Statistical Learning Theory:
How to ensure that ML algorithms work?

Rodrigo Fernandes de Mello

Universidade de São Paulo

Instituto de Ciências Matemáticas e de Computação

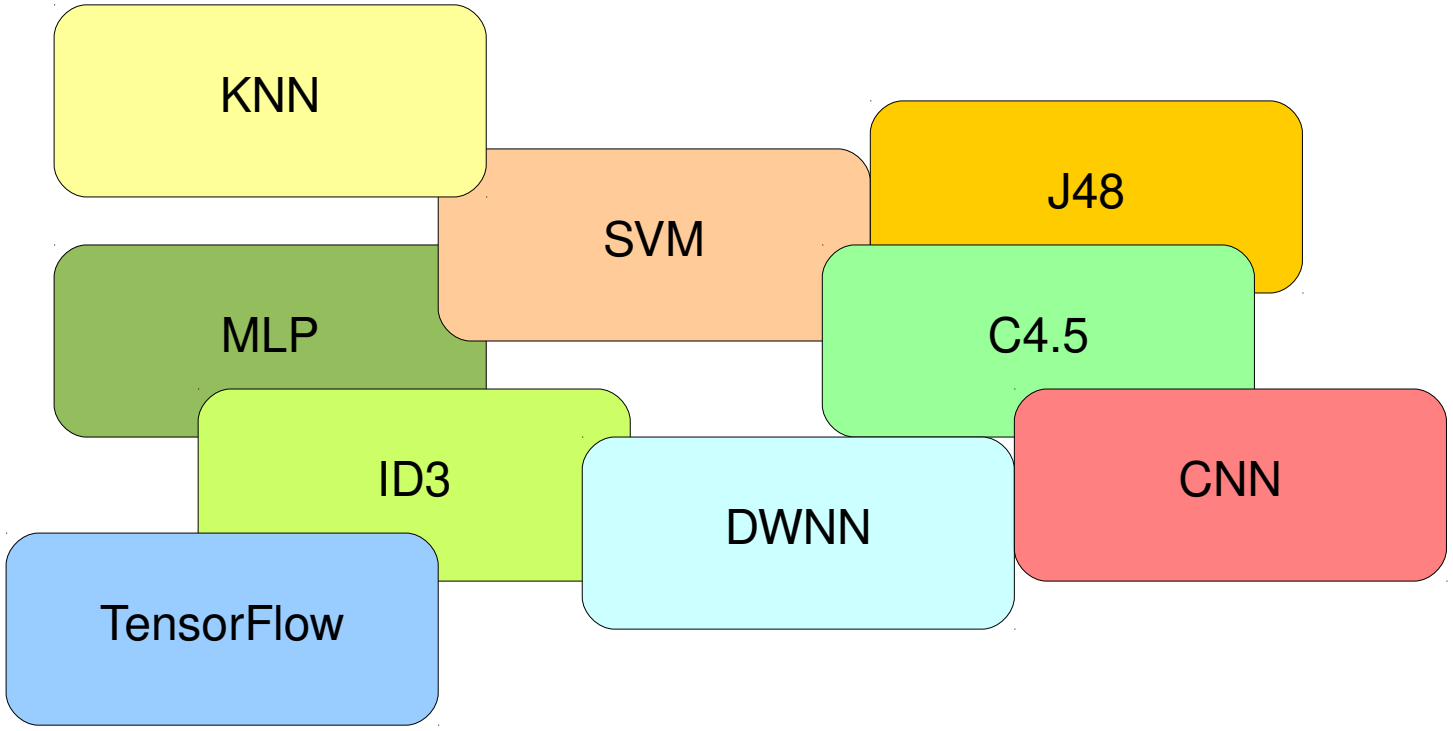
Departamento de Ciências de Computação

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

mello@icmc.usp.br

Statistical Learning Theory

