engaging the ethics of data science in practice. *Communications of the ACM* 60, 11 (2017).
IDOINE, C., KRENSKY, P., LINDEN, A., AND BRETHENOUX, E. Magic quadrant for data science and machine learning platforms. Tech. rep., Gartner Research (2019)
KROLL, J. The fallacy of inscrutability. *Philosophical Transactions of the Royal Society* 376 (2018).
VEALE, M., AND BINNS, R. Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society* 4, 2 (2017).

INTRODUCTION.

Decisions based on machine learning have a **substantial impact** on our everyday lives.

However, deploying machine learning in practice **remains a challenge** for most companies.

As a result, industrial machine learning today is far from being **sustainable.**

AMODEI, D., OLAH, C., STEINHARDT, J., CHRISTIANO, P., SCHULMAN,J., AND MANÉ, D. Concrete problems in AI safety. arXiv:1606.06565 (2016).

BAROCAS, S., AND SELBST, A. D. Big data's disparate impact. *California Law Review* 104, 3 (2016).

BERK, R., HEIDARI, H., JABBARI, S., KEARNS, M., AND ROTH, A. Fairness in criminal justice risk assessments: The state of the art. *Sociological Methods & Research* (2018).

BOLUKBASI, T., CHANG, K. W., ZOU, J., SALIGRAMA, V., AND KALAI,A. Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. In *Proceedings of the 29th International Conference on Neural Information Processing Systems* (Barcelona, Spain, 2016).

BOSTROM, N. Ethical issues in advanced artificial intelligence. In *Science Fiction and Philosophy: From Time Travel to Superintelligence*, Wiley-Blackwell, New Jersey, NJ, USA (2009).

PODESTA, J, PRITZKER, P., MONIZ, E., HOLDREN, J., AND ZIENTS,J. Big data: Seizing opportunities, preserving values. Tech.rep., Executive Office of the President. The White House (2014).

INTRODUCTION.

INTRODUCTION.

Which are the constraints that prevent a sustainable machine learning deployment?

How can we adapt trained machine learning models to changes in their environment?

How can we modify models which display several shortcomings but which have already been served into production?

How is this problem formalized?

Which tools do we have at our disposal to solve it?

Which control mechanisms can be enforced to prevent undesired negative impacts of machine learning?

CONTENTS.

CONTENTS.

CONTENTS.

CONTENTS.

CONTENTS.

CONTENTS.

CONTENTS.

CONTENTS.

01

MACHINE LEARNING ACCOUNTABILITY

SURVIVAL OF THE FITTEST.

The **level of adaptation to their environment** plays a key role in ensuring preservation of living creatures. The same can be said for machine learning models.

BAROCAS, S. AND BOYD, D. Engaging the ethics of data science in practice. *Communications of the ACM* 60, 11 (2017).
DARWIN, C. *On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life*. John Murray, London, UK (1859).
KROLL, J. The fallacy of inscrutability. *Philosophical Transactions of the Royal Society* 376 (2018).
VEALE, M. AND BINNS, R. Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society* 4, 2 (2017).

SURVIVAL OF THE FITTEST.

The **level of adaptation to their environment** plays a key role in ensuring preservation of living creatures. The same can be said for machine learning models.

Machine learning models interact which a **large number of elements** that tend to change in time.

SURVIVAL OF THE FITTEST.

The **level of adaptation to their environment** plays a key role in ensuring preservation of living creatures. The same can be said for machine learning models.

Machine learning models interact which a **large number of elements** that tend to change in time.

Technological infrastructure
Data governance
Business alignment
Ethics and business rules
Market trends and globalization
Third party providers
Regulatory framework

THE NEED FOR ACCOUNTABILITY.

In recent years, an increasing number of voices have publicly denounced the **shortcomings of machine learning** and their potential negative impact.

AMODEI, D., OLAH, C., STEINHARDT, J., CHRISTIANO, P., SCHULMAN,J., AND MANÉ, D. Concrete problems in AI safety. arXiv:1606.06565 (2016).

BAROCAS, S., AND SELBST, A. D. Big data's disparate impact. *California Law Review* 104, 3 (2016).

BOSTROM, N. Ethical issues in advanced artificial intelligence. In *Science Fiction and Philosophy: From Time Travel to Superintelligence*. Wiley-Blackwell, New Jersey, NJ, USA (2009)

GLOBAL FUTURE COUNCIL ON HUMAN RIGHTS. How to prevent discriminatory out-comes in machine learning. Tech. rep., World Economic Forum (2016).

PODESTA, J, PRITZKER, P., MONIZ, E., HOLDREN, J., AND ZIENTS, J. Big data: Seizing opportunities, preserving values. Tech.rep., Executive Office of the President. The White House (2014)

THE NEED FOR ACCOUNTABILITY.

In recent years, an increasing number of voices have publicly denounced the **shortcomings of machine learning** and their potential negative impact.

As a result, there is a growing **demand for accountability**.

ANGWIN, J. Make algorithms accountable. *The New York Times* (2016).

EUROPEAN PARLIAMENT. Civil law rules on robotics. European Parliament resolution of 16 February 2017 with recommendations to the Commission on civil law rules on robotics 2015/2103(INL). No.: P8TA-PROV(2017)00 51. (2017).

EXECUTIVE OFFICE OF THE PRESIDENT. The national artificial intelligence research and development strategic plan. Tech. rep., National Science and Technology Council (2016).

GOODMAN, B. W. A step towards accountable algorithms?: Algorithmic discrimination and the European Union general data protection. In *Proceedings of the 29th International Conference on Neural Information Processing Systems* (Barcelona, Spain, 2016)

THE NEED FOR ACCOUNTABILITY.

In recent years, an increasing number of voices have publicly denounced the **shortcomings of machine learning** and their potential negative impact.

As a result, there is a growing **demand for accountability**.

Instrument through which agents can be held accountable of the potential negative consequences of automatic decisions

ANGWIN, J. Make algorithms accountable. *The New York Times* (2016).

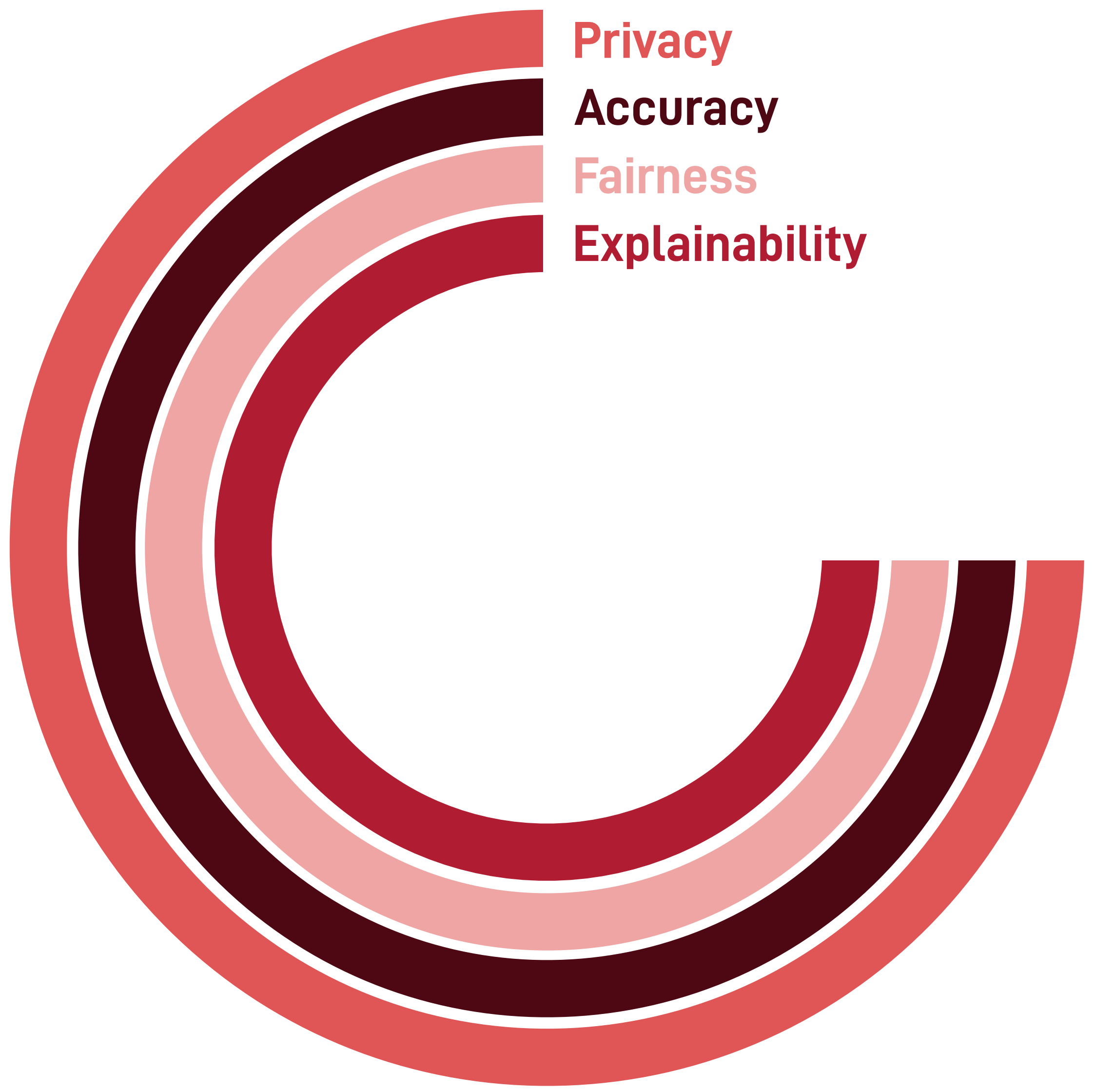
EUROPEAN PARLIAMENT. Civil law rules on robotics. European Parliament resolution of 16 February 2017 with recommendations to the Commission on civil law rules on robotics 2015/2103(INL). No.: P8TA-PROV(2017)00 51. (2017).

EXECUTIVE OFFICE OF THE PRESIDENT. The national artificial intelligence research and development strategic plan. Tech. rep., National Science and Technology Council (2016).

GOODMAN, B. W. A step towards accountable algorithms?: Algorithmic discrimination and the European Union general data protection. In *Proceedings of the 29th International Conference on Neural Information Processing Systems* (Barcelona, Spain, 2016)

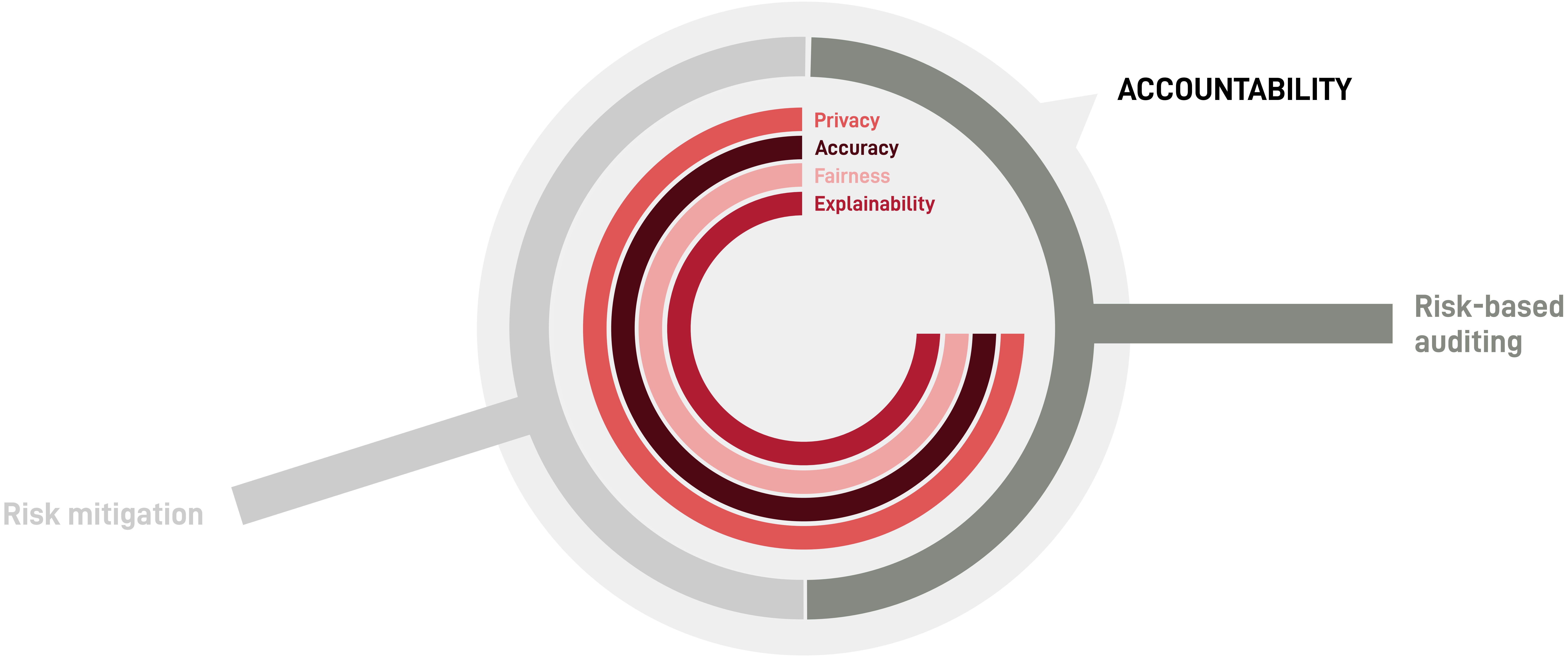
MACHINE LEARNING ACCOUNTABILITY IN PRACTICE.

MACHINE LEARNING ACCOUNTABILITY IN PRACTICE.

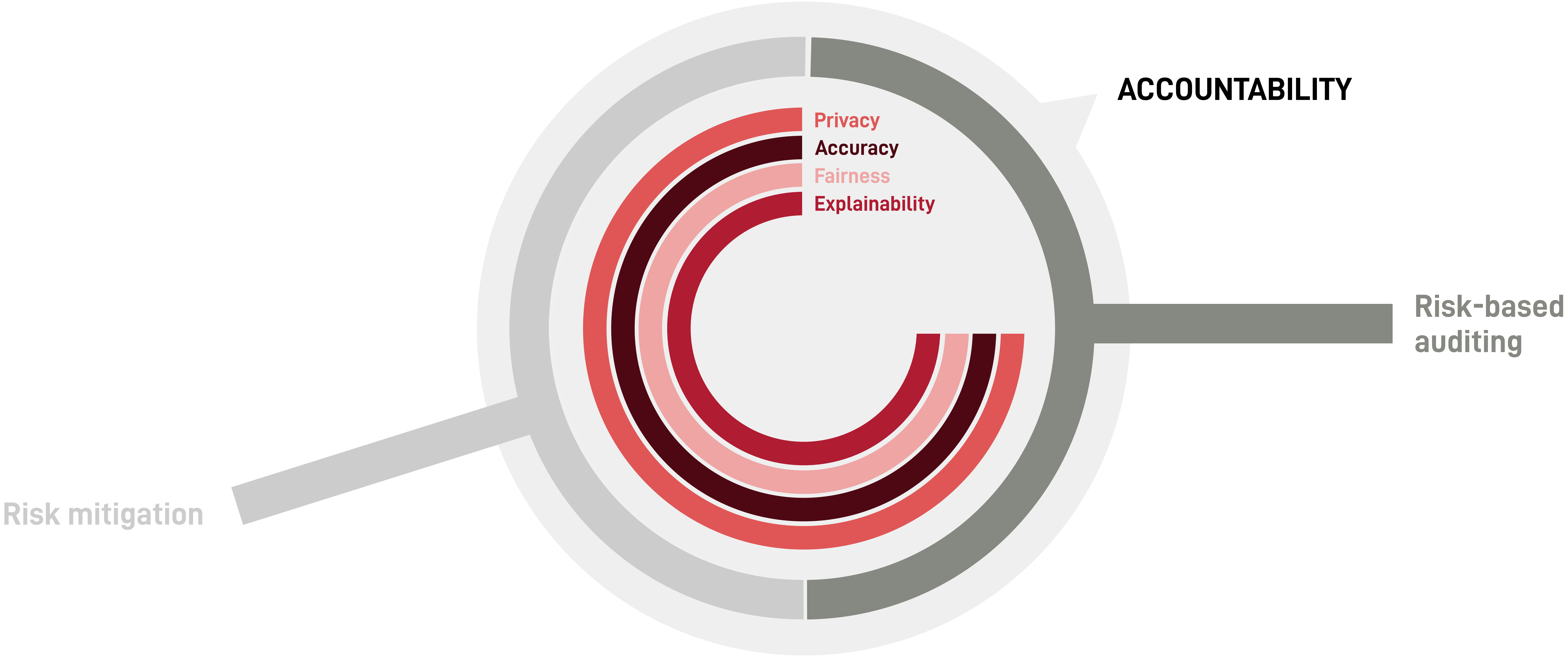


MACHINE LEARNING ACCOUNTABILITY IN PRACTICE.

MACHINE LEARNING ACCOUNTABILITY IN PRACTICE.



MACHINE LEARNING ACCOUNTABILITY IN PRACTICE.



LUCA, M., KLEINBERG, J., AND MULLAINATHAN, S. Algorithms need managers, too. *Harvard Business Review* (2016).

SCULLEY, D., HOLT, G., GOLOVIN, D., DAVYDOV, E., PHILLIPS, T.,EBNER, D., CHAUDHARY, V., AND YOUNG, M. Machine learning: The high interest credit card of technical debt. In *SE4ML: Software Engineering for Machine Learning* (Montreal, Canada, 2014).

CONTENTS.

