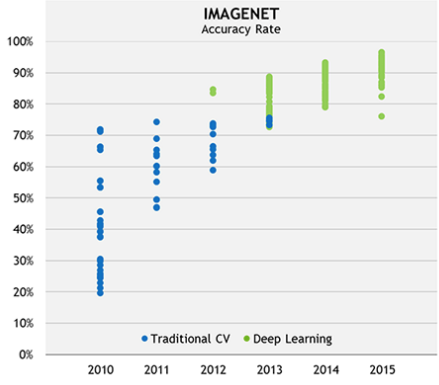
# Robustness in large-scale machine learning and its relevance to AI-enabled ECG

**Antônio Horta Ribeiro**
Uppsala University
Sweden

Imperial College, UK
July, 2023

# End-to-end learning

**Imagenet**



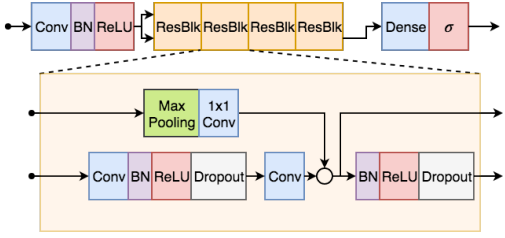

**Left:** dataset samples. **Right:** Models accuracy on the benchmark.

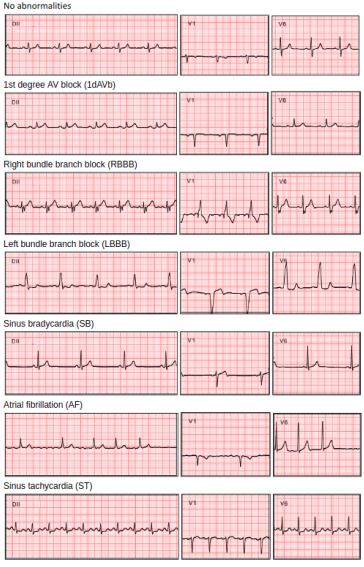
J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in 2009 IEEE conference on computer vision and pattern recognition, 2009, pp. 248–255.

# Automatic diagnosis of the ECG

- The Telehealth Center of Minas Gerais

- CODE dataset: historical data 2010 to 2017.
  - 2.3 M ECGs from $n = 1.6$M patients
- Develop and evaluate deep neural network




**A. H. Ribeiro** , M.H. Ribeiro, Paixão, G.M.M., et al. "Automatic diagnosis of the 12-lead ECG using a deep neural network," Nature Communications, 2020

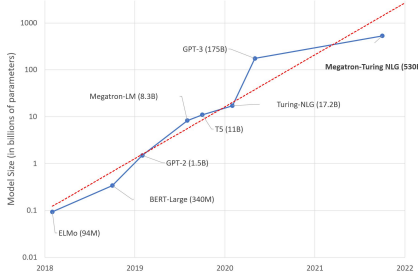
# Learning theoretical understanding of deep learning

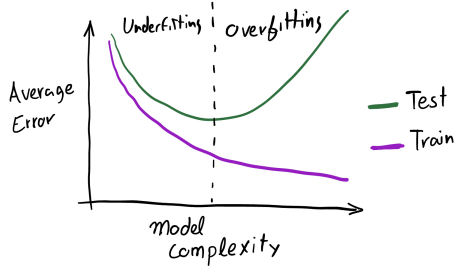
**Beyond Occam's Razor in System Identification: Double-Descent when Modeling Dynamics**
**Antônio H. Ribeiro**, Johannes N. Hendriks, Adrian G. Wills, Thomas B. Schön.
*IFAC Symposium on System Identification (SYSID), 2021.*
*Honorable mention: Young author award*

**Deep networks for system identification: a Survey.**
Gianluigi Pillonetto, Aleksandr Aravkin, Daniel Gedon, Lennart Ljung, **Antonio H. Ribeiro**, Thomas B. Schön.
*Under review Automatica* (2023)

# Generalization of deep neural networks

- Bias-variance tradeoff.
- Model size in DNN




- Deep neural networks can fit randomly labeled data and still generalize.

