

The Statistical Learning Theory in Practical Problems

Rodrigo Fernandes de Mello

Universidade de São Paulo

Instituto de Ciências Matemáticas e de Computação

Departamento de Ciências de Computação

<http://www.icmc.usp.br/~mello>

mello@icmc.usp.br

Theoretical Aspects

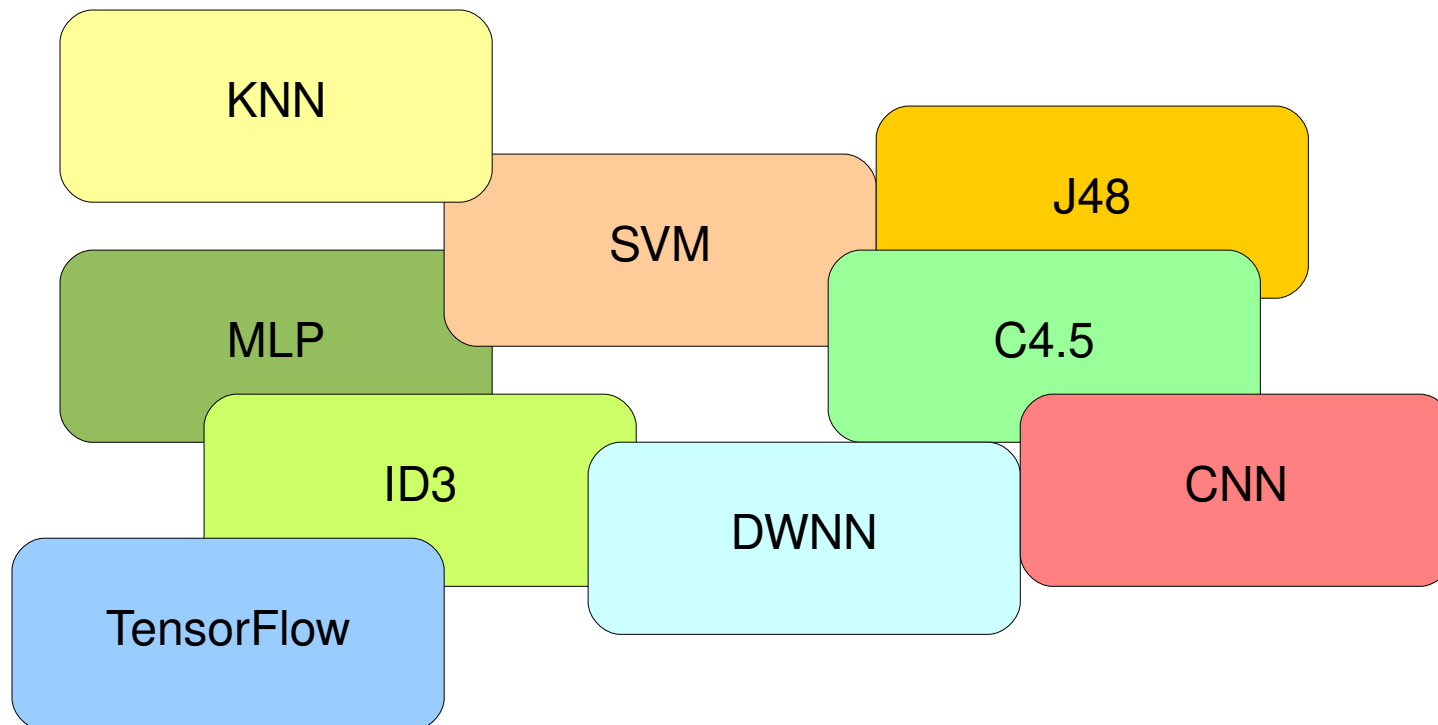
- Interested in introducing my research interests

Theoretical Aspects

- Interested in introducing my research interests
 - What a better way than the Statistical Learning Theory?