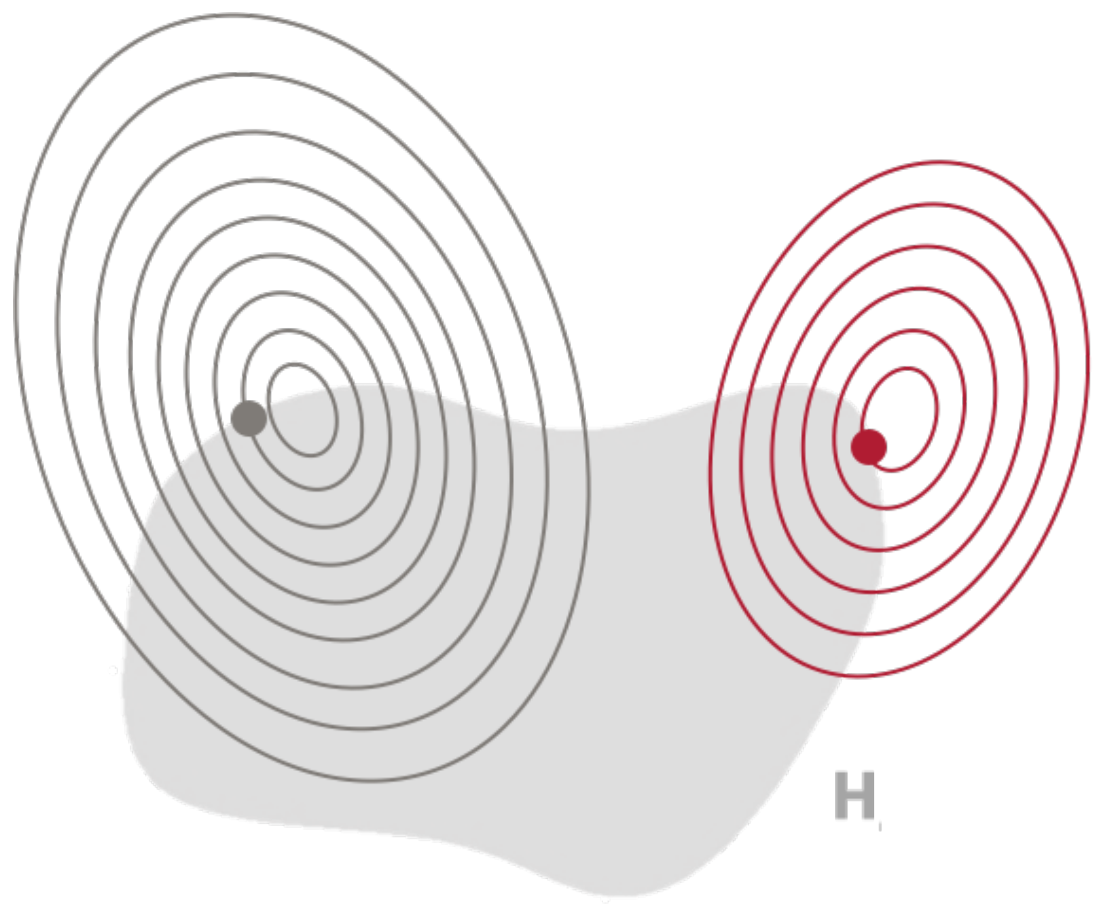
CONTENTS.

02

ENVIRONMENTAL ADAPTATION AND DIFFERENTIAL REPLICATION

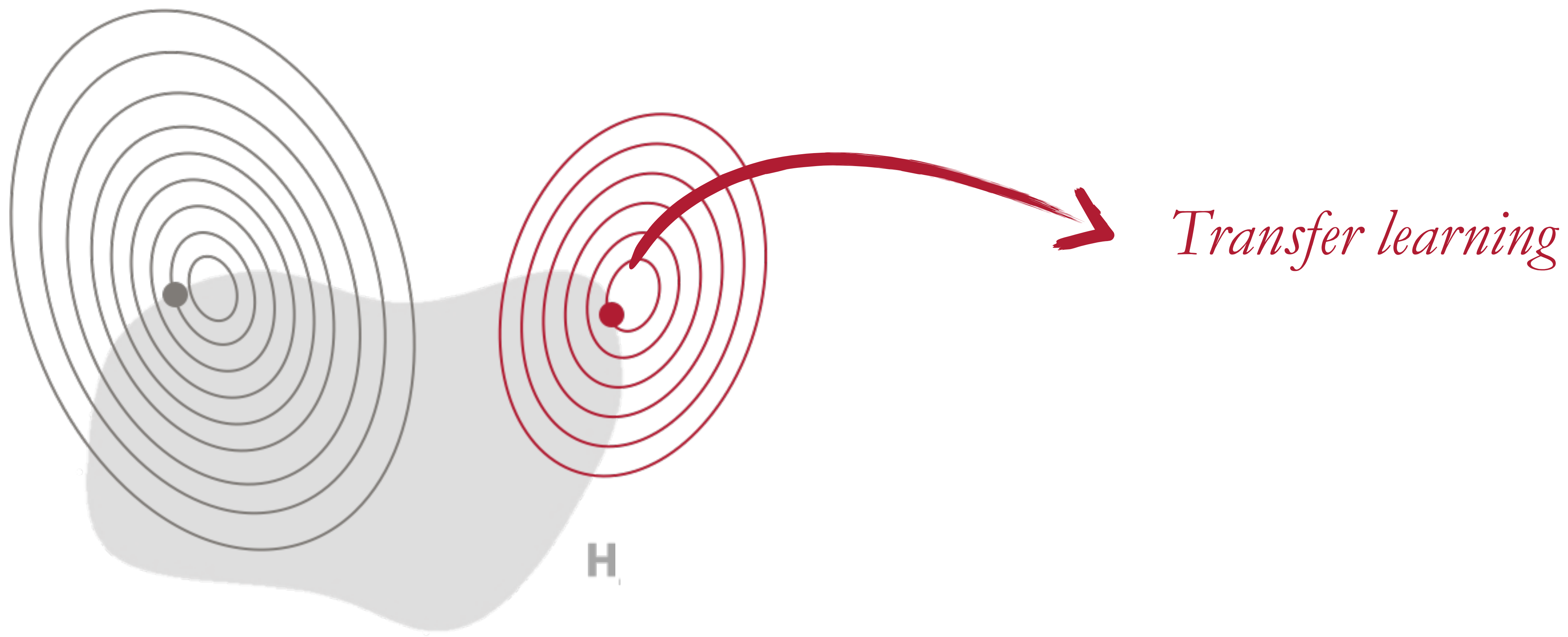
ENVIRONMENTAL ADAPTATION.

ENVIRONMENTAL ADAPTATION.



LI, D., YANG, Y., SONG, Y., AND HOSPEDALES, T. Deeper, broader and artier domain generalization. In *Proceedings of the IEEE International Conference on Computer Vision* (2017)
TORREY, L., AND SHAVLIK, J. Transfer learning. In *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques.* IGI Global, (PA, USA, 2010)
PAN, S., AND YANG, Q. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering* 22, 10 (2010).

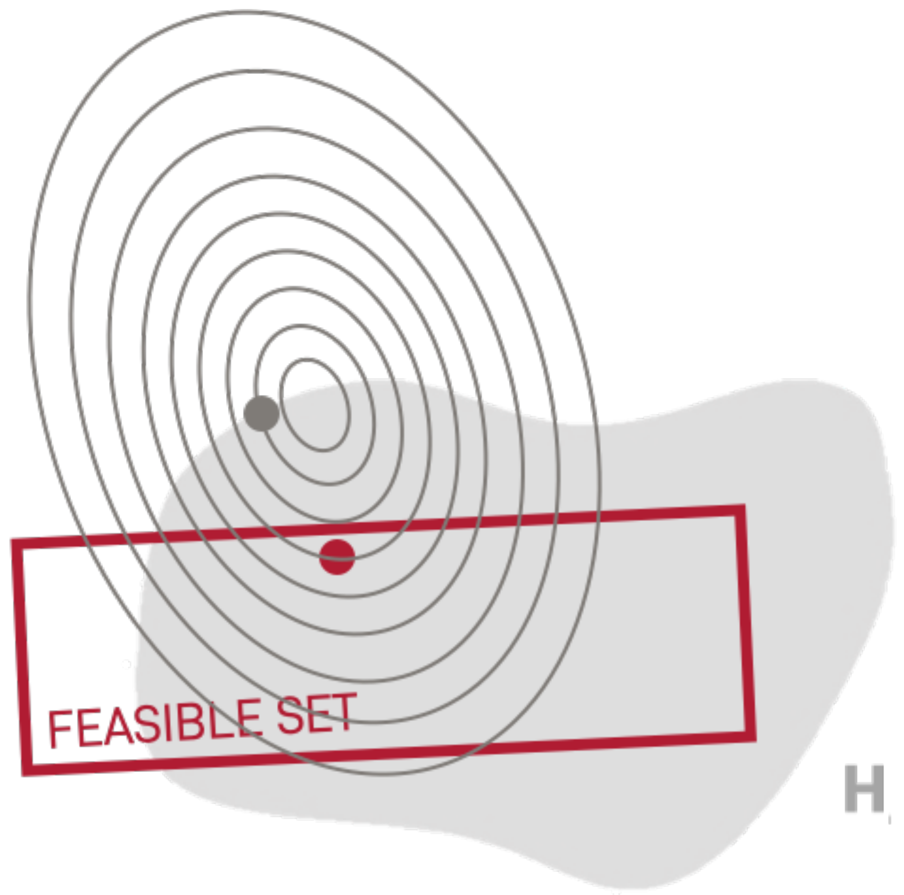
ENVIRONMENTAL ADAPTATION.



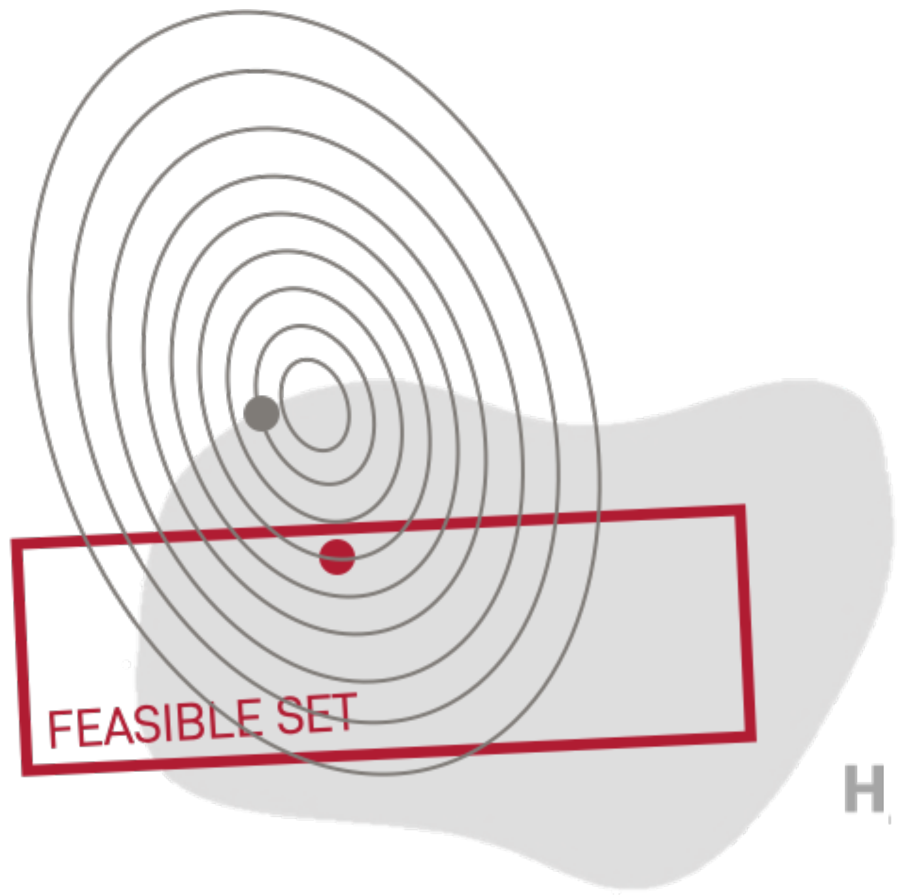
LI, D., YANG, Y., SONG, Y., AND HOSPEDALES, T. Deeper, broader and artier domain generalization. In *Proceedings of the IEEE International Conference on Computer Vision* (2017)
TORREY, L., AND SHAVLIK, J. Transfer learning. In *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques.* IGI Global, (PA, USA, 2010)
PAN, S., AND YANG, Q. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering* 22, 10 (2010).

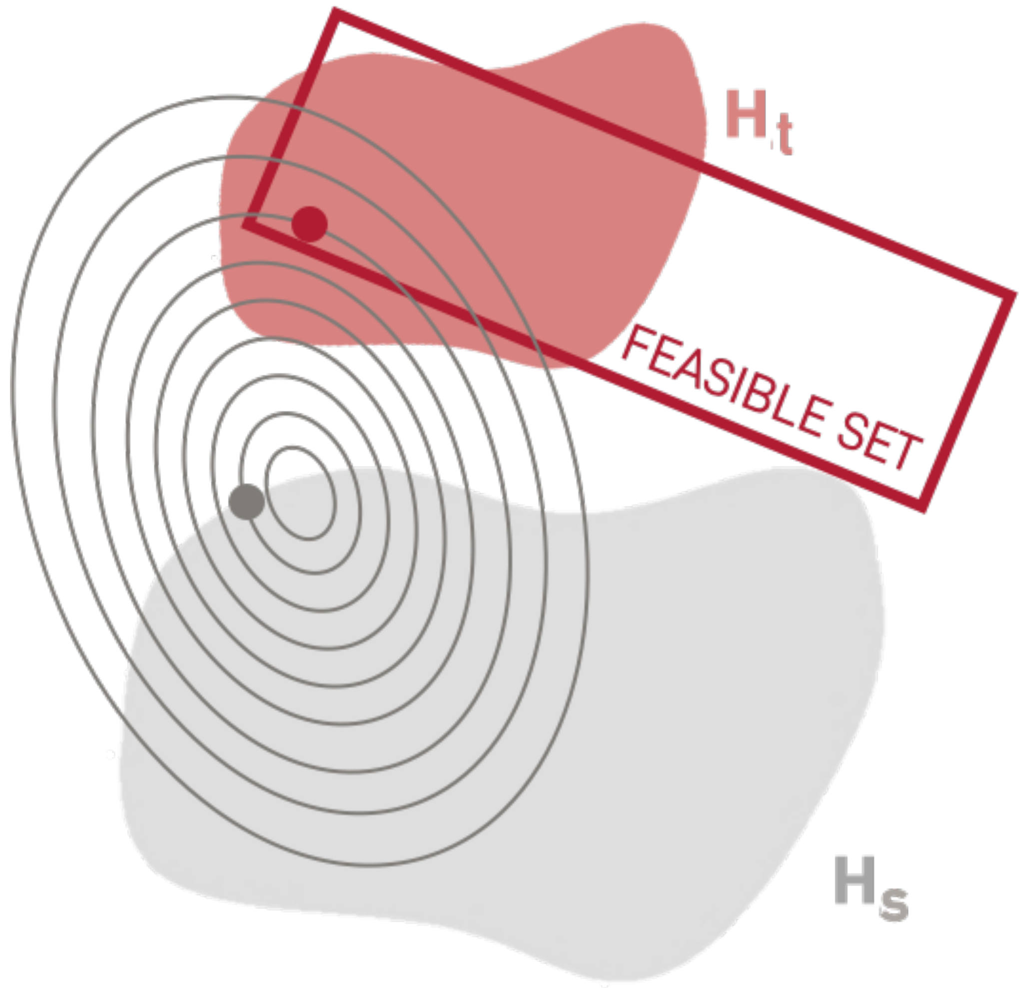
ENVIRONMENTAL ADAPTATION.

ENVIRONMENTAL ADAPTATION.