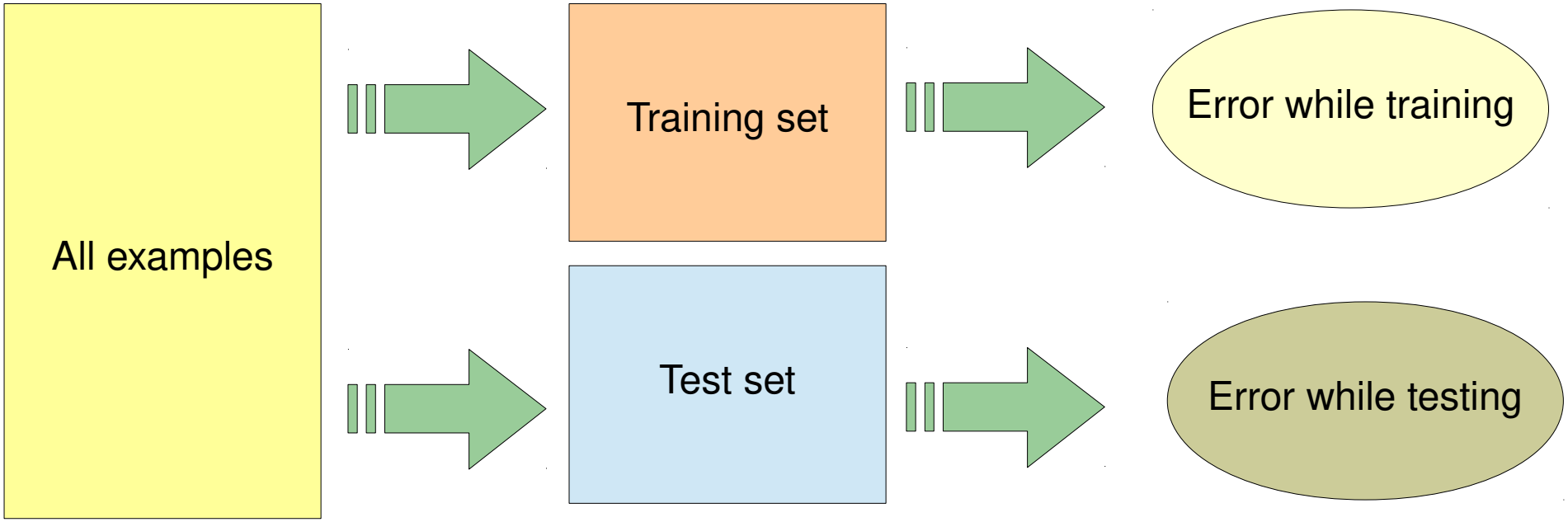
- So many classification algorithms:
  - How can we conclude about learning?



Statistical Learning Theory

- Vapnik proposed the Statistical Learning Theory
  - Defined in the context of **supervised learning**
  - Learning guarantees and conditions

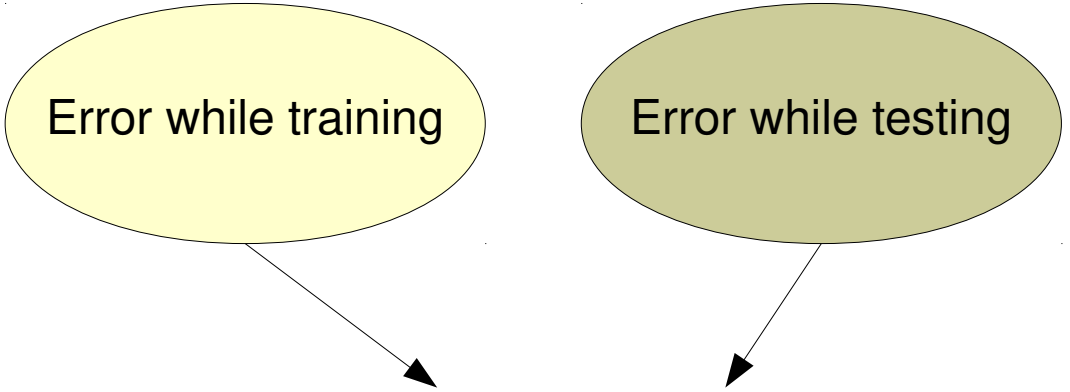
- What is the main call?