C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals. Understanding deep learning requires rethinking generalization. ICLR, 2017

# The principle of parsimony

- *"Everything should be made as simple as possible, but not simpler"*(**Albert Einsten**)
- *"Plurality should not be posited without necessity"*
  (**William of Ockham**)
- *Of two competing theories, the simpler explanation of an entity is to be preferred* **(Occam's razor)**
- *"It is superfluous to suppose that what can be accounted for by a few principles has been produced by many."*(**Summa Theologica, Thomas Aquinas**)
- *"To think is to forget a difference, to generalize, to abstract. In the overly replete world of Funes, there were nothing but details."*
  (**Funes, the Memorious, Jorge Luis Borges**)

# Simple model of study

- Nonlinear transformation $\phi(x)$, input to feature space

$$\phi : \mathbb{R}^{\#inputs} \mapsto \mathbb{R}^{\#parameters}.$$

- Linear model:

$$\hat{y} = \widehat{\beta}^{\top} \phi(x)$$

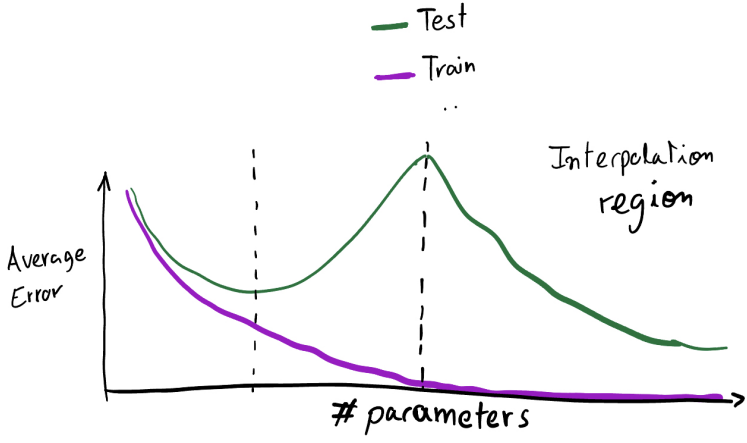
- Estimation procedure:

$$\min_{\beta} \sum_{i=1}^{\#train} (y_i - \widehat{\beta}^{\top} \phi(x_i))^2$$

- Optimization procedure: Gradient descent.

$$\beta^{i+1} = \beta^{i} - \gamma \nabla V(\beta^{i})$$

# Double-descent and benign overfitting




M. Belkin, D. Hsu, S. Ma, and S. Mandal, "Reconciling modern machine-learning practice and the classical bias–variance trade-off," Proceedings of the National Academy of Sciences, vol. 116, no. 32, pp. 15849–15854, 2019, doi: 10.1073/pnas.1903070116.

# The importance of implicit regularization

## Solutions of a linear system

The system

$$X\beta = y$$

has:

- no solution if $\#parameters < \#train$
- one unique solution if $\#parameters = \#train$
- multiple solution if $\#parameters > \#train$

Gradient descent converges to the minimum-norm solution:

$$\min_{\theta} \|\beta\|_2 \quad \text{subject to} \quad X\beta = y.$$

# Results

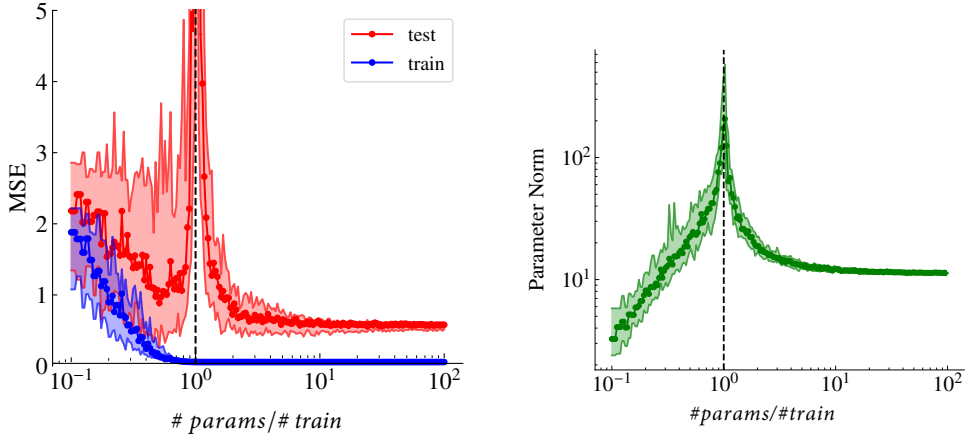



Figure: **Double-descent in system identification.** *Right:* MSE = Mean square error. *Right:* Parameter norm double descent curve.

**Beyond Occam's Razor in System Identification: Double-Descent when Modeling Dynamics**
**Antônio H. Ribeiro**, Johannes N. Hendriks, Adrian G. Wills, Thomas B. Schön.
*IFAC Symposium on System Identification (SYSID), 2021. Honorable mention: Young author award*

# Adversarial examples and robustness

**Regularization properties of adversarially-trained linear regression**
**Antônio H. Ribeiro**, Dave Zachariah, Francis Bach, Thomas B. Schön.
*Submited NeurIPS* (2023)

**Overparameterized Linear Regression under Adversarial Attack.**
**Antônio H. Ribeiro**, Thomas B. Schön.
*IEEE Transactions on Signal Processing* (2023)

# Adversarial attacks

- Neural networks can be vulnerable:

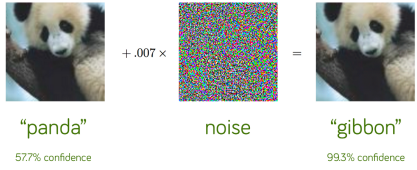



**Figure:** Effect of adversarial training on image classification.

Source: I. J. Goodfellow, J. Shlens, C. Szegedy , *“Explaining and Harnessing Adversarial Examples”*, ICLR 2015

- Neural networks in ECG applications display the same behavior:

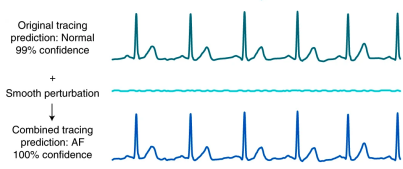



**Figure:** Effect of adversarial training on ECG Classification.

Source: Han, X., Hu, Y., Foschini, L. et al. Deep learning models for electrocardiograms are susceptible to adversarial attack. Nature Medicine 26, 360–363 (2020). https://doi.org/10.1038/s41591-020-0791-x

# Can large models be robust?

1. Enlarging the function classes we are able to find models that are smoother.

   S. Bubeck and M. Sellke. A Universal Law of Robustness via Isoperimetry. arXiv:2105.12806, June2021

2. Robustness-accuracy tradeoff

   A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu. Towards Deep Learning Models Resistantto Adversarial Attacks.Proceedings of the International Conference for Learning Representations (ICLR), 2018.

# Framework: Linear regression

*Simplest case where adversarial vulnerability has been observed.*

I. J. Goodfellow, J. Shlens, C. Szegedy , *"Explaining and Harnessing Adversarial Examples"*, ICLR 2015
D. Tsipras, S. Santurkar, L. Engstrom, A. Turner, and A. Ma, *"Robustness May Be At Odds with Accuracy,"* ICLR, p. 23, 2019.

- Training dataset:
$$(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n) \Rightarrow \widehat{\beta}$$
- Model prediction
$$\widehat{y} = \widehat{\beta}^{\mathsf{T}} x$$
- $\text{Error}(\widehat{\beta}) = y - x^{\mathsf{T}}\widehat{\beta}$
- $\text{Adv-error}(\widehat{\beta}) = \max_{\|\Delta x\| \leq \delta} \left( y - (x + \Delta x)^{\mathsf{T}} \widehat{\beta} \right)$