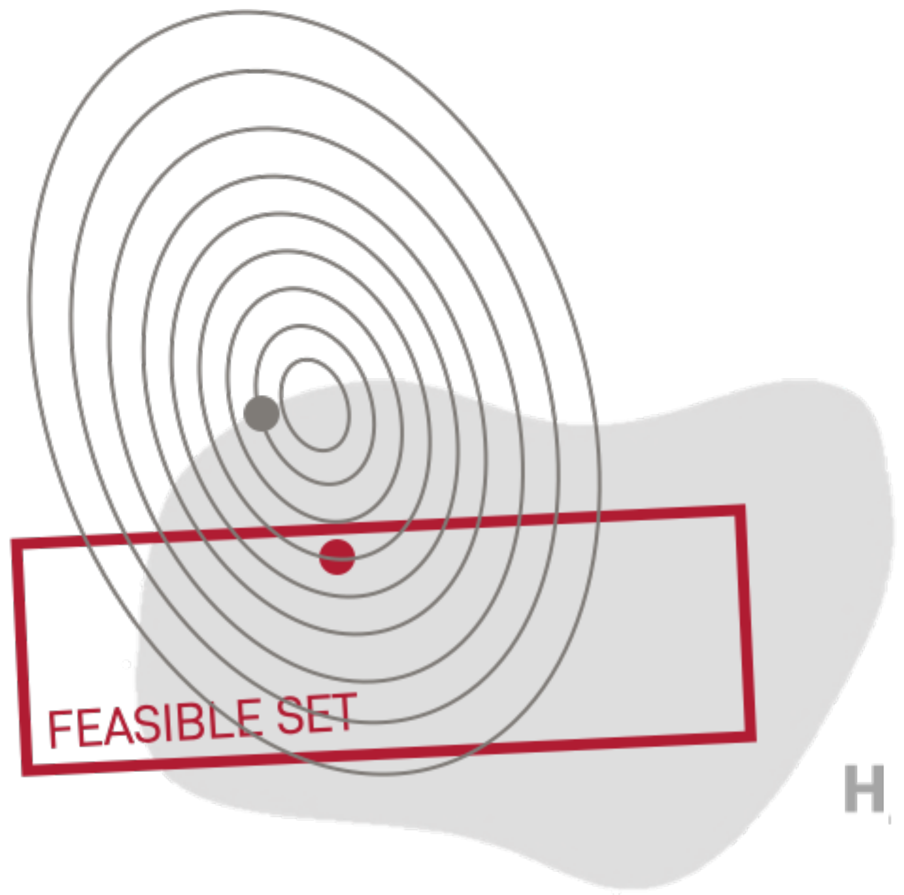


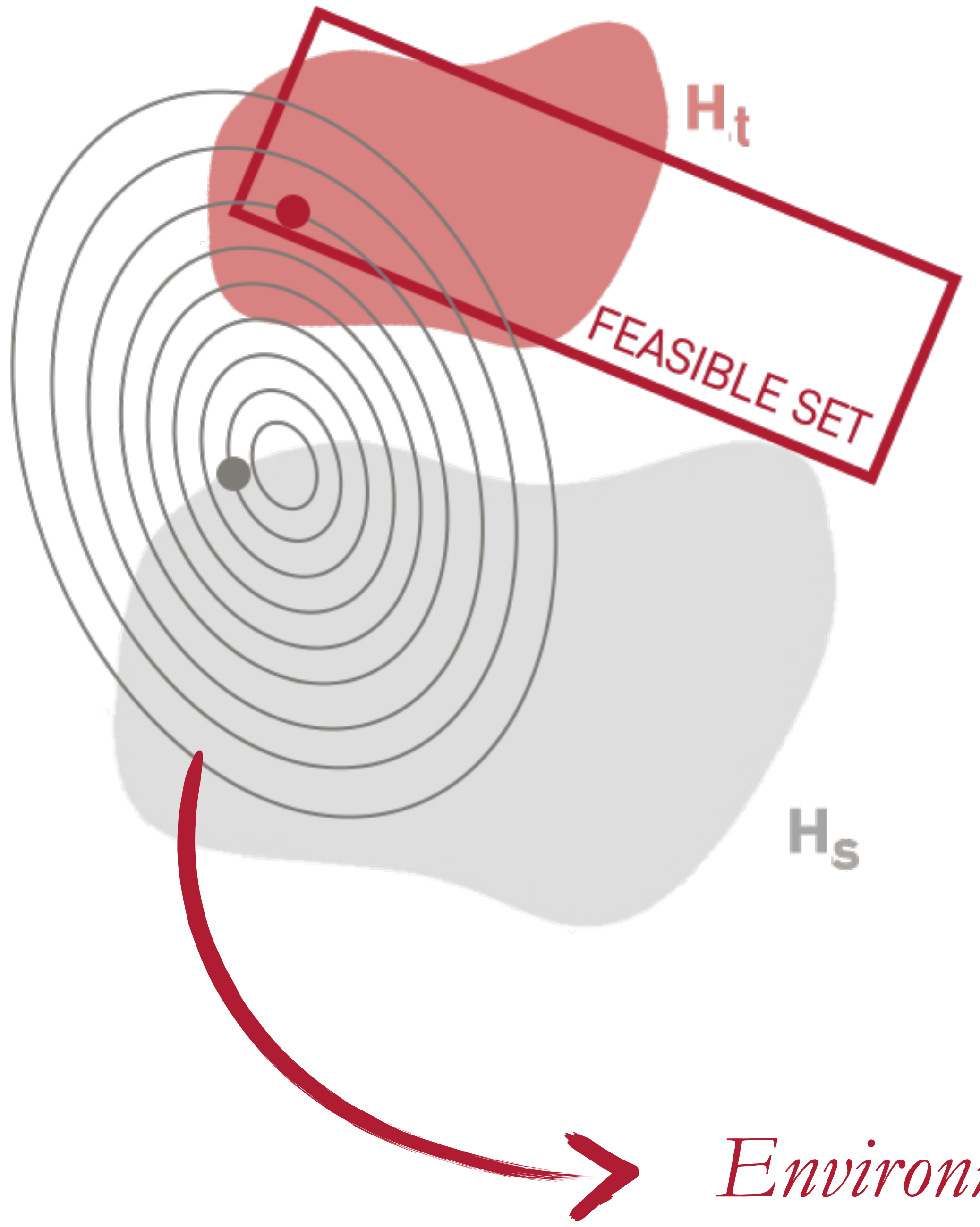
ENVIRONMENTAL ADAPTATION.





ENVIRONMENTAL ADAPTATION.





Environmental adaptation

DIFFERENTIAL REPLICATION.

The need for adaptation can be understood as a need to **transform one form of knowledge representation to another**, which we can control and which is therefore more suitable under certain circumstances.

BREIMAN, L. Statistical modeling: The two cultures. *Statistical Science* 16, 3 (2001).
BUCILUĂ, C., CARUANA, R., AND NICULESCU-MIZIL, A. Model compression. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (PA, USA, 2006).
DOMINGOS, P. Knowledge acquisition from examples via multiple models. In *Proceedings of the 14th International Conference on Machine Learning* (Miami, FL, USA, 1997).

DIFFERENTIAL REPLICATION.

The need for adaptation can be understood as a need to **transform one form of knowledge representation to another**, which we can control and which is therefore more suitable under certain circumstances.

Differential replication allows us to **reuse the knowledge** acquired by an existing model to train a second generation that can better adapt to the new environmental conditions.

BREIMAN, L. Statistical modeling: The two cultures. *Statistical Science* 16, 3 (2001).
BUCILUĂ, C., CARUANA, R., AND NICULESCU-MIZIL, A. Model compression. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (PA, USA, 2006).
DOMINGOS, P. Knowledge acquisition from examples via multiple models. In *Proceedings of the 14th International Conference on Machine Learning* (Miami, FL, USA, 1997).

Inheritance of the decision behavior

DIFFERENTIAL REPLICATION.

The need for adaptation can be understood as a need to **transform one form of knowledge representation to another**, which we can control and which is therefore more suitable under certain circumstances.

Differential replication allows us to **reuse the knowledge** acquired by an existing model to train a second generation that can better adapt to the new environmental conditions.

BREIMAN, L. Statistical modeling: The two cultures. *Statistical Science* 16, 3 (2001).
BUCILUĂ, C., CARUANA, R., AND NICULESCU-MIZIL, A. Model compression. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (PA, USA, 2006).
DOMINGOS, P. Knowledge acquisition from examples via multiple models. In *Proceedings of the 14th International Conference on Machine Learning* (Miami, FL, USA, 1997).

Inheritance of the decision behavior

DIFFERENTIAL REPLICATION.

New traits and characteristics

The need for adaptation can be understood as a need to **transform one form of knowledge representation to another**, which we can control and which is therefore more suitable under certain circumstances.

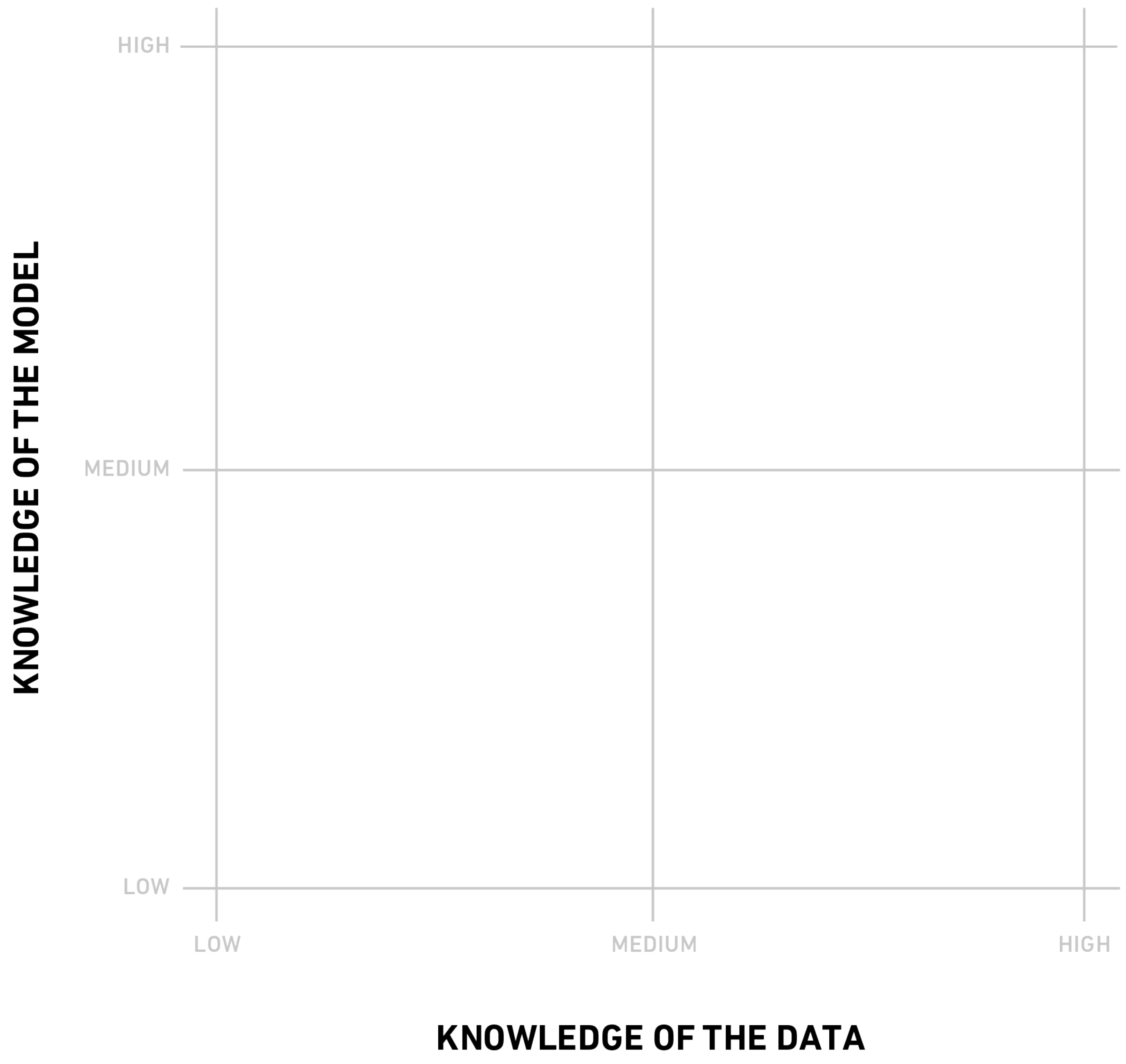
Differential replication allows us to **reuse the knowledge** acquired by an existing model to train a second generation that can better adapt to the new environmental conditions.

BREIMAN, L. Statistical modeling: The two cultures. *Statistical Science* 16, 3 (2001).

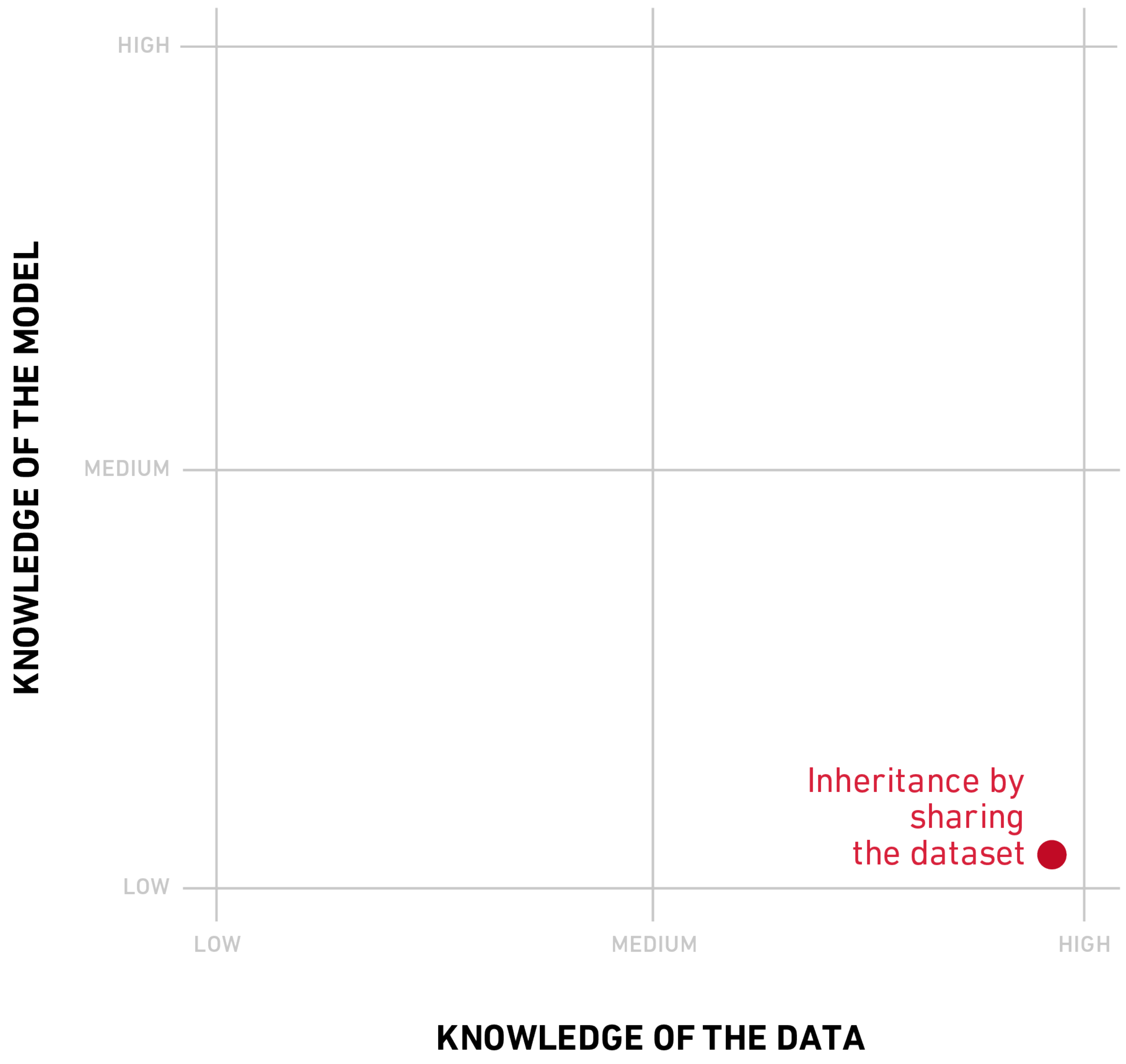
BUCILUĂ, C., CARUANA, R., AND NICULESCU-MIZIL, A. Model compression. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (PA, USA, 2006).

DOMINGOS, P. Knowledge acquisition from examples via multiple models. In *Proceedings of the 14th International Conference on Machine Learning* (Miami, FL, USA, 1997).

MECHANISMS FOR DIFFERENTIAL REPLICATION.



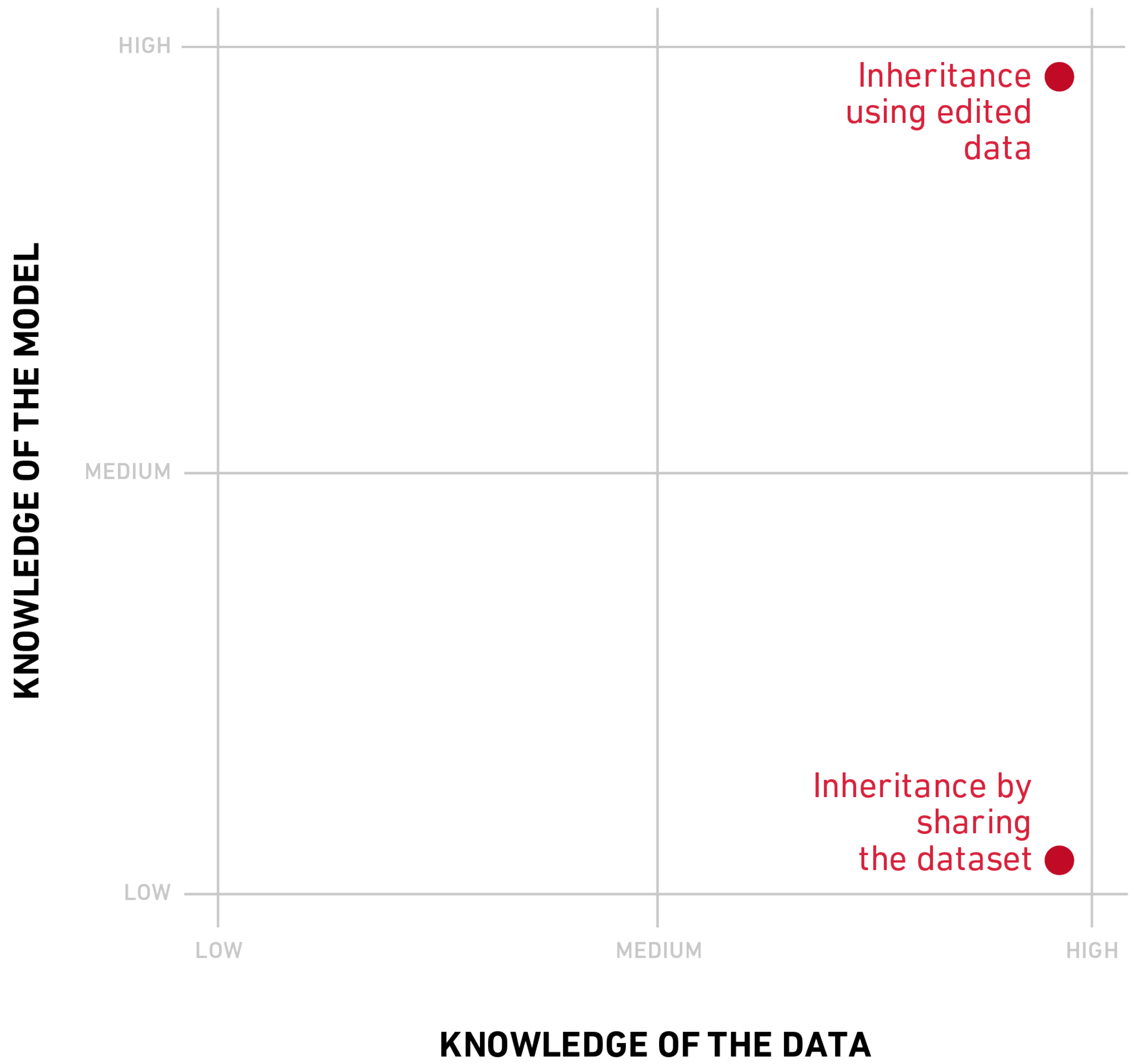
MECHANISMS FOR DIFFERENTIAL REPLICATION.



BARQUE, M., MARTIN, S., VIANIN, J., GENOUD, D., AND WANNIER, D. Improving wind power prediction with retraining machine learning algorithms. In *International Workshop on Big Data and Information Security* (Jakarta, Indonesia, 2018).

MENA, J., PUJOL, O., AND VITRIÀ, J. Uncertainty-based rejection wrappers for black-box classifiers. *IEEE Access* 8 (2020).