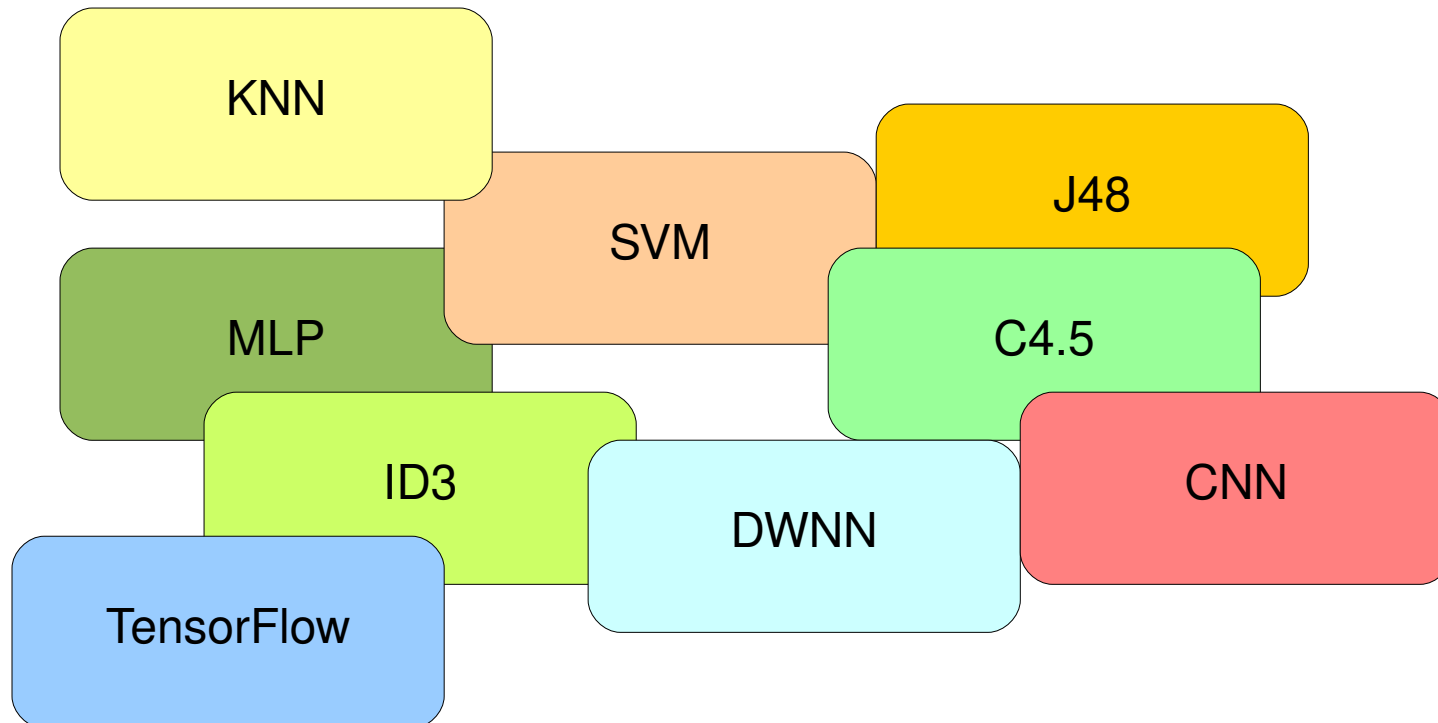
Statistical Learning Theory

- So many classification algorithms:
 - How can we assess any of those?



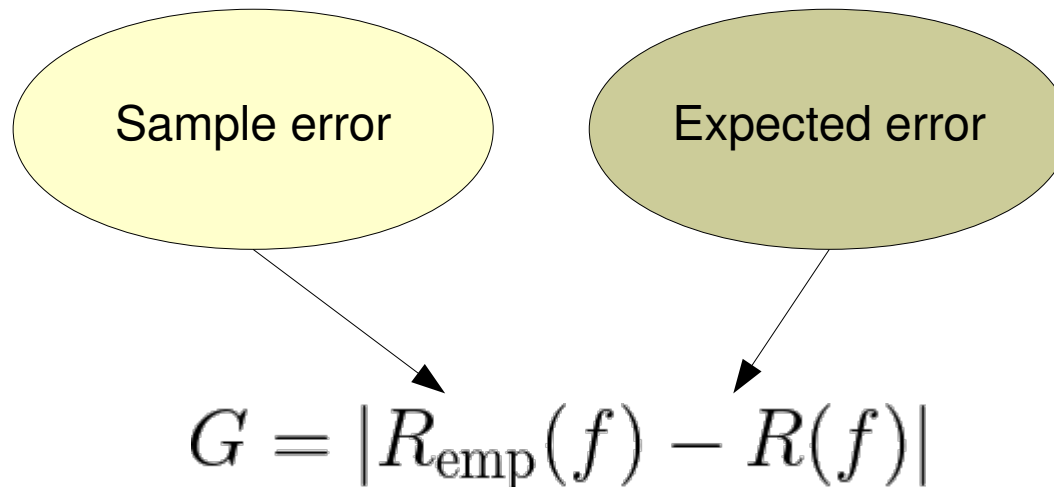
Statistical Learning Theory

- So many classification algorithms:
 - How can we assess any of those?
 - K-fold cross validation, leave-one-out, ...
 - How can we prove any of those learn?



Statistical Learning Theory

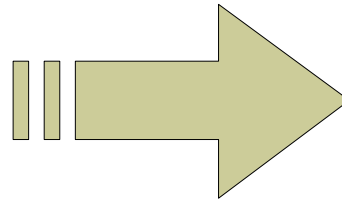
- Vapnik proposed the Statistical Learning Theory
 - Defined in the context of **supervised learning**
 - Learning guarantees and conditions
- What is the main call?