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?

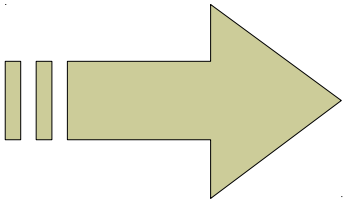


$$G = |R_{\text{train}}(f) - R_{\text{test}}(f)|$$

This is the concept of Generalization

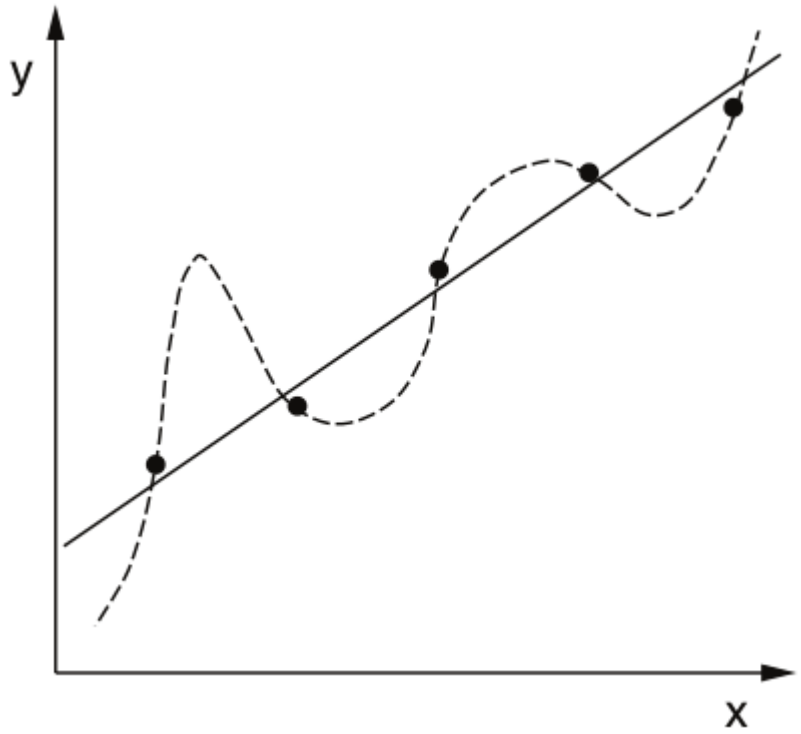
Statistical Learning Theory

- So what is generalization?
  - Consider you are a teacher and define a textbook for any course



Statistical Learning Theory: Bias-Variance Dilemma

- An example based on **regression**:



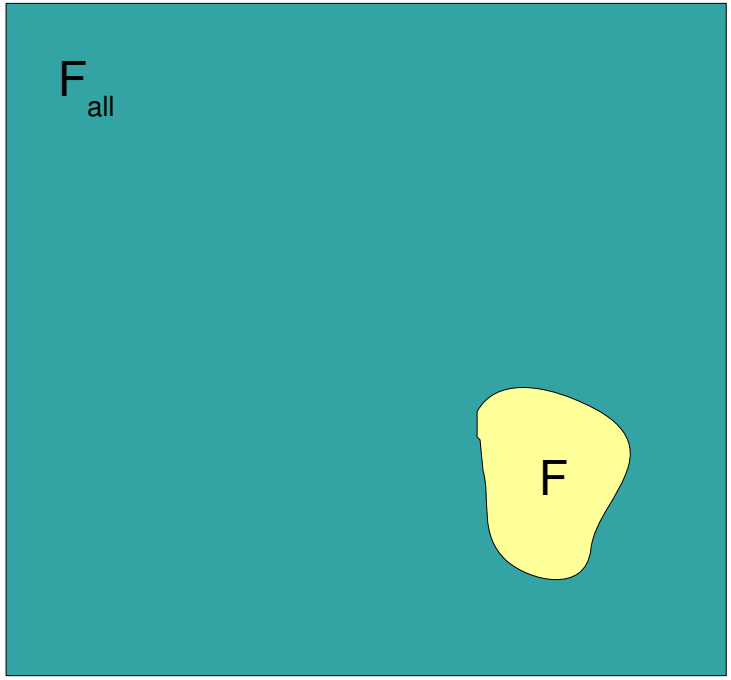
Which function is the best?

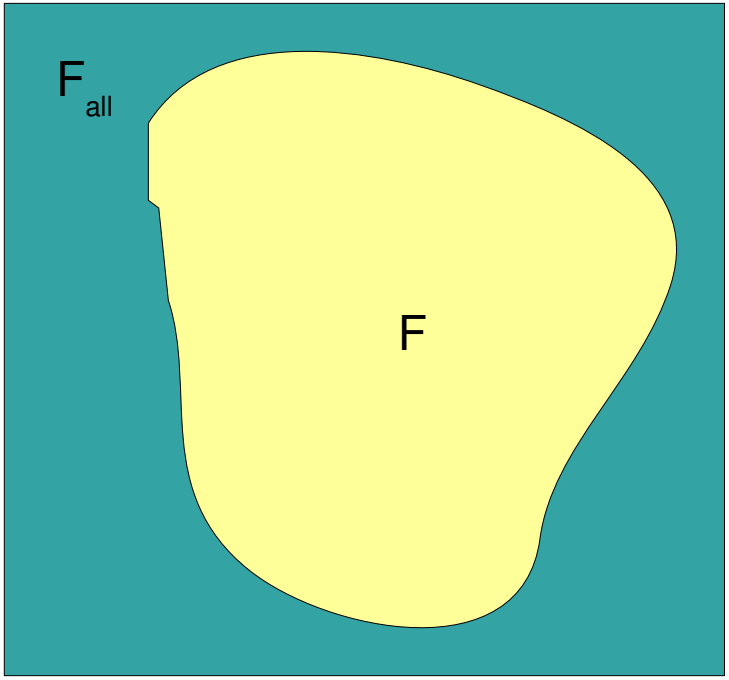
To answer it we need to assess their Expected Risks

Figure from Luxburg and Scholkopf, Statistical Learning Theory: Models, Concepts, and Results

Statistical Learning Theory: Bias-Variance Dilemma

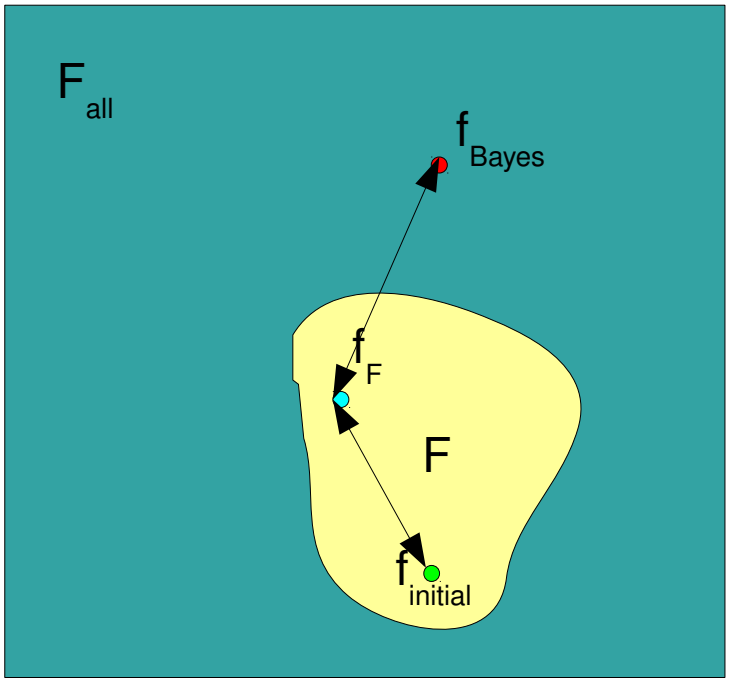
- The dichotomy associated to the Bias-Variance Dilemma