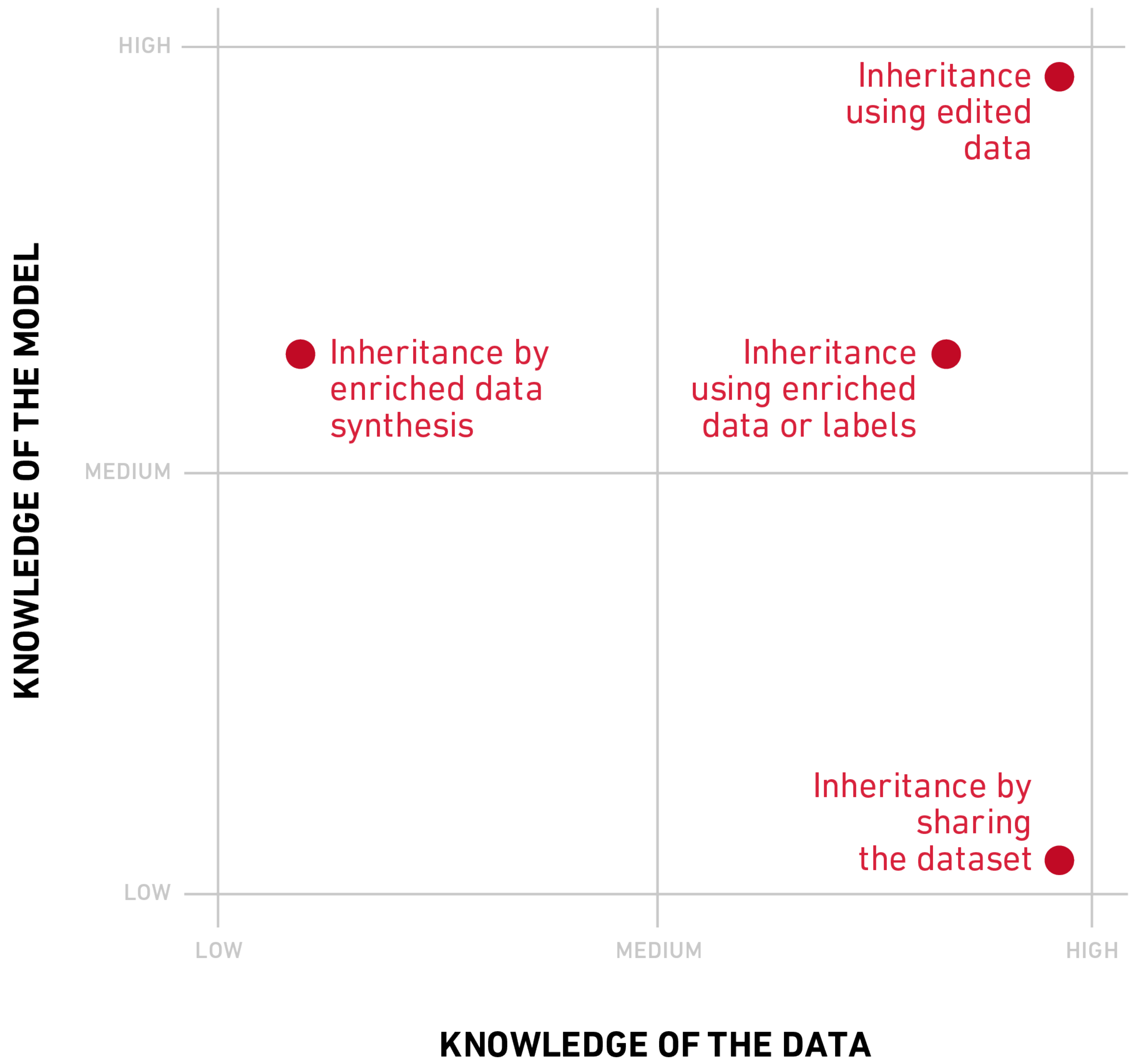
MECHANISMS FOR DIFFERENTIAL REPLICATION.



BHATTACHARYA, B., POULSEN, R., AND TOUSSAINT, G. Application of proximity graphs to editing nearest neighbor decision rule. In *International Symposium on Information Theory* (Santa Monica, CA, USA, 1981).

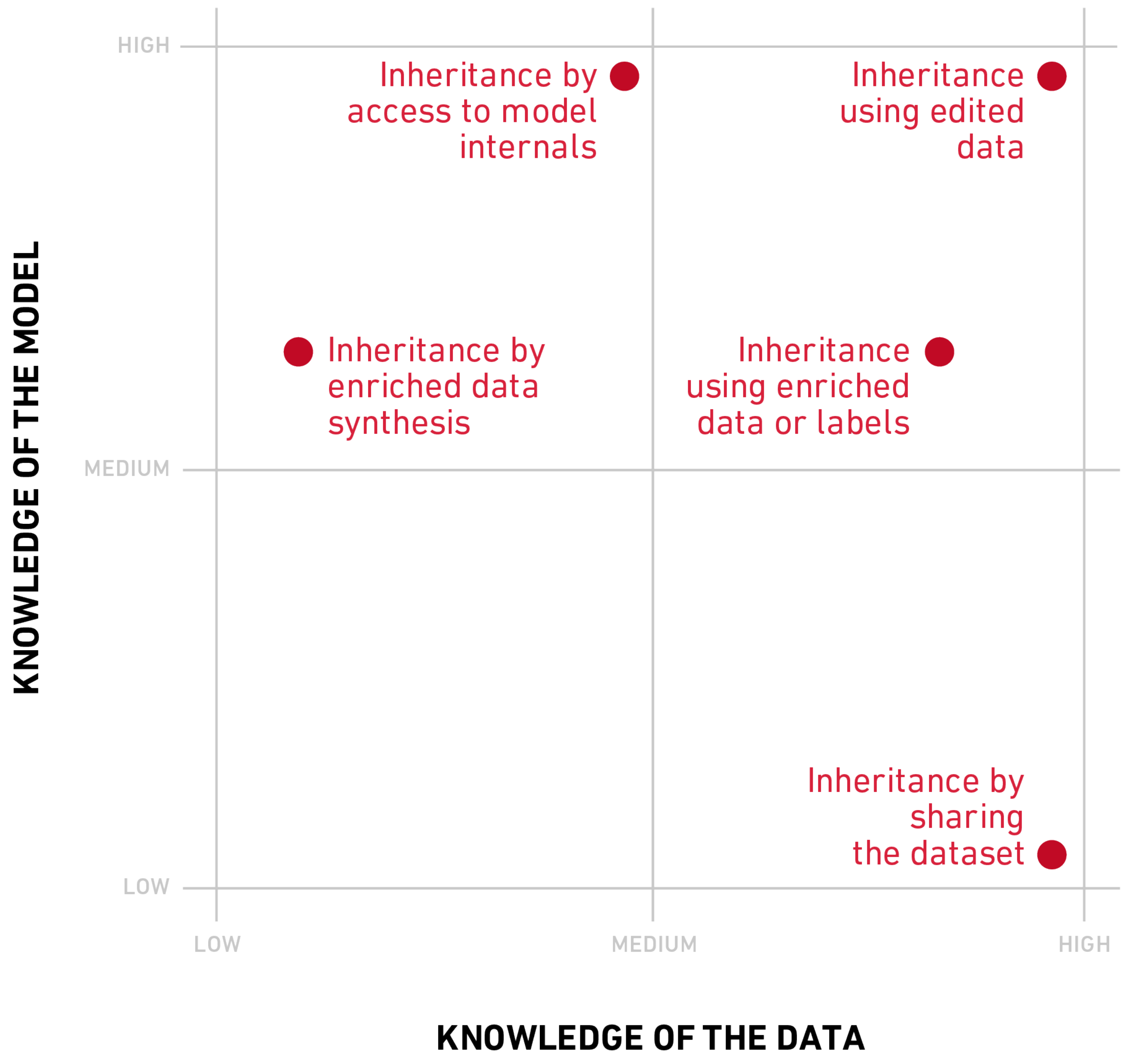
MUKHERJEE, K. Application of the gabriel graph to instance based learning. Master's thesis, Simon Fraser University (2004).

MECHANISMS FOR DIFFERENTIAL REPLICATION.



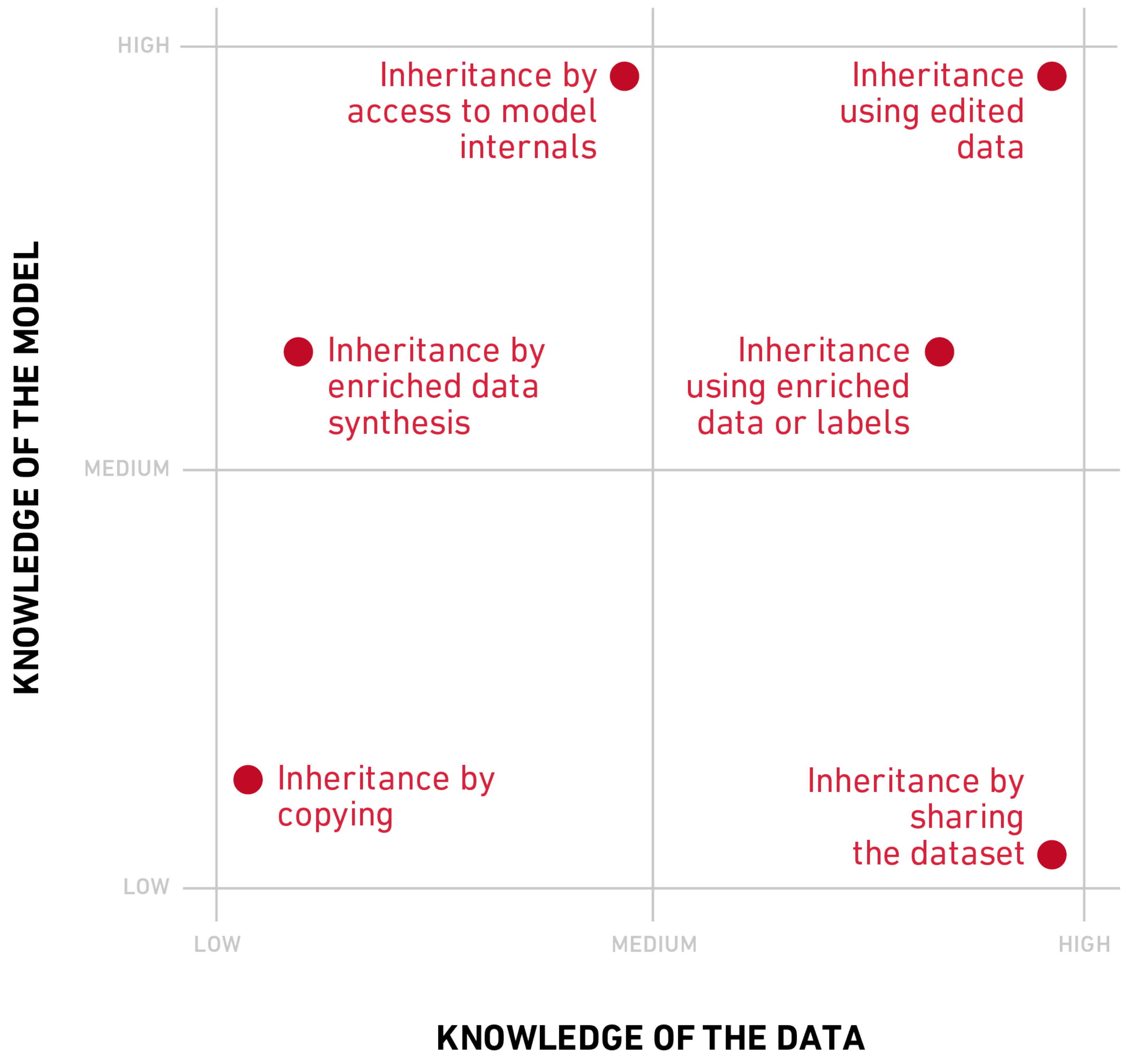
BAGHERINEZHAD, H., HORTON, M., RASTEGARI, M., AND FARHADI,A. Label refinery: Improving imagenet classification through label progression. arXiv:1805.02641 (2018).
HINTON, G., VINYALS, O., AND DEAN, J. Distilling the knowledge in a neural network. In *NIPS Deep Learning and Representation Learning Workshop* (2015).
MÜLLER, R., KORNBLITH, S., AND HINTON, G. When does label smoothing help? In *Proceedings of the 33rd Conference on Neural Information Processing Systems* (Vancouver, Canada, 2019).
NAYAK, G. K., MOPURI, K. R., SHAJ, V., BABU, R. V., AND CHAKRABORTY, A. Zero-shot knowledge distillation in deep networks.arXiv:1905.08114 (2019).

MECHANISMS FOR DIFFERENTIAL REPLICATION.



AGUILAR, G., LING, Y., ZHANG, Y., YAO, B., XING, FAN, X., AND GUO, C. Knowledge distillation from internal representations. arXiv:1910.03723 (2019).
CHENG, X., RAO, Z., CHEN, Y., AND ZHANG, Q. Explaining knowledge distillation by quantifying the knowledge. arXiv:2003.03622 (2020).

MECHANISMS FOR DIFFERENTIAL REPLICATION.



CONTENTS.

CONTENTS.

03

INHERITANCE BY COPYING

A THEORY FOR COPYING.

Model

$$f_{\mathcal{O}} : \mathcal{X} \rightarrow \mathcal{T}$$

A THEORY FOR COPYING.

Input space

Model $f_{\mathcal{O}} : \mathcal{X} \to \mathcal{T}$

Target space

A THEORY FOR COPYING.

Input space

Model $$f_\theta : \mathcal{X} \rightarrow \mathcal{T}$$

Target space

Training data $$\mathscr{D} = \{(\boldsymbol{x}_i, t_i)\}_{i=1}^{M} \mid \boldsymbol{x}_i \in \mathcal{X}, t_i \in \mathcal{T}$$

A THEORY FOR COPYING.

Input space

Model $f_{\Theta} : \mathcal{X} \to \mathcal{T}$

Target space

$\mathcal{X} = \mathbb{R}^d$

Training data $\mathscr{D} = \{(\boldsymbol{x}_i, t_i)\}_{i=1}^{M} \mid \boldsymbol{x}_i \in \mathcal{X}, t_i \in \mathcal{T}$

$\mathcal{T} = \mathbb{Z}_k$

A THEORY FOR COPYING.

Input space

Model $f_{\mathcal{O}} : \mathcal{X} \rightarrow \mathcal{T}$ *Target space*

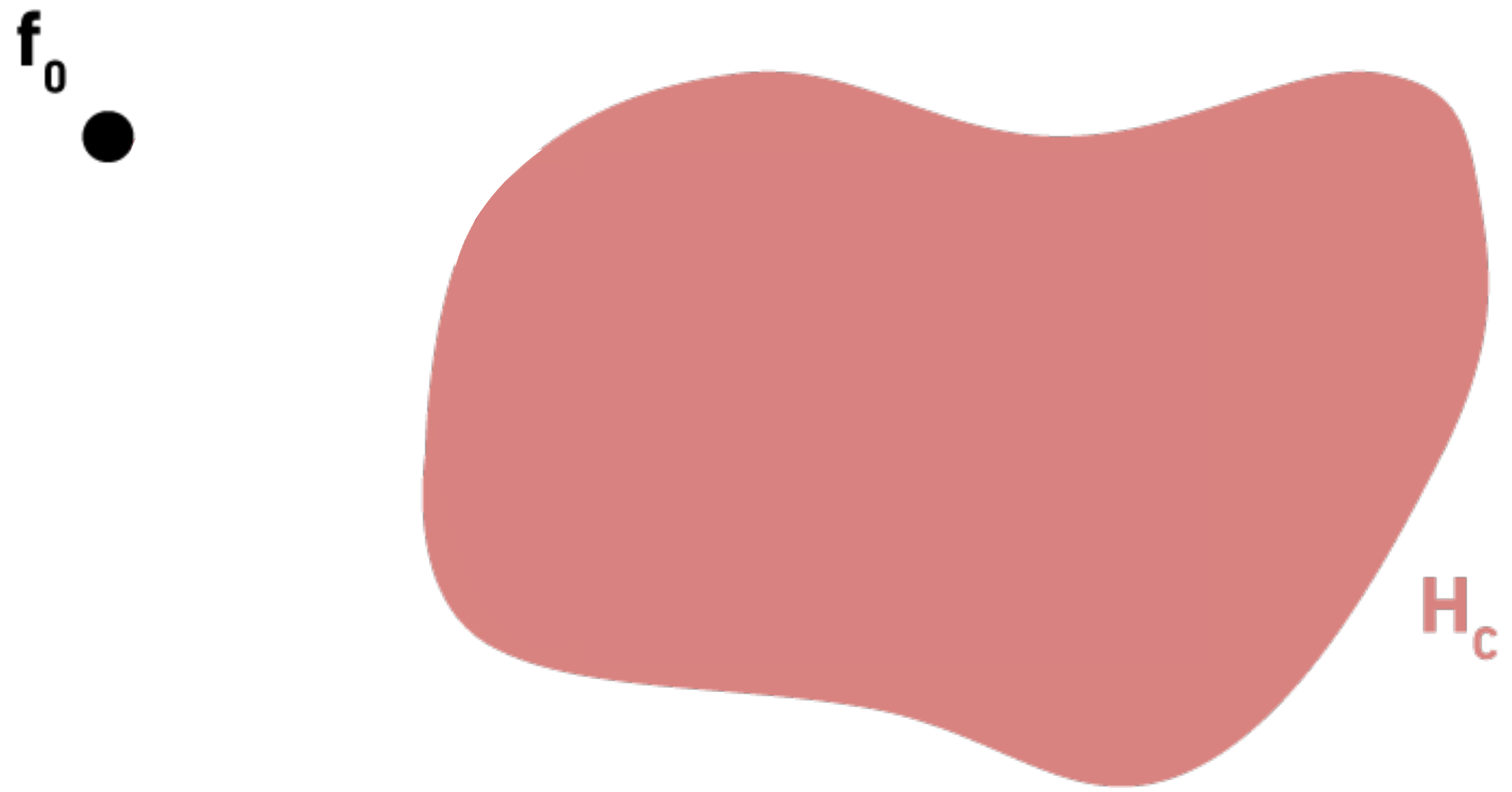
$\mathcal{X} = \mathbb{R}^d$

Training data $\mathscr{D} = \{(\boldsymbol{x}_i, t_i)\}_{i=1}^{M} \mid \boldsymbol{x}_i \in \mathcal{X}, t_i \in \mathcal{T}$