# Minimum $\ell_2$-norm interpolator under adversarial attacks

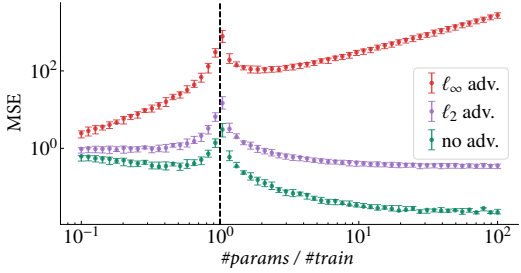



**Figure:** Adv. risk. minimum $\ell_2$-norm interpolator

**Overparameterized Linear Regression under Adversarial Attack.**
**Antônio H. Ribeiro**, Thomas B. Schön.
*IEEE Transactions on Signal Processing* (2023)

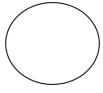
# Adversarial training in linear models

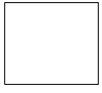
Linear regression:

$$\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} (y_i - x_i^{\mathsf{T}} \beta)^2$$

Adversarial training:

$$\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} \max_{\|\Delta x_i\| \leq \delta} (y_i - (x_i + \Delta x_i)^{\mathsf{T}} \beta)^2$$





$\{\|\Delta x\|_2 \leq \delta\}$ $\{\|\Delta x\|_\infty \leq \delta\}$

# Minimum-norm interpolator and adversarial training

Adversarial training

$$\min \frac{1}{n} \sum_{i=1}^{n} \max_{\|\Delta x\| \leq \delta} (y_i - (x_i + \Delta x)^\mathsf{T} \beta)^2$$

**Theorem**

Adversarial training is minimized at the minimum norm interpolator

$$\min_{\beta} \|\beta\|_* \quad \text{subject to} \quad X\beta = y$$

iff $0 < \delta < \bar{\delta}$.

**Regularization properties of adversarially-trained linear regression**
**Antônio H. Ribeiro**, Dave Zachariah, Francis Bach, Thomas B. Schön.
*Submited NeurIPS* (2023)

# Adversarial training and robustness

- Explanation for robustness.

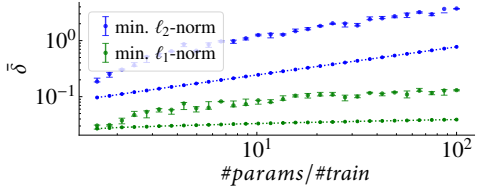


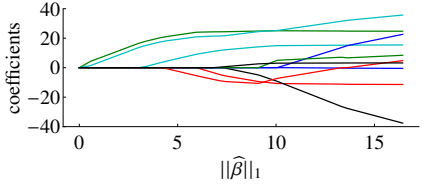

Figure: Threshold $\bar{\delta}$ *vs* number of features $m$.

- Mismatched training and evaluation can result in brittleness.

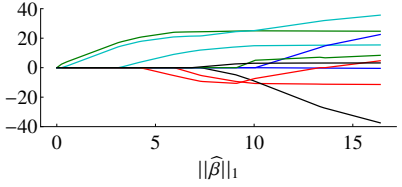
# Conclusions and future directions

*Is adversarial training useful on its own?*

- Similarities between $\ell_\infty$-adversarial training and lasso.




Lasso regression




$\ell_\infty$-adversarial training

- *Pivotal*

**Pivotal method: definition**

adversarial radius $\delta$ can be chosen without the knowledge of the noise variance.

- It allows for a default choice of the adversarial that works well out-of-the-box
- Efficient solver: interactive reweighted ridge regression

# AI-ECG: three directions

1. Automatic diagnosis;
2. Screening;
3. Prognosis.

**Figure** Automated ECG interpretation Glasgow (1971).

Macfarlane, P.W.; Kennedy, J. "Automated ECG Interpretation—A Brief History from High Expectations to Deepest Networks." Hearts 2021.