This is the concept of Generalization

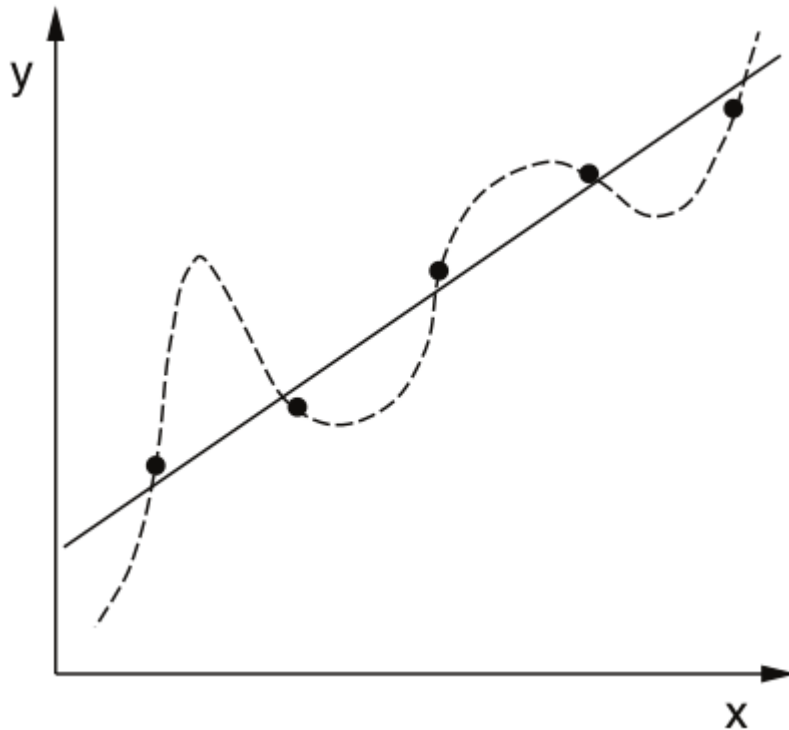
Statistical Learning Theory

- So what is generalization?
 - Consider someone has given you a textbook



Statistical Learning Theory: Bias-Variance Dilemma

- An example based on **regression**:

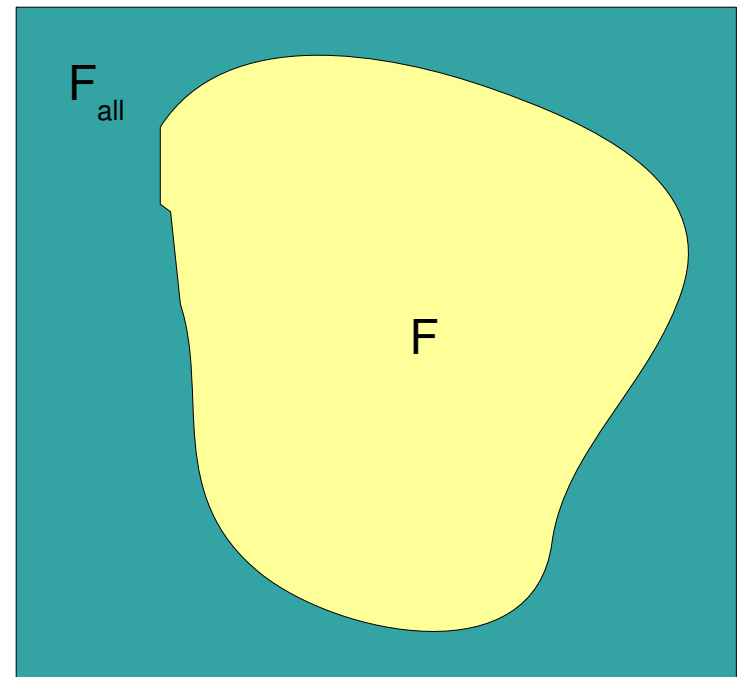
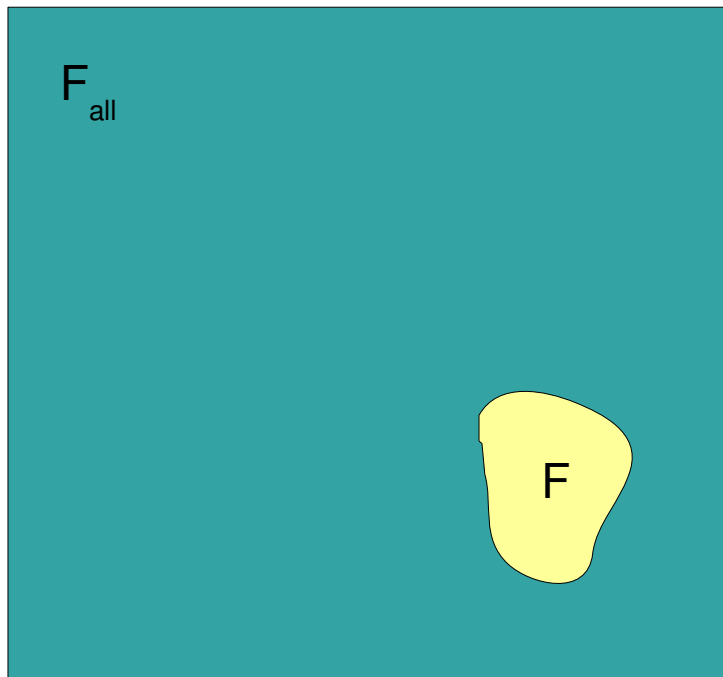


Which function is the best?

**To answer it we need to
assess their Expected Risks**

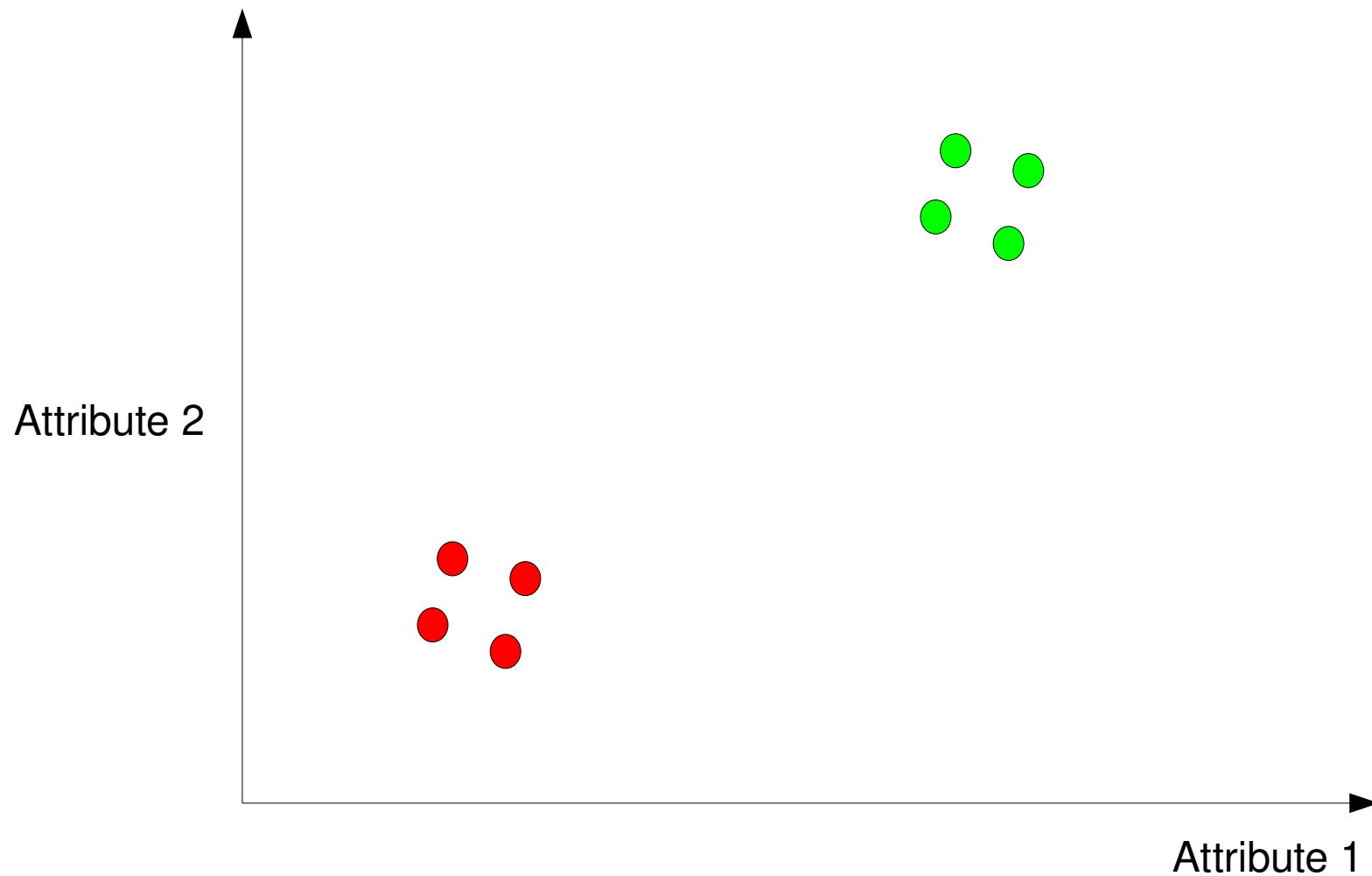
Statistical Learning Theory: Bias-Variance Dilemma

- The dichotomy associated to the Bias-Variance Dilemma



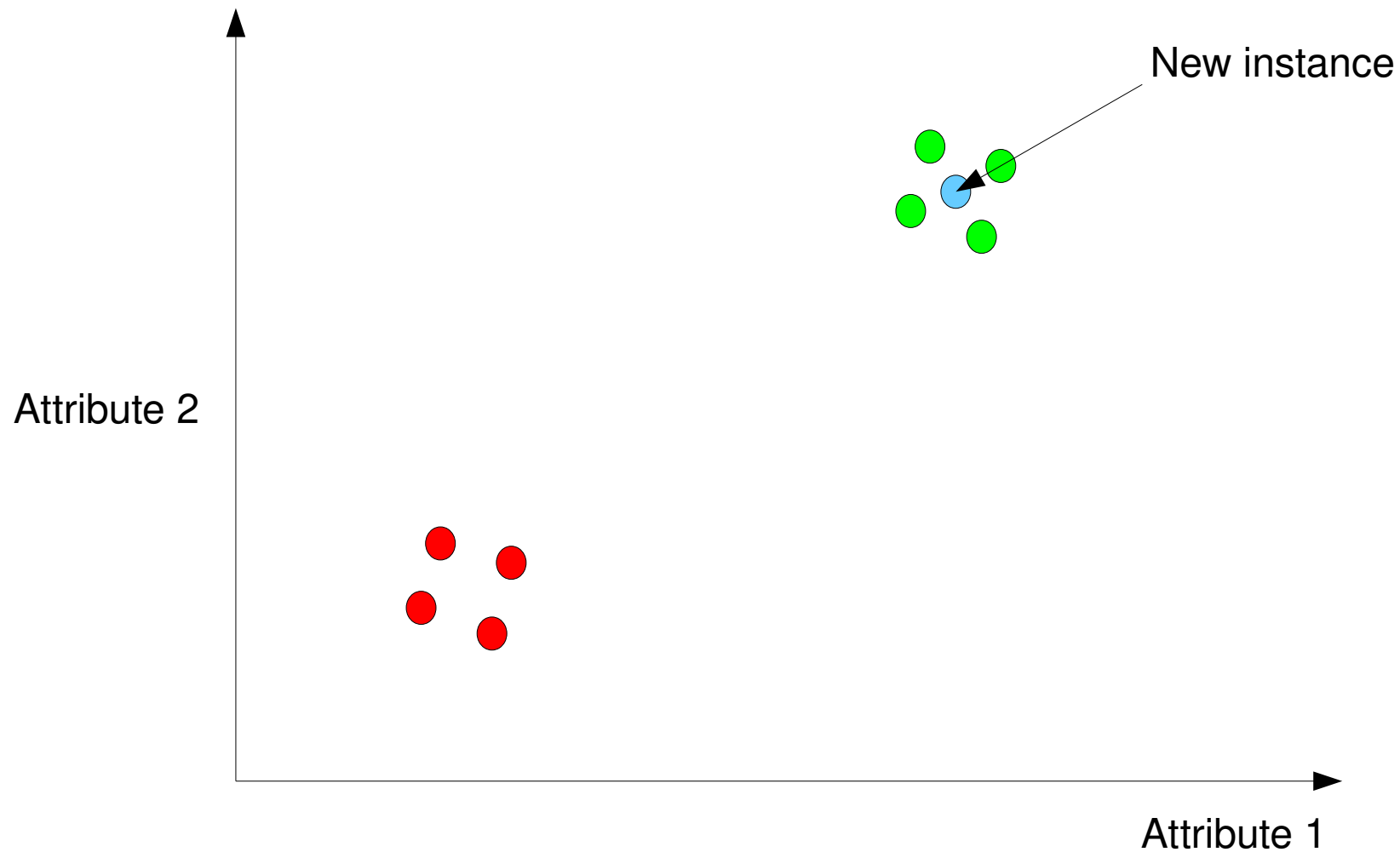
Distance-Weighted Nearest Neighbors

- Based on the same principles as the k-Nearest Neighbors



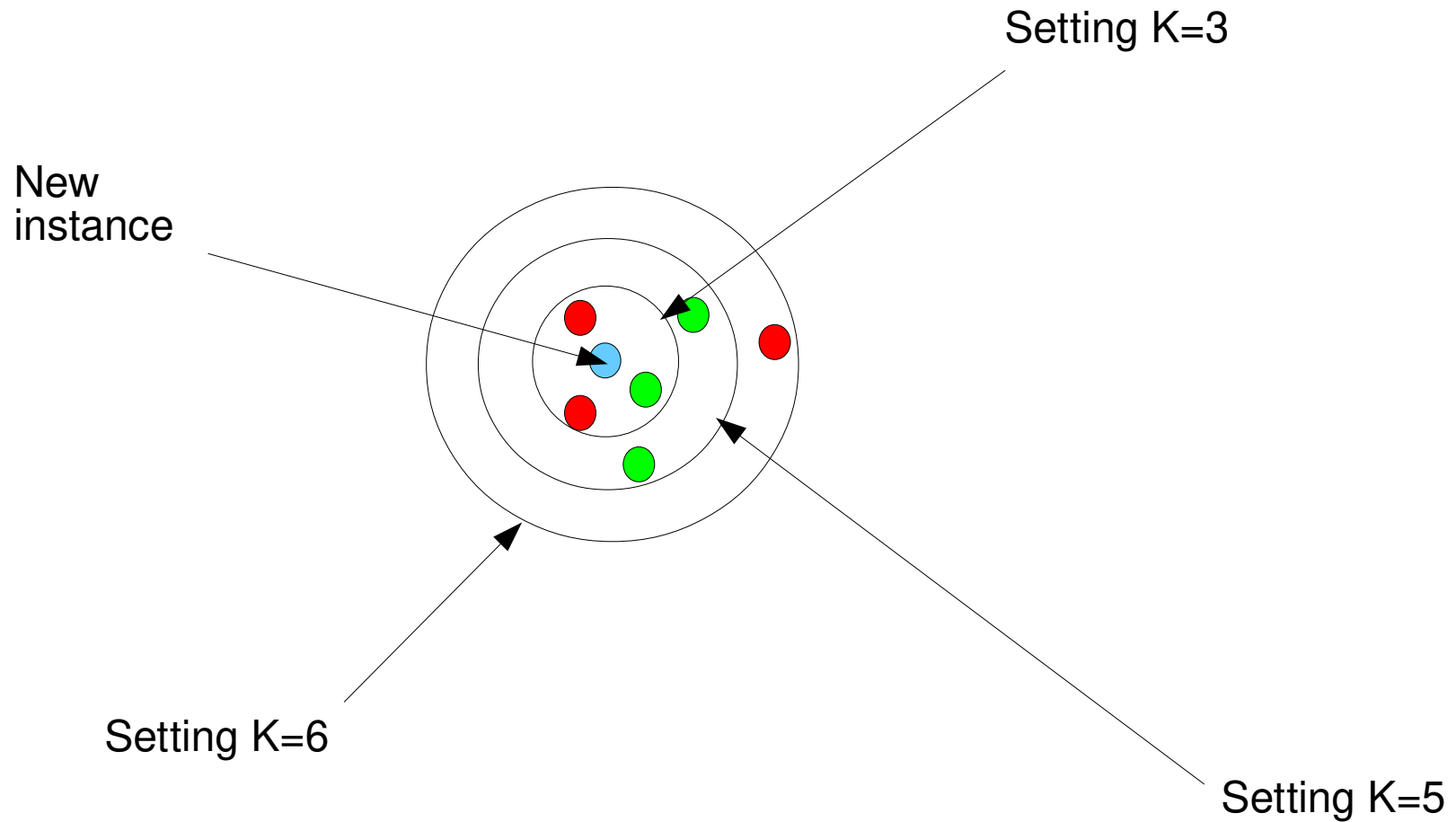
Distance-Weighted Nearest Neighbors

- Based on the same principles as the k-Nearest Neighbors



Distance-Weighted Nearest Neighbors

- Based on the same principles as the k-Nearest Neighbors



Distance-Weighted Nearest Neighbors

- It is based on Radial functions centered at the new instance a.k.a. query point
- Classification output:

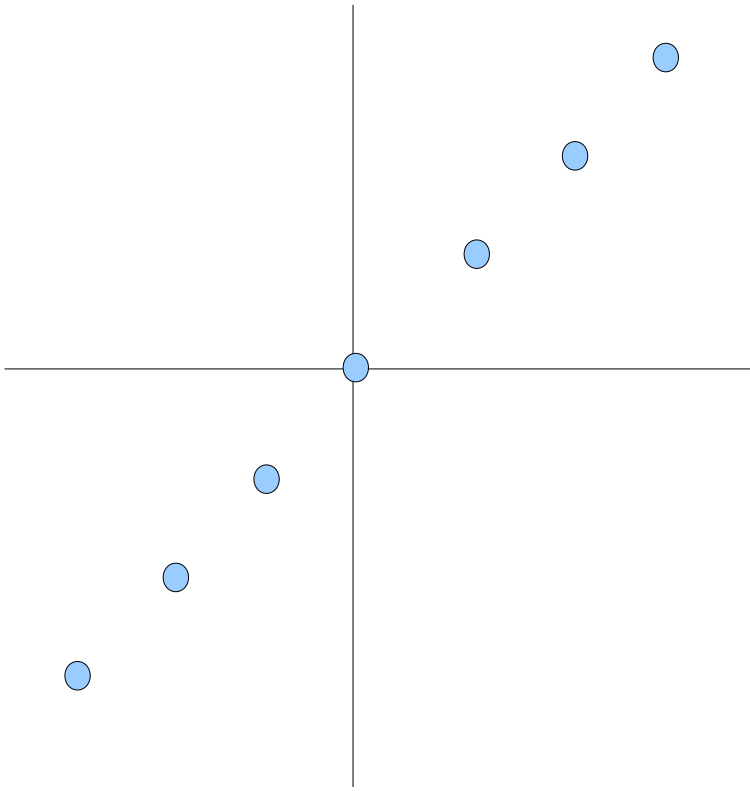
$$f(\mathbf{x}) = \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i}$$

- Given the weight function:

$$w_i = \exp \left(-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{2\sigma^2} \right)$$

Distance-Weighted Nearest Neighbors

- After implementing, test it on this simple example of an identity function:



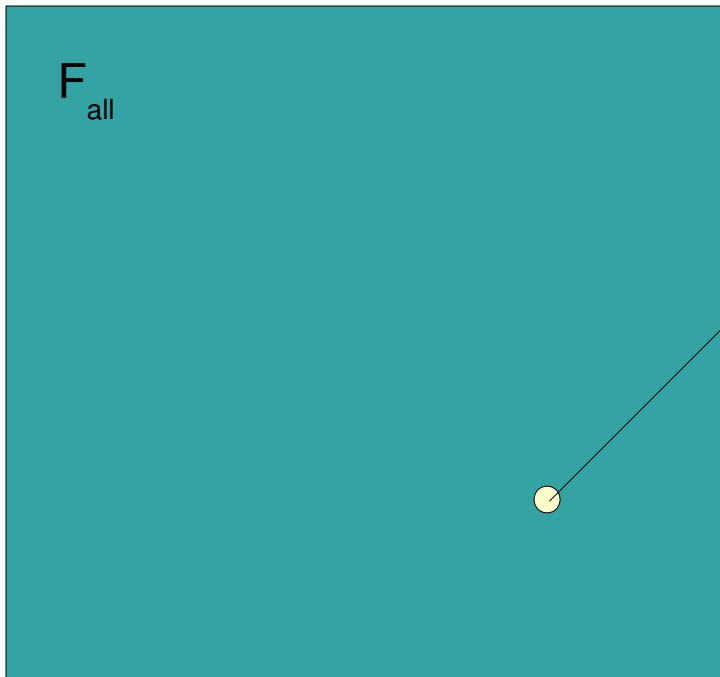
Two main questions:

- What happens if sigma is too big?
- What happens if sigma is too small?

So, how can we define the best value for sigma?

Distance-Weighted Nearest Neighbors

- When sigma tends to infinity

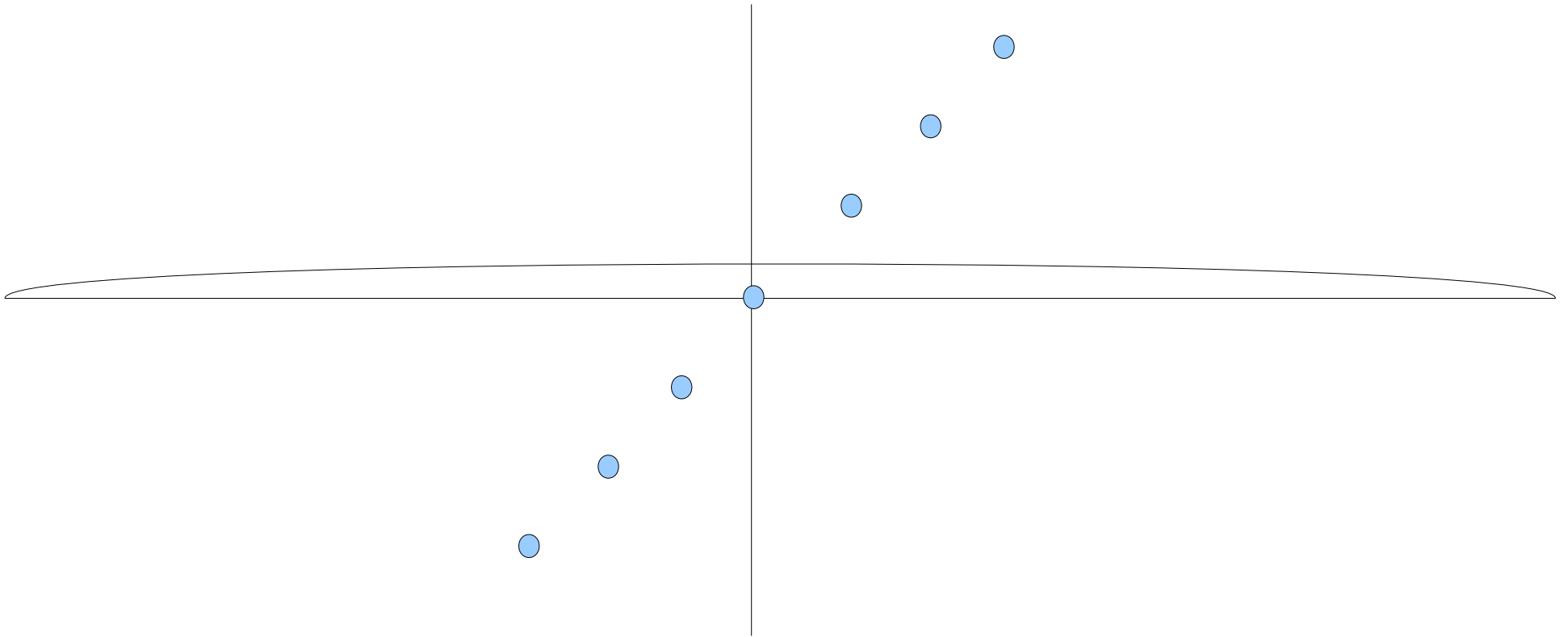


The space of admissible functions (bias) will contain a single function

In this case, the average function

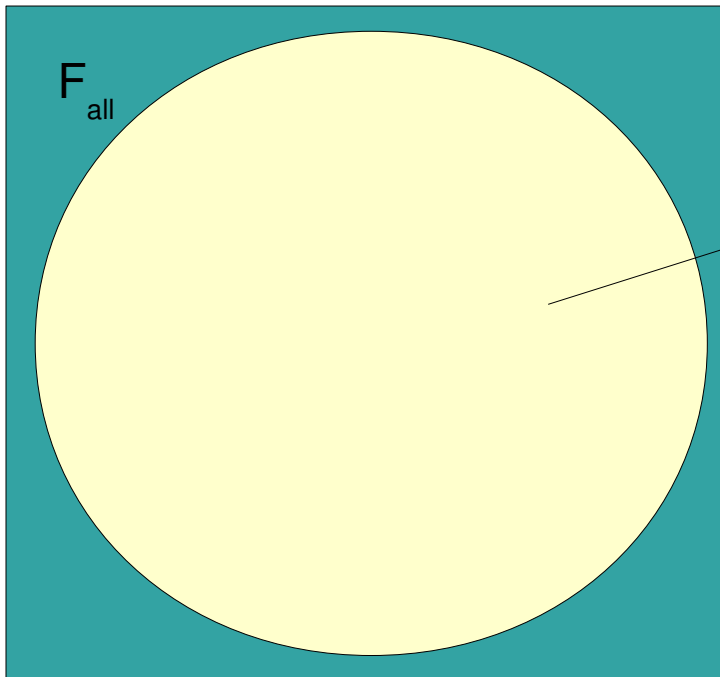
Distance-Weighted Nearest Neighbors

- When sigma tends to infinity



Distance-Weighted Nearest Neighbors

- When sigma tends to zero



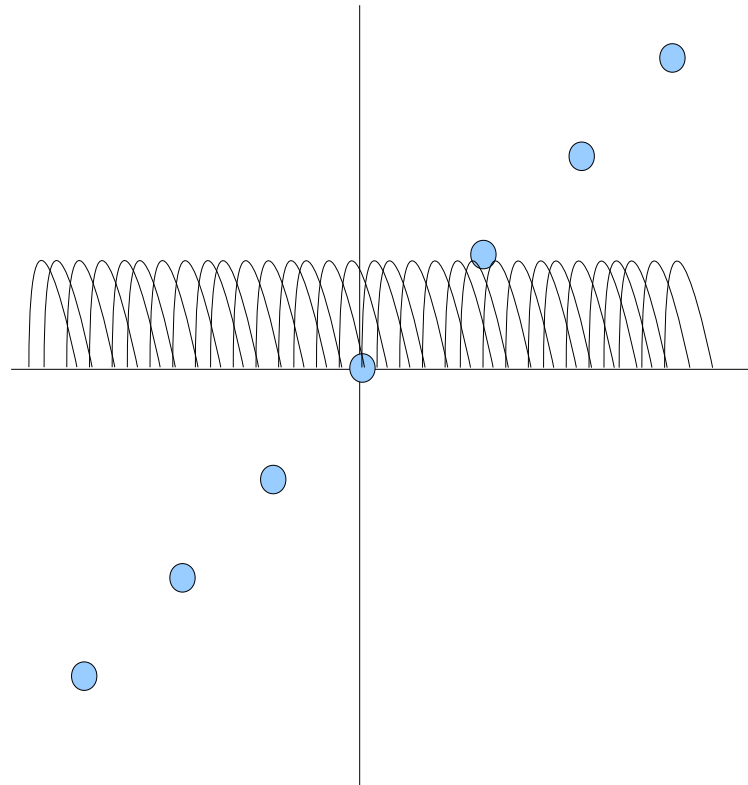
The space of admissible functions (bias) will tend to the whole space

What is the problem with that?

It will most probably contain at least one memory-based classifier

Distance-Weighted Nearest Neighbors

- When sigma tends to zero



- Putting in simple terms:

Theoretical Aspects

- Putting in simple terms:
 - Interested in proving learning in different scenarios

- Putting in simple terms:
 - Interested in proving learning in different scenarios
 - Building up a theoretical framework to prove time series modeling
 - Idea: use it to ensure concept drift detection

Theoretical Aspects


- Putting in simple terms:
 - Interested in proving learning in different scenarios
 - Building up a theoretical framework to prove time series modeling
 - Idea: use it to ensure concept drift detection
 - Theoretical Framework to design Deep Learning architectures

Living in the Past...

- Some practical results:
 - Bionic hand
 - Hardware under development at the northeast of Brazil
 - Mainly to support people living in extreme poverty
 - Time Series Visualization tool
 - www.tsviz.com.br
 - Cover song identification
 - Supported the ECAD system, Brazil
 - Copyright office of songs in Brazil
 - Go Digital, Brazil – incorporated by Acxiom, USA
 - Machine Learning for geopositioning systems
 - Marketing based on ML




- Vapnik, V. The Nature of the Statistical Learning, Springer, 2011
- Luxburg and Scholkopf, Statistical Learning Theory: Models, Concepts, and Results. Handbook of the History of Logic. Volume 10: Inductive Logic. Volume Editors: Dov M. Gabbay, Stephan Hartmann and John Woods, Elsevier, 2009
- Schölkopf, B., Smola, A. J., Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, MIT, 2002



Search

[Home](#) [Subjects](#) [Services](#) [Products](#) [Springer Shop](#) [About us](#)

[» Computer Science](#) [» Artificial Intelligence](#)



© 2018

Machine Learning

A Practical Approach on the Statistical Learning Theory

Authors: **Fernandes de Mello**, Rodrigo, **Antonelli Ponti**, Moacir