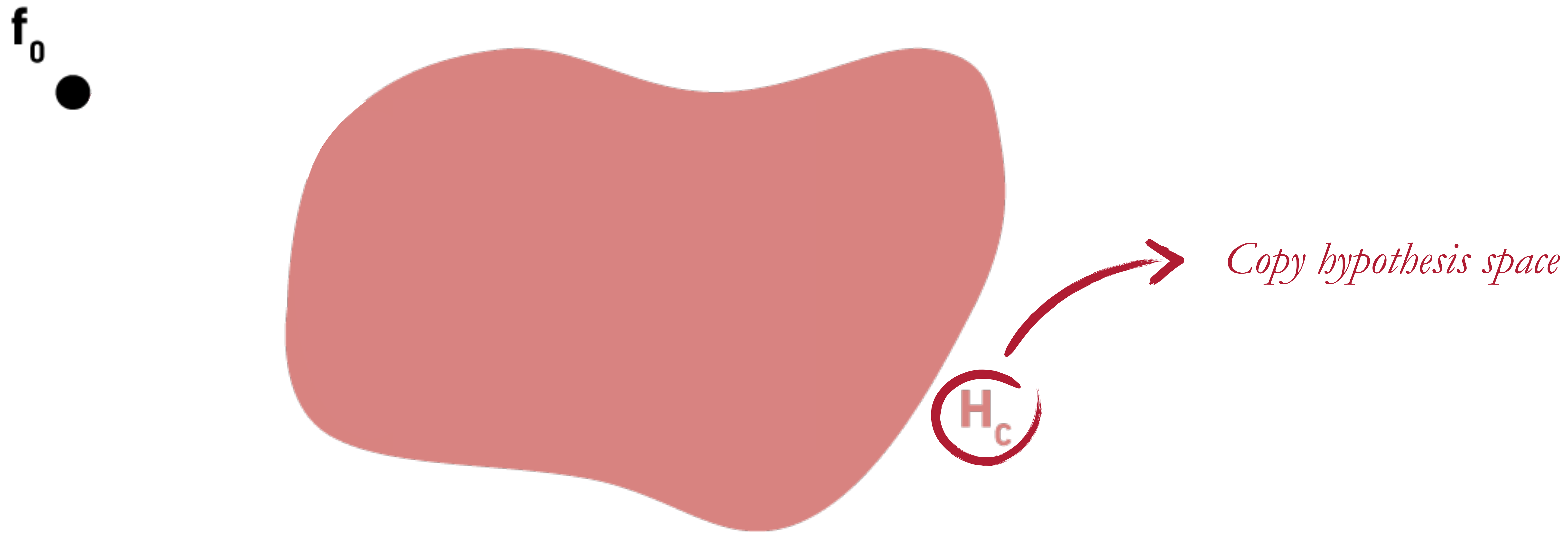
$\mathcal{T} = \mathbb{Z}_k$

Copy $f_{\mathcal{C}}(\theta) \in \mathcal{H}_{\mathcal{C}}$

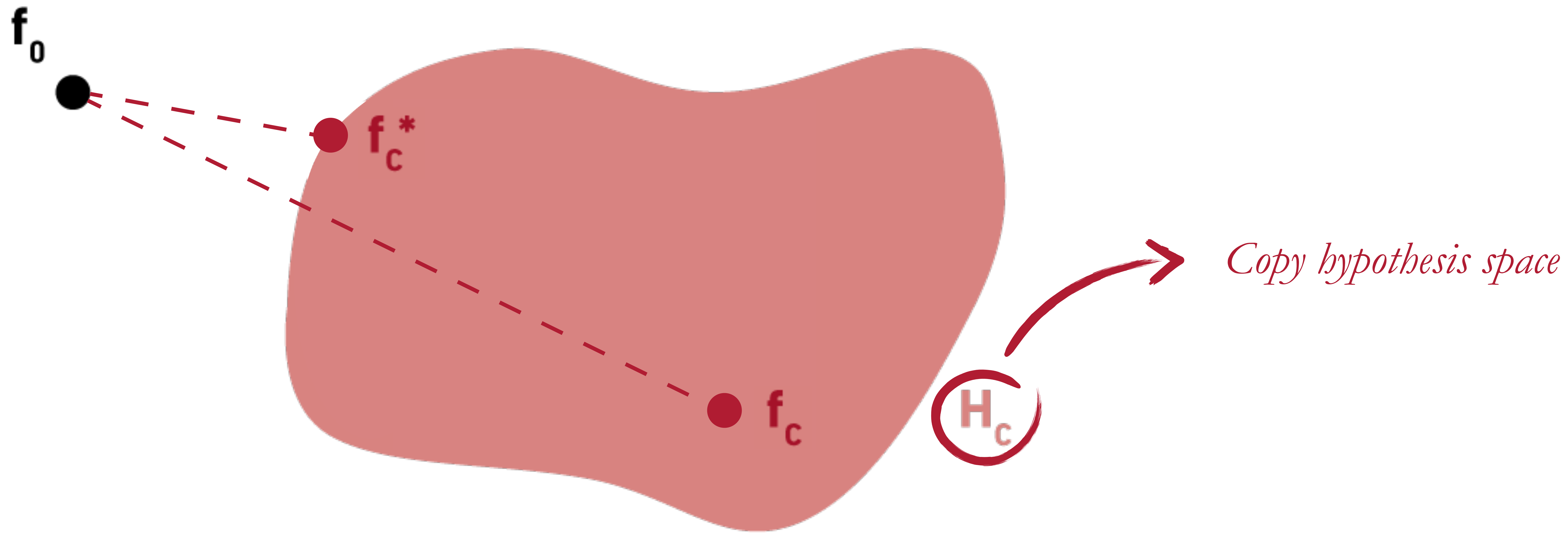
THE COPY HYPOTHESIS SPACE.



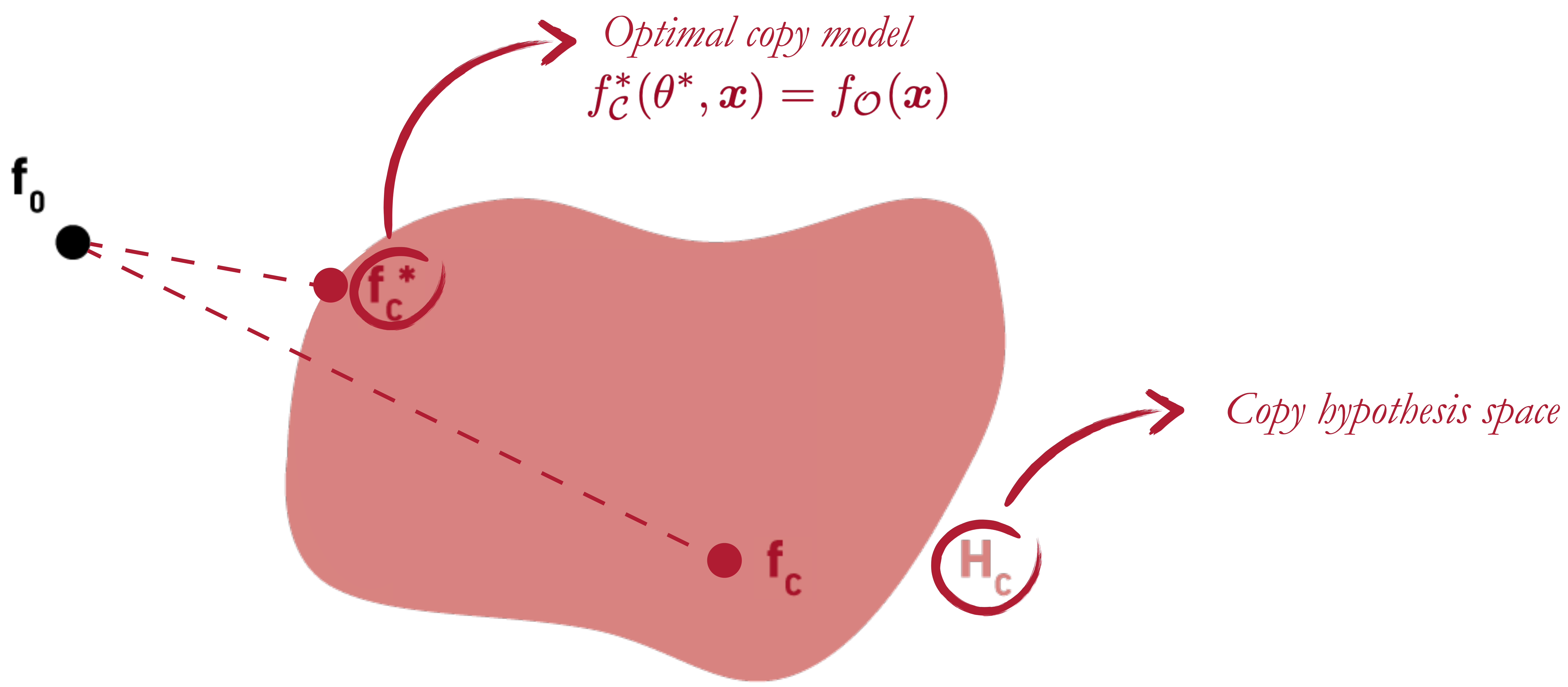
THE COPY HYPOTHESIS SPACE.



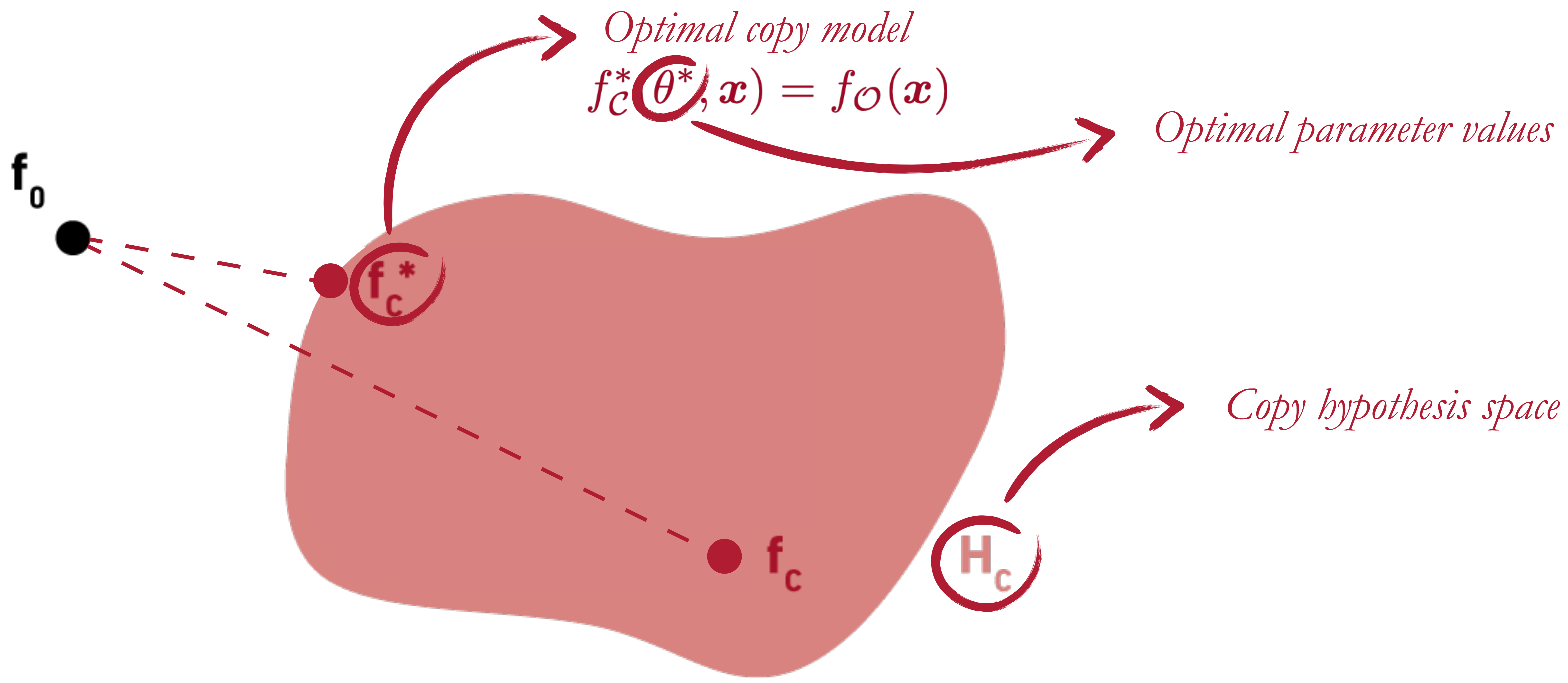
THE COPY HYPOTHESIS SPACE.



THE COPY HYPOTHESIS SPACE.



THE COPY HYPOTHESIS SPACE.



THE NEED FOR UNLABELLED DATA.

Copying problem

$$\theta^* = \arg\max_{\theta} \mathsf{P}(\theta | f_{\mathcal{O}})$$

THE NEED FOR UNLABELLED DATA.

Copying problem

We envisage a scenario where ***model internals are not open for inspection*** *and the* ***training data are unknown*** *or lost*

$$\theta^* = \arg\max_{\theta} \mathsf{P}(\theta | f_{\mathcal{O}})$$

THE NEED FOR UNLABELLED DATA.

Copying problem

$$\theta^* = \arg\max_{\theta} \mathrm{P}(\theta | f_{\mathcal{O}})$$

*We envisage a scenario where **model internals are not open for inspection** and the **training data are unknown** or lost*

Unlabelled set

$$\boldsymbol{Z} = \{\boldsymbol{z}_j\}_{j=1}^{N} \mid \boldsymbol{z}_j \in \mathcal{X}$$

THE NEED FOR UNLABELLED DATA.

Copying problem

$$\theta^* = \arg\max_\theta \int_{\boldsymbol{z}\sim P_Z} \mathsf{P}(\theta | f_{\mathcal{O}}(\boldsymbol{z})) dP_Z$$

We envisage a scenario where ***model internals are not open for inspection*** *and the* ***training data are unknown*** *or lost*

Unlabelled set

$$\boldsymbol{Z} = \{\boldsymbol{z}_j\}_{j=1}^N \mid \boldsymbol{z}_j \in \mathcal{X}$$

THE NEED FOR UNLABELLED DATA.

Copying problem

$$\theta^* = \arg\max_{\theta} \int_{z \sim P_Z} \mathsf{P}(\theta | f_{\mathcal{O}}(z)) dP_Z$$

We envisage a scenario where ***model internals are not open for inspection*** *and the* ***training data are unknown*** *or lost*

Generating probability distribution

Unlabelled set

$$\boldsymbol{Z} = \{z_j\}_{j=1}^{N} \mid z_j \in \mathcal{X}$$

COPYING UNDER EMPIRICAL RISK MINIMIZATION.

$$(\theta^*, \boldsymbol{Z}^*) = \arg \min_{\theta, \boldsymbol{z}_j \in \boldsymbol{Z}} \left[\frac{1}{N} \sum_{j=1}^{N} \gamma_1 \ell_1(f_{\mathcal{C}}(\boldsymbol{z}_j, \theta), f_{\mathcal{O}}(\boldsymbol{z}_j)) + \gamma_2 \ell_2(\theta, \theta^+) \right]$$

VAPNIK, V.N. *The Nature of Statistical Learning Theory.* Springer, Berlin, Heidelberg (2000).

COPYING UNDER EMPIRICAL RISK MINIMIZATION.

$$(\theta^*, \boldsymbol{Z}^*) = \arg \min_{\theta, \boldsymbol{z}_j \in \boldsymbol{Z}} \left[\underbrace{\frac{1}{N} \sum_{j=1}^{N} \gamma_1 \ell_1(f_{\mathcal{C}}(\boldsymbol{z}_j, \theta), f_{\mathcal{O}}(\boldsymbol{z}_j))}_{} + \gamma_2 \ell_2(\theta, \theta^+) \right]$$

Empirical fidelity error

$$R^{\mathcal{F}}_{emp}(f_{\mathcal{C}}(\boldsymbol{z}, \theta), f_{\mathcal{O}}(\boldsymbol{z}))$$

VAPNIK, V.N. *The Nature of Statistical Learning Theory.* Springer, Berlin, Heidelberg (2000).

COPYING UNDER EMPIRICAL RISK MINIMIZATION.

$$(\theta^*, \boldsymbol{Z}^*) = \arg \min_{\theta, \boldsymbol{z}_j \in \boldsymbol{Z}} \Bigg[\underbrace{\frac{1}{N} \sum_{j=1}^{N} \gamma_1 \ell_1(f_{\mathcal{C}}(\boldsymbol{z}_j, \theta), f_{\mathcal{O}}(\boldsymbol{z}_j))}_{\substack{\text{Empirical fidelity error} \\ R^{\mathcal{F}}_{emp}(f_{\mathcal{C}}(\boldsymbol{z}, \theta), f_{\mathcal{O}}(\boldsymbol{z}))}} + \underbrace{\gamma_2 \ell_2(\theta, \theta^+)}_{\substack{\text{Regularization} \\ \Omega(\theta)}} \Bigg]$$

VAPNIK, V.N. *The Nature of Statistical Learning Theory.* Springer, Berlin, Heidelberg (2000).

COPYING UNDER EMPIRICAL RISK MINIMIZATION.

$$(\theta^*, \boldsymbol{Z}^*) = \arg\min_{\theta, \boldsymbol{z}_j \in \boldsymbol{Z}} \left[\underbrace{\frac{1}{N} \sum_{j=1}^{N} \gamma_1 \ell_1(f_{\mathcal{C}}(\boldsymbol{z}_j, \theta), f_{\mathcal{O}}(\boldsymbol{z}_j))}_{\substack{\textit{Empirical fidelity error} \\ R^{\mathcal{F}}_{emp}(f_{\mathcal{C}}(\boldsymbol{z}, \theta), f_{\mathcal{O}}(\boldsymbol{z}))}} + \underbrace{\gamma_2 \ell_2(\theta, \theta^+)}_{\substack{\textit{Regularization} \\ \Omega(\theta)}} \right]$$

$$= \arg\min_{\theta, \boldsymbol{z}_j \in \boldsymbol{Z}} \left[R^{\mathcal{F}}_{emp}(f_{\mathcal{C}}(\boldsymbol{z}, \theta), f_{\mathcal{O}}(\boldsymbol{z})) + \Omega(\theta) \right]$$

VAPNIK, V.N. *The Nature of Statistical Learning Theory.* Springer, Berlin, Heidelberg (2000).

COPYING UNDER EMPIRICAL RISK MINIMIZATION.

Optimal set of synthetic samples

$$(\theta^*, \boldsymbol{Z}^*) = \arg \min_{\theta, \boldsymbol{z}_j \in \boldsymbol{Z}} \Bigg[\underbrace{\frac{1}{N} \sum_{j=1}^{N} \gamma_1 \ell_1(f_{\mathcal{C}}(\boldsymbol{z}_j, \theta), f_{\mathcal{O}}(\boldsymbol{z}_j))}_{\substack{\text{Empirical fidelity error} \\ R^{\mathcal{F}}_{emp}(f_{\mathcal{C}}(\boldsymbol{z},\theta), f_{\mathcal{O}}(\boldsymbol{z}))}} + \underbrace{\gamma_2 \ell_2(\theta, \theta^+)}_{\substack{\text{Regularization} \\ \Omega(\theta)}} \Bigg]$$

$$= \arg \min_{\theta, \boldsymbol{z}_j \in \boldsymbol{Z}} \Big[R^{\mathcal{F}}_{emp}(f_{\mathcal{C}}(\boldsymbol{z}, \theta), f_{\mathcal{O}}(\boldsymbol{z})) + \Omega(\theta) \Big]$$

VAPNIK, V.N. *The Nature of Statistical Learning Theory.* Springer, Berlin, Heidelberg (2000).

COPYING UNDER EMPIRICAL RISK MINIMIZATION.

Synthetic dataset

$\mathscr{Z}^* = \{(\boldsymbol{z}_j^*, f_{\mathcal{O}}(\boldsymbol{z}_j^*))\}_{j=1}^{N}$

Optimal set of synthetic samples

$$(\theta^*, \boldsymbol{Z}^*) = \arg\min_{\theta, \boldsymbol{z}_j \in \boldsymbol{Z}} \left[\frac{1}{N} \sum_{j=1}^{N} \gamma_1 \ell_1(f_{\mathcal{C}}(\boldsymbol{z}_j, \theta), f_{\mathcal{O}}(\boldsymbol{z}_j)) + \gamma_2 \ell_2(\theta, \theta^+) \right]$$

Empirical fidelity error $R^{\mathcal{F}}_{emp}(f_{\mathcal{C}}(\boldsymbol{z}, \theta), f_{\mathcal{O}}(\boldsymbol{z}))$

Regularization $\Omega(\theta)$

$$= \arg\min_{\theta, \boldsymbol{z}_j \in \boldsymbol{Z}} \left[R^{\mathcal{F}}_{emp}(f_{\mathcal{C}}(\boldsymbol{z}, \theta), f_{\mathcal{O}}(\boldsymbol{z})) + \Omega(\theta) \right]$$

VAPNIK, V.N. *The Nature of Statistical Learning Theory.* Springer, Berlin, Heidelberg (2000).

SOLVING THE COPYING PROBLEM.

SOLVING THE COPYING PROBLEM.

1 The problem is always **separable**

SOLVING THE COPYING PROBLEM.

1. The problem is always **separable**
2. We can potentially generate **infinite samples**

SOLVING THE COPYING PROBLEM.

SOLVING THE COPYING PROBLEM.

Unconstrained problem

$$\underset{\theta, \boldsymbol{Z}}{\text{minimize}} \quad R^{\mathcal{F}}_{emp}(f_{\mathcal{C}}(\boldsymbol{z}, \theta), f_{\mathcal{O}}(\boldsymbol{z}))$$

SOLVING THE COPYING PROBLEM.

Unconstrained problem

$$\underset{\theta, \boldsymbol{Z}}{\text{minimize}} \quad R^{\mathcal{F}}_{emp}(f_{\mathcal{C}}(\boldsymbol{z}, \theta), f_{\mathcal{O}}(\boldsymbol{z}))$$

Constrained problem

$$\begin{aligned} \underset{\theta, \boldsymbol{Z}}{\text{minimize}} \quad & \Omega(\theta) \\ \text{subject to} \quad & \| R^{\mathcal{F}}_{emp}(f_{\mathcal{C}}(\boldsymbol{z}, \theta), f_{\mathcal{O}}(\boldsymbol{z})) - R^{\mathcal{F}}_{emp}(f^{\dagger}_{\mathcal{C}}(\boldsymbol{z}, \theta), f_{\mathcal{O}}(\boldsymbol{z})) \| < \epsilon \end{aligned}$$

SOLVING THE COPYING PROBLEM.

Unconstrained problem

$$\underset{\theta, \boldsymbol{Z}}{\text{minimize}} \quad R_{emp}^{\mathcal{F}}(f_{\mathcal{C}}(\boldsymbol{z}, \theta), f_{\mathcal{O}}(\boldsymbol{z}))$$

Constrained problem

$$\begin{aligned} \underset{\theta, \boldsymbol{Z}}{\text{minimize}} \quad & \Omega(\theta) \\ \text{subject to} \quad & \| R_{emp}^{\mathcal{F}}(f_{\mathcal{C}}(\boldsymbol{z}, \theta), f_{\mathcal{O}}(\boldsymbol{z})) - R_{emp}^{\mathcal{F}}(f_{\mathcal{C}}^{\dagger}(\boldsymbol{z}, \theta), f_{\mathcal{O}}(\boldsymbol{z})) \| < \epsilon \end{aligned}$$

SOLVING THE COPYING PROBLEM.

Unconstrained problem

$$\underset{\theta, \boldsymbol{Z}}{\text{minimize}} \quad R^{\mathcal{F}}_{emp}(f_{\mathcal{C}}(\boldsymbol{z}, \theta), f_{\mathcal{O}}(\boldsymbol{z}))$$

Constrained problem

$$\begin{aligned} \underset{\theta, \boldsymbol{Z}}{\text{minimize}} \quad & \Omega(\theta) \\ \text{subject to} \quad & \| R^{\mathcal{F}}_{emp}(f_{\mathcal{C}}(\boldsymbol{z}, \theta), f_{\mathcal{O}}(\boldsymbol{z})) - R^{\mathcal{F}}_{emp}(f^{\dagger}_{\mathcal{C}}(\boldsymbol{z}, \theta), f_{\mathcal{O}}(\boldsymbol{z})) \| < \epsilon \end{aligned}$$

Tolerance

SOLVING THE COPYING PROBLEM.

Unconstrained problem

$$\underset{\theta, \boldsymbol{Z}}{\text{minimize}} \quad R_{emp}^{\mathcal{F}}(f_{\mathcal{C}}(\boldsymbol{z}, \theta), f_{\mathcal{O}}(\boldsymbol{z}))$$

Constrained problem

$$\begin{aligned} \underset{\theta, \boldsymbol{Z}}{\text{minimize}} \quad & \Omega(\theta) \\ \text{subject to} \quad & \| R_{emp}^{\mathcal{F}}(f_{\mathcal{C}}(\boldsymbol{z}, \theta), f_{\mathcal{O}}(\boldsymbol{z})) - R_{emp}^{\mathcal{F}}(f_{\mathcal{C}}^{\dagger}(\boldsymbol{z}, \theta), f_{\mathcal{O}}(\boldsymbol{z})) \| < \epsilon \end{aligned}$$

Tolerance

SOLVING THE COPYING PROBLEM.

SOLVING THE COPYING PROBLEM.

Model

SOLVING THE COPYING PROBLEM.

Model

Synthetic dataset

SOLVING THE COPYING PROBLEM.

Model

Copy

Synthetic dataset

SOLVING THE COPYING PROBLEM.

Model

Copy

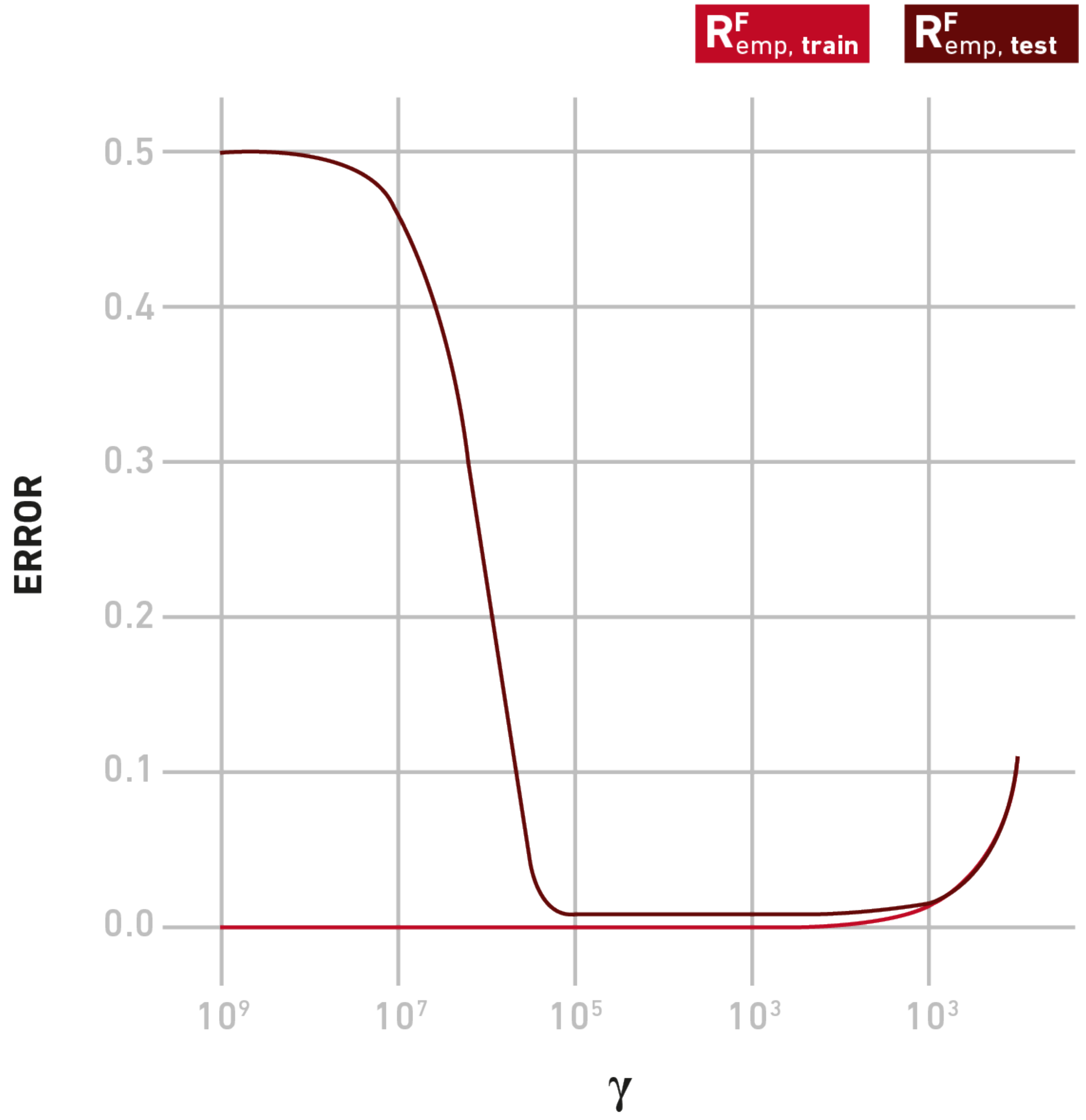
Synthetic dataset

SOLVING THE COPYING PROBLEM.

Model

Copy

Synthetic dataset



THE SINGLE-PASS APPROACH.

THE SINGLE-PASS APPROACH.

1 Finding the optimal set of **synthetic samples**

$$\boldsymbol{Z}^* = \arg\min_{\boldsymbol{Z}} R^{\mathcal{F}}_{emp}(f_{\mathcal{C}}(\boldsymbol{z}, \theta), f_{\mathcal{O}}(\boldsymbol{z}))$$

THE SINGLE-PASS APPROACH.

1 Finding the optimal set of **synthetic samples**

$$\boldsymbol{Z}^* = \arg\min_{\boldsymbol{Z}} R^{\mathcal{F}}_{emp}(f_{\mathcal{C}}(\boldsymbol{z}, \theta), f_{\mathcal{O}}(\boldsymbol{z}))$$

2 Optimizing the copy **parameters**

$$\begin{aligned} \underset{\theta}{\text{minimize}} \quad & \Omega(\theta) \\ \text{subject to} \quad & \| R^{\mathcal{F}}_{emp}(f_{\mathcal{C}}(\boldsymbol{z}, \theta), f_{\mathcal{O}}(\boldsymbol{z})) - R^{\mathcal{F}}_{emp}(f^{\dagger}_{\mathcal{C}}(\boldsymbol{z}, \theta), f_{\mathcal{O}}(\boldsymbol{z})) \| < \epsilon \end{aligned}$$

CONTENTS.

CONTENTS.

04

EMPIRICAL VALIDATION

THE CONTEXT.

THE CONTEXT.

Credit default prediction for non-client mortgage loans

Credit default prediction for non-client mortgage loans

THE CONTEXT.

Credit default has **significant cost implications** for financial institutions.

DOCHERTY, A., AND VIORT, F. *Better Banking: Understanding and Addressing the Failures in Risk Management, Governance and Regulation*. John Wiley & Sons, Ltd, New Jersey, NJ, USA (2013).

Credit default prediction for non-client mortgage loans

THE CONTEXT.

Credit default has **significant cost implications** for financial institutions.

Increasing efforts are devoted to **develop complex models** able to learn this problem.

RUDIN, C. Please stop explaining black box models for high-stakes decisions. In *Workshop on Critiquing and Correcting Trends in Machine Learning* (Montreal, Canada, 2018).
S&P DOW JONES INDICES. S&P EXPERIAN CONSUMER CREDIT DEFAULT INDICES SHOW DEFAULT RATES STABLE IN AUGUST 2018. TECH. REP. (2018).

Credit default prediction for non-client mortgage loans

THE CONTEXT.

Credit default has **significant cost implications** for financial institutions.

Increasing efforts are devoted to **develop complex models** able to learn this problem.

However, credit scoring models are required by law to be **interpretable**.

E. U. COMMISION. Legislation. *OJ* (2016).
GOODMAN, B., AND FLAXMAN, S. European union regulations on algorithmic decision-making and a right to explanation. *AI Magazine* 38, 3 (2017).

Credit default prediction for non-client mortgage loans

THE CONTEXT.

Credit default has **significant cost implications** for financial institutions.

Increasing efforts are devoted to **develop complex models** able to learn this problem.

However, credit scoring models are required by law to be **interpretable**.

In this context, **logistic regression models** are widely established.

THE CONTEXT.

Credit default prediction for non-client mortgage loans

Credit default has **significant cost implications** for financial institutions.

Increasing efforts are devoted to **develop complex models** able to learn this problem.

However, credit scoring models are required by law to be **interpretable**.

In this context, **logistic regression models** are widely established.

These models requires a sophisticated **pre-processing** to account for non-linear effects in the data.

THE CONTEXT.

Credit default prediction for non-client mortgage loans

Credit default has **significant cost implications** for financial institutions.

Increasing efforts are devoted to **develop complex models** able to learn this problem.

However, credit scoring models are required by law to be **interpretable**.

In this context, **logistic regression models** are widely established.

These models requires a sophisticated **pre-processing** to account for non-linear effects in the data.

Credit default prediction for non-client mortgage loans

THE CONTEXT.

Credit default has **significant cost implications** for financial institutions.

Increasing efforts are devoted to **develop complex models** able to learn this problem.

However, credit scoring models are required by law to be **interpretable**.

In this context, **logistic regression models** are widely established.

These models requires a sophisticated **pre-processing** to account for non-linear effects in the data.

Non-decomposability

Credit default prediction for non-client mortgage loans

THE CONTEXT.

Credit default has **significant cost implications** for financial institutions.

Increasing efforts are devoted to **develop complex models** able to learn this problem.

However, credit scoring models are required by law to be **interpretable**.

In this context, **logistic regression models** are widely established.

These models requires a sophisticated **pre-processing** to account for non-linear effects in the data.

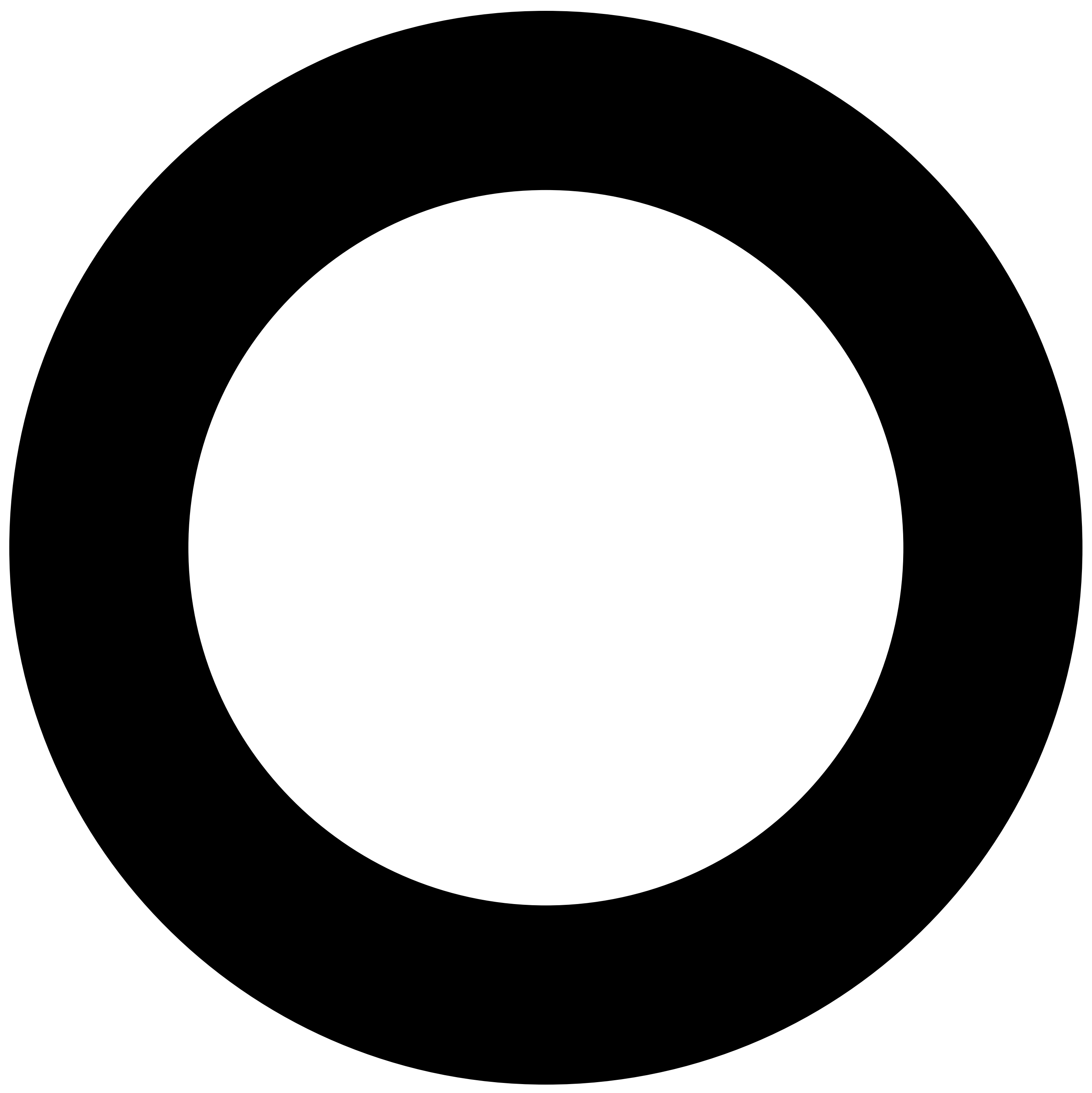
Non-decomposability

Increased time-to-market delivery

THE DATA.

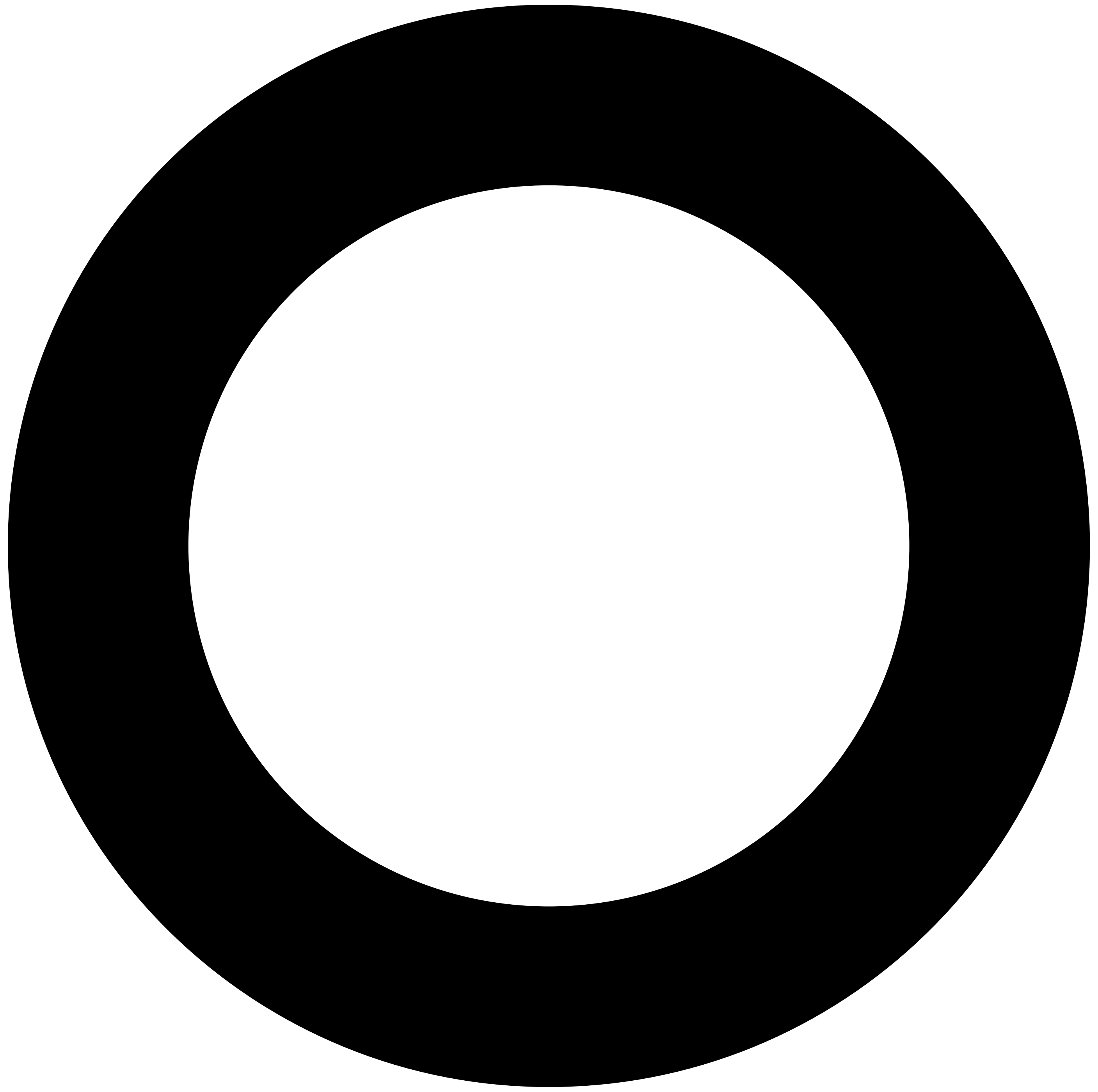
THE DATA.

Non-client mortgage loan applications



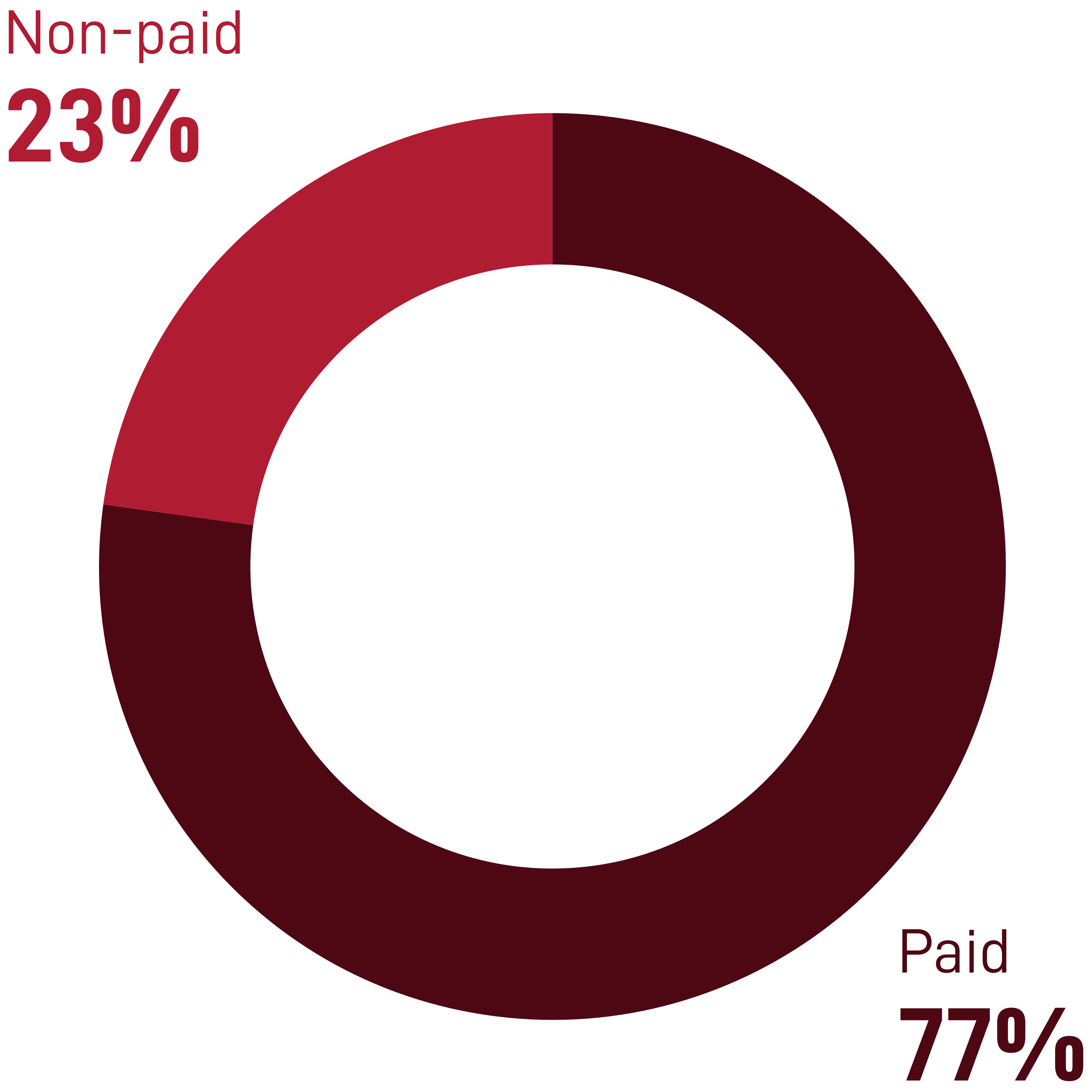
THE DATA.

Non-client mortgage loan applications

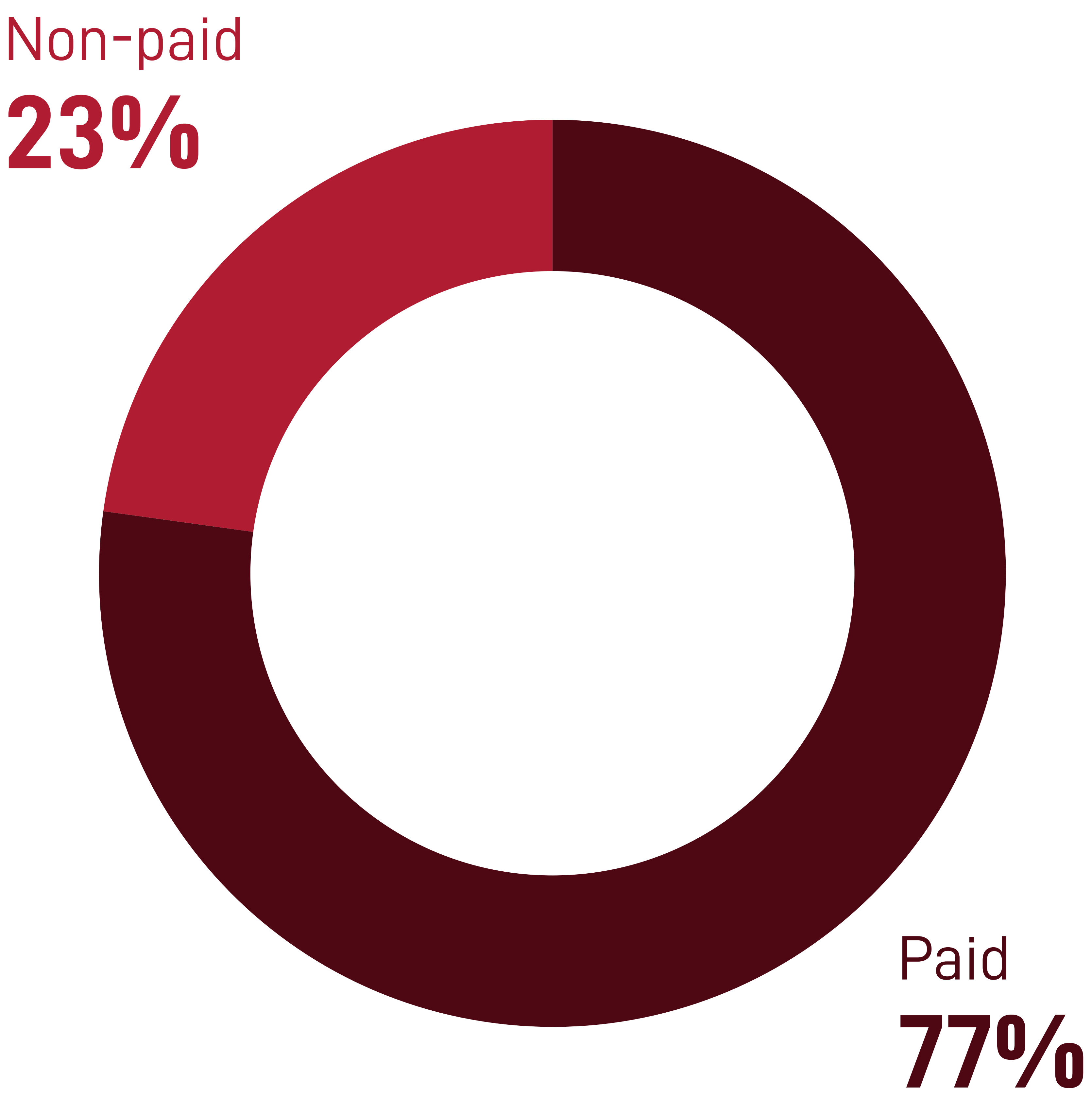


No previous active contract with the bank at the time of loan application

THE DATA.



THE DATA.



Attribute	Description
age	Age
studies	Level of studies
n_family_unit	Members of the family unit
zip_code	Municipality
municipality	ZIP code
indebtedness	Level of indebtedness
p_default	Ratio of defaulted contracts
economy_level	Level of economy
est_income	Estimated income
est_soc_income	Estimated socio-demographic income
est_mila_income	Estimated income based on MILA model
poverty_index	Marginalization / poverty index
credit_amount	Amount of credit
property_value	Property value
value_m2	Value per square meter
loan_to_value	Loan to value
duration	Duration of the loan
installment	Monthly installment

THE SCENARIOS.

THE SCENARIOS.

SCENARIO 1 Deobfuscation of the attribute pre-processing

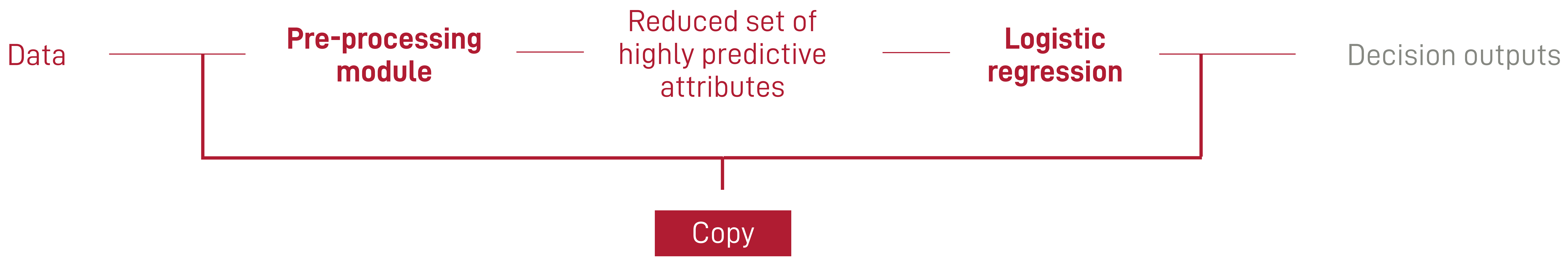
THE SCENARIOS.

SCENARIO 1 Deobfuscation of the attribute pre-processing

Data — **Pre-processing module** — Reduced set of highly predictive attributes — **Logistic regression** — Decision outputs

THE SCENARIOS.

SCENARIO 1 Deobfuscation of the attribute pre-processing



THE SCENARIOS.

SCENARIO 2 Fast regulatory-compliant model building

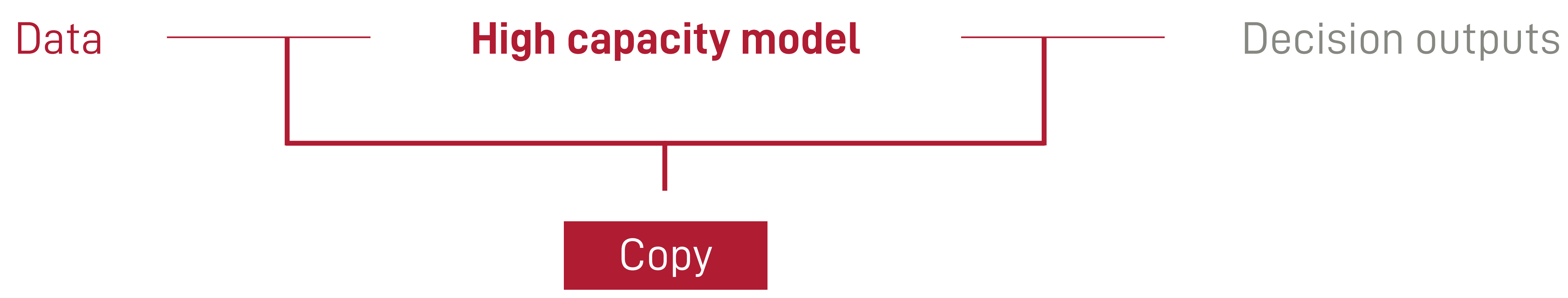
THE SCENARIOS.

SCENARIO 2 Fast regulatory-compliant model building

Data —— **High capacity model** —— Decision outputs

THE SCENARIOS.

SCENARIO 2 Fast regulatory-compliant model building



THE PERFORMANCE METRICS.

THE PERFORMANCE METRICS.

Empirical fidelity error

$$R_{emp}^{\mathcal{F},\mathscr{L}} = \frac{1}{N}\sum_{j=1}^{N}\mathbb{I}[f_{\mathcal{O}}(\boldsymbol{z}_j) \neq f_{\mathcal{C}}(\boldsymbol{z}_j)]$$

THE PERFORMANCE METRICS.

Empirical fidelity error

$$R_{emp}^{\mathcal{F},\mathscr{L}} = \frac{1}{N}\sum_{j=1}^{N} \mathbb{I}[f_{\mathcal{O}}(\boldsymbol{z}_j) \neq f_{\mathcal{C}}(\boldsymbol{z}_j)]$$

Copy accuracy

$$\mathcal{A}_{\mathcal{C}} = \frac{1}{M}\sum_{i=1}^{M} \mathbb{I}[t_i = f_{\mathcal{C}}(\boldsymbol{x}_i)]$$

THE PERFORMANCE METRICS.

Empirical fidelity error

$$R_{emp}^{\mathcal{F},\mathscr{L}} = \frac{1}{N}\sum_{j=1}^{N} \mathbb{I}[f_{\mathcal{O}}(\boldsymbol{z}_j) \neq f_{\mathcal{C}}(\boldsymbol{z}_j)]$$

Copy accuracy

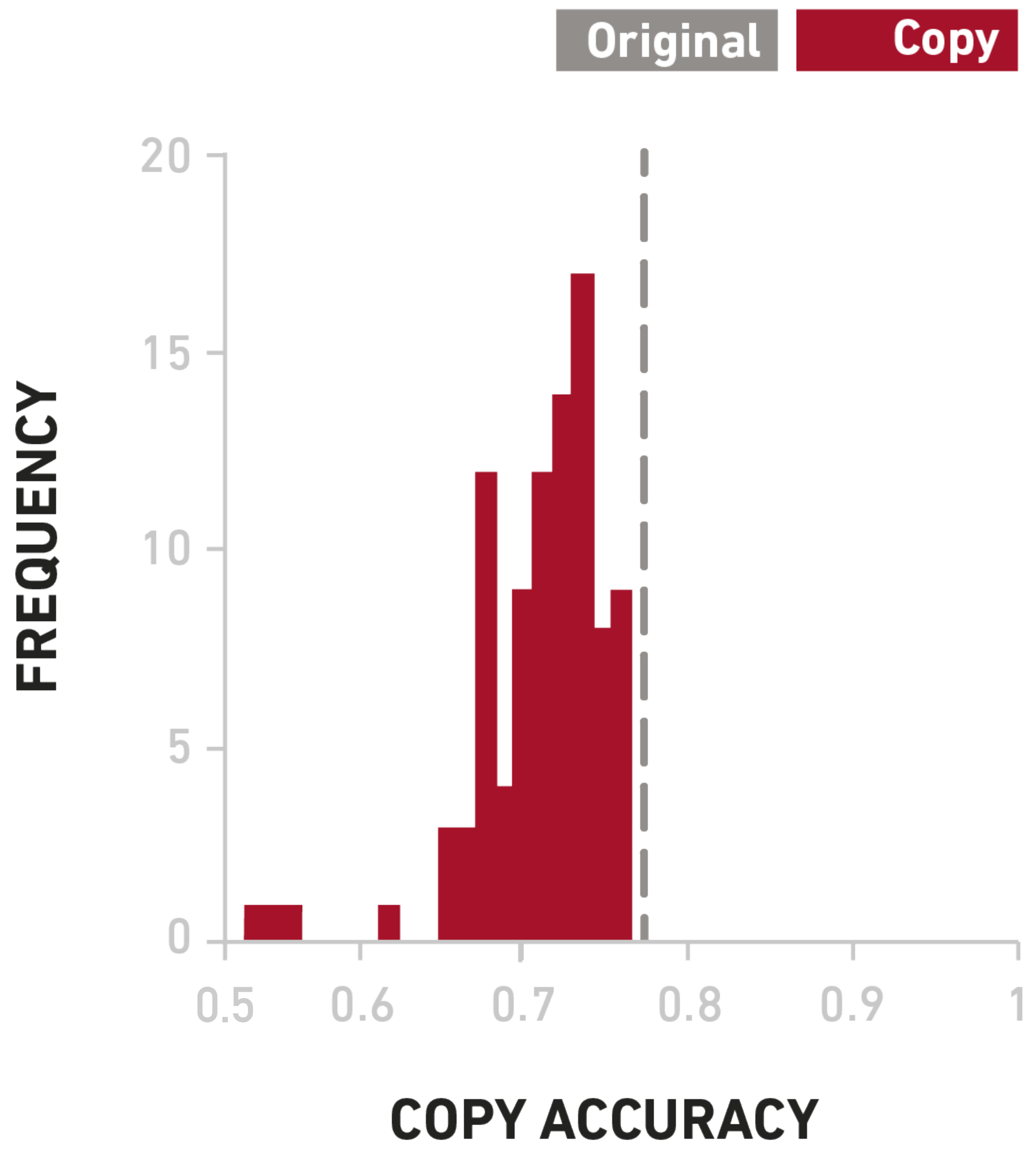
$$\mathcal{A}_{\mathcal{C}} = \frac{1}{M}\sum_{i=1}^{M} \mathbb{I}[t_i = f_{\mathcal{C}}(\boldsymbol{x}_i)]$$

THE RESULTS.

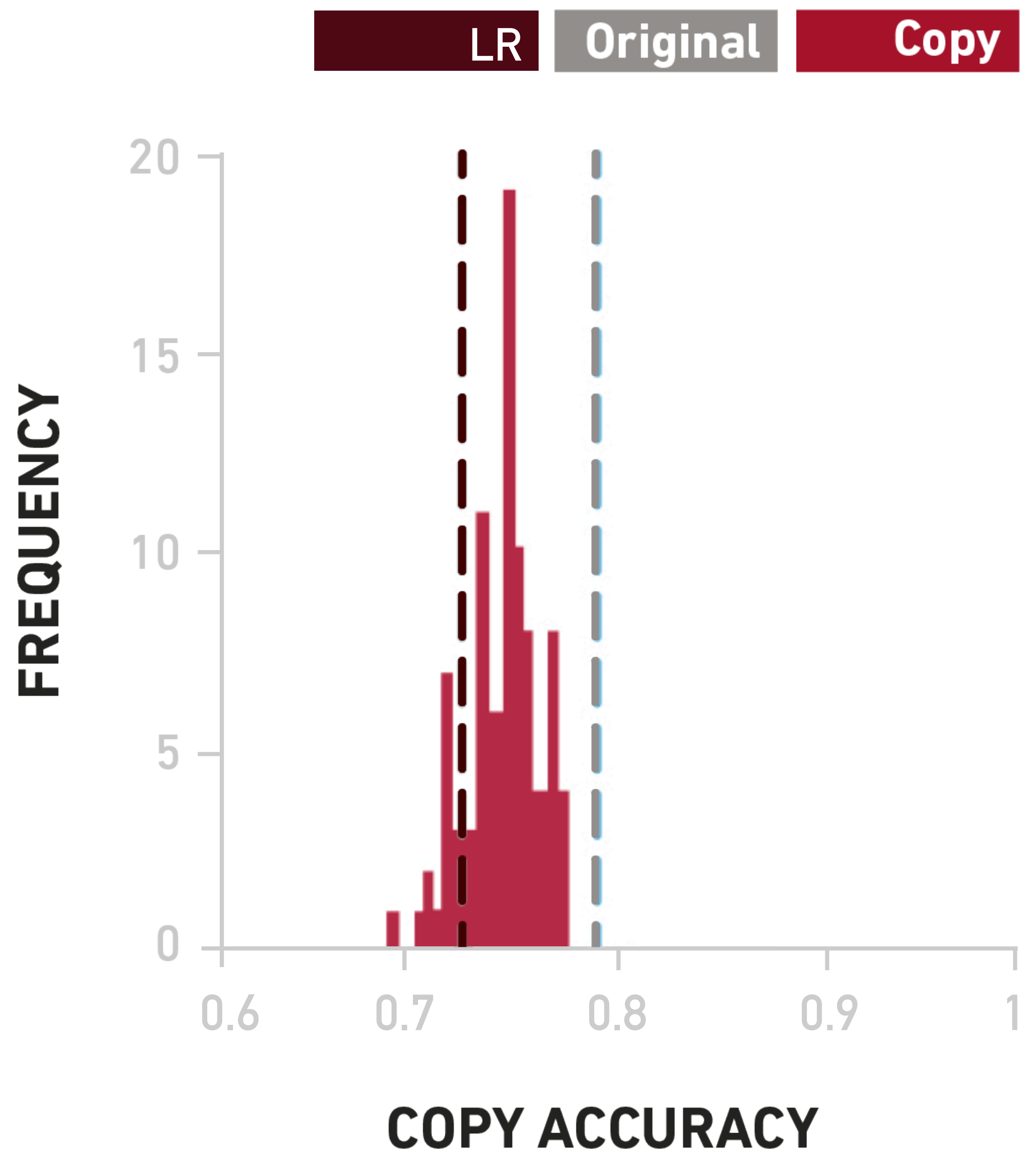
0.5

THE RESULTS.

SCENARIO 1



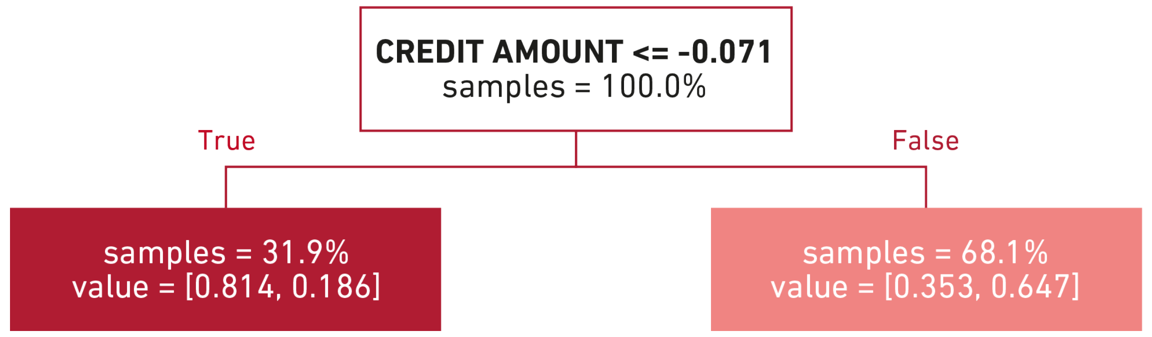
SCENARIO 2



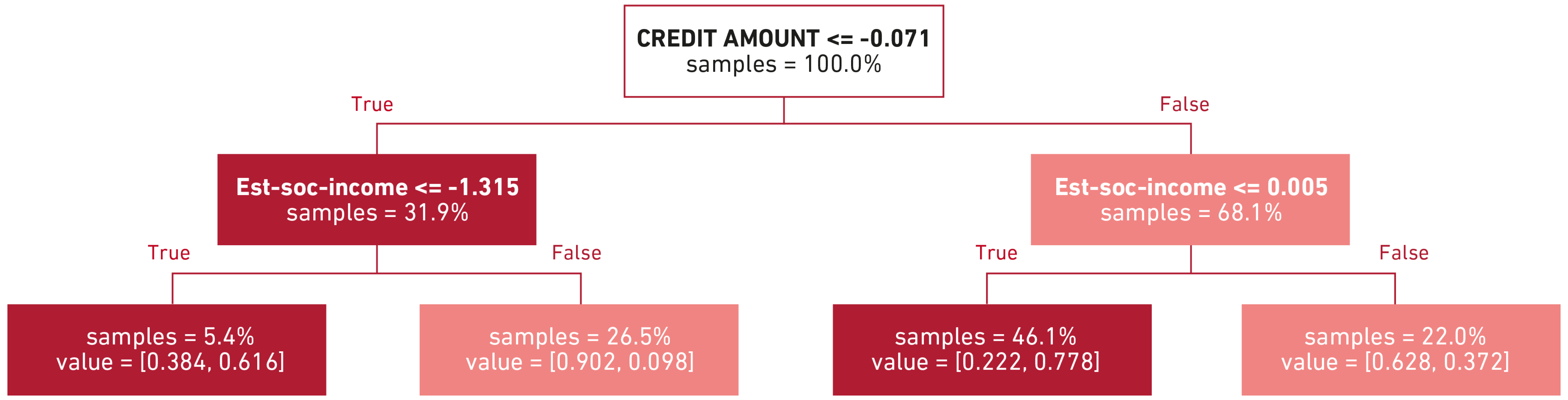
THE INSIGHTS.

THE INSIGHTS.

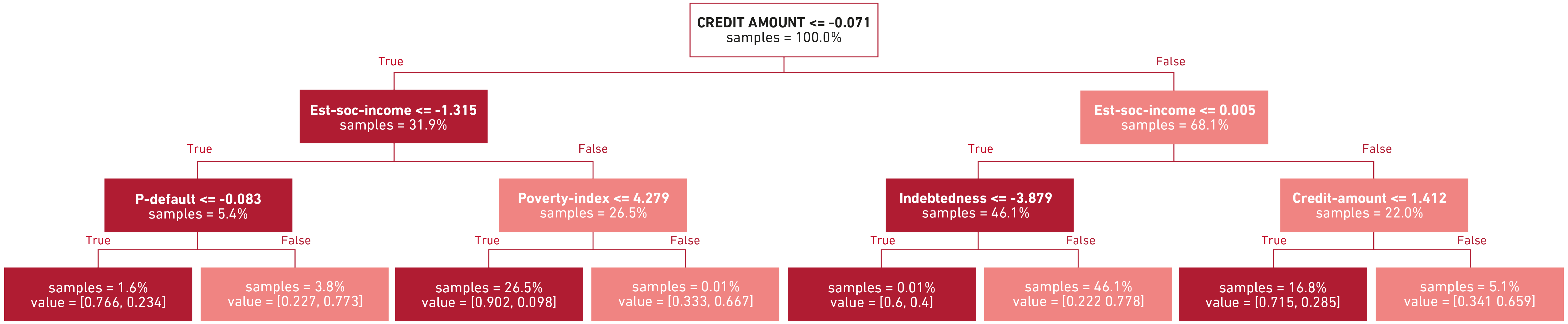
Depth 1



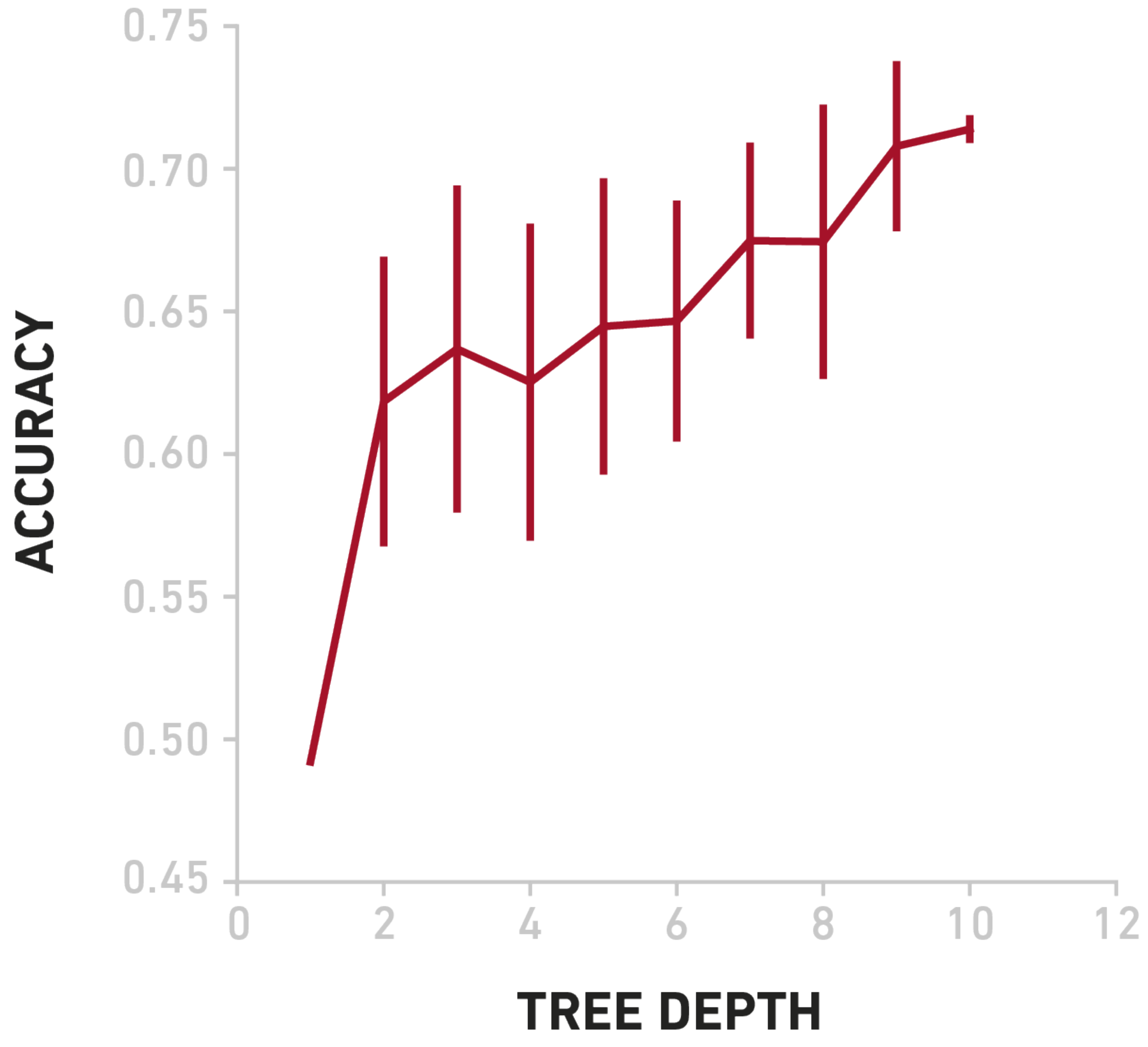
Depth 2



Depth 3



THE INSIGHTS.



CONTENTS.

CONTENTS.

CONCLUSIONS

FROM THEORY TO PRACTICE.

Putting the theoretical postulates of machine learning to practice **remains a challenge**. We need to develop new knowledge to ensure a more **sustainable** use of machine learning in practice.

FROM THEORY TO PRACTICE.

Putting the theoretical postulates of machine learning to practice **remains a challenge**. We need to develop new knowledge to ensure a more **sustainable** use of machine learning in practice.

Formalize the problem ***environmental adaptation*** *and discuss the mechanisms that allow* ***differential replication*** *through different forms of inheritance.*

FROM THEORY TO PRACTICE.

Putting the theoretical postulates of machine learning to practice **remains a challenge**. We need to develop new knowledge to ensure a more **sustainable** use of machine learning in practice.

1

Formalize the problem ***environmental adaptation*** *and discuss the mechanisms that allow* ***differential replication*** *through different forms of inheritance.*

2

Develop the theory behind ***inheritance by copying*** *to replicate the decision behavior of a model using another in scenarios with limited knowledge.*

FROM THEORY TO PRACTICE.

Putting the theoretical postulates of machine learning to practice **remains a challenge**. We need to develop new knowledge to ensure a more **sustainable** use of machine learning in practice.

1

Formalize the problem ***environmental adaptation*** *and discuss the mechanisms that allow* ***differential replication*** *through different forms of inheritance.*

2

Develop the theory behind ***inheritance by copying*** *to replicate the decision behavior of a model using another in scenarios with limited knowledge.*

3

Evaluate the feasibility of this technique in practice to ensure ***actionable accountability*** *of machine learning against rapidly changing conditions.*

FUTURE WORK.

FUTURE WORK.

Study the projection onto the space of **causal** and **privacy-preserving** models.

FUTURE WORK.

Study the projection onto the space of **causal** and **privacy-preserving** models.

Develop the **dual-pass approach** to solve the copying problem.

FUTURE WORK.

Study the projection onto the space of **causal** and **privacy-preserving** models.

Develop the **dual-pass approach** to solve the copying problem.

Devise additional mechanisms to ensure a **sustainable machine learning deployment.**

THANK YOU FOR YOUR ATTENTION!

Instituto de Ciencia de Datos e Inteligencia Artificial
December 22, 2021

esade
RAMON LLULL UNIVERSITY