# Screening for Chagas diseases

- 6 million people infected.
- Diagnosed with blood test.
- early diagnosis and treatment halt progression.
- Low detection rates

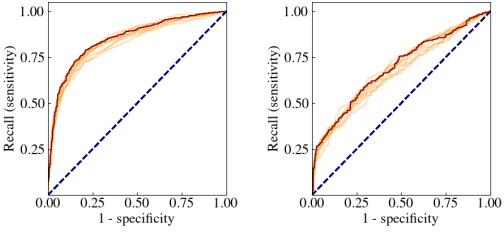
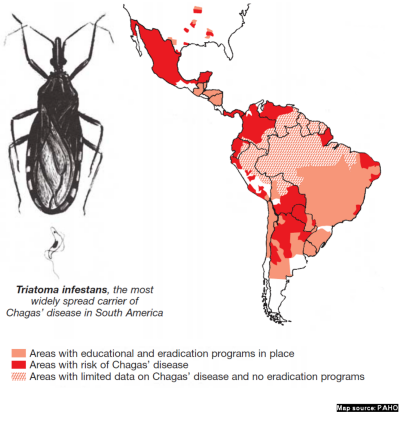



**Figure:** Performance on predicting Chagas Disease.




**Screening for Chagas disease from the electrocardiogram using a deep neural network**
Carl Jidling, Daniel Gedon, Thomas B. Schön, Claudia Di Lorenzo Oliveira, Clareci Silva Cardoso, Ariela Mota Ferreira, Luana Giatti, Sandhi Maria Barreto, Ester C. Sabino, Antônio L. P. Ribeiro, **Antônio H. Ribeiro**
Plos Neglected Tropical Diseases (2023)

# Prognosis and ECG-age

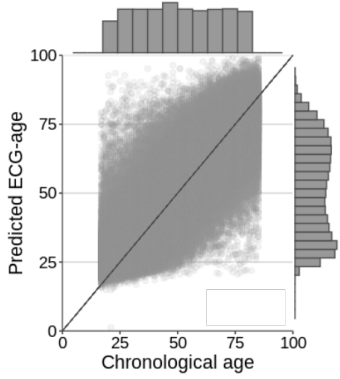



Figure: Predicted vs estimated age. MAE= 8.38 years.

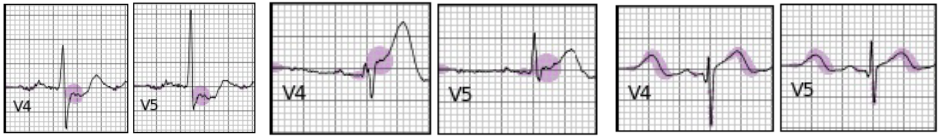
$\Delta$ age = ECG-age - age

| All ECGS | |
|---|---|
| $\Delta$ age $<$ - 8 y | 0.78 |
| $\Delta$ age $>$ 8 y | 1.79 |
| **Only normals** | |
| $\Delta$ age $<$ - 8 y | 0.66 |
| $\Delta$ age $>$ 8 y | 1.53 |

**Deep neural network estimated electrocardiographic-age as a mortality predictor**
Emilly M. Lima, **Antônio H. Ribeiro**, Gabriela MM Paixão, et. al.
Nature Communications. (2021)

# Challenges

- **Interpretability** Attempt to draw real electrocardiographic knowledge from the AI-ECG effort.




**Grad-CAM plots**. **(Left)** STEMI with typical ST-segment elevation highlighted. **(Middle)** STEMI with another feature highlighted is not typical. **(Right)** NSTEMI with ST-segment depression highlighted.

**Development and validation of deep learning ECG-based prediction of myocardial infarction in emergency department patients.**
Stefan Gustafsson, Daniel Gedon, Erik Lampa, **Antônio H. Ribeiro**, Martin J. Holzmann, Thomas B. Schön, Johan Sundström.
Scientific Reports (2022)

- **Robustness**. Ability to work in real situations.

*ML algorithms don't need to be really interpreable to be useful in clinical practice. But they need to be robust!*

# Current AI-ECG project

- Classification
  - CODE-v2.0 - 50 classes
- Screening
  - Predicting atrial fibrillation from sinus rythm
  - Left Ventricular Systolic Dysfunction
  - Detecting Miocardial Infarction - NSTEMI
  - Chagas disease
  - Electrolyte Prediction from the ECG.
- Prognosis
  - Validating algorithms in longitudinal studies.

**Thank you!**

✉ antonio.horta.ribeiro@it.uu.se
🌐 antonior92.github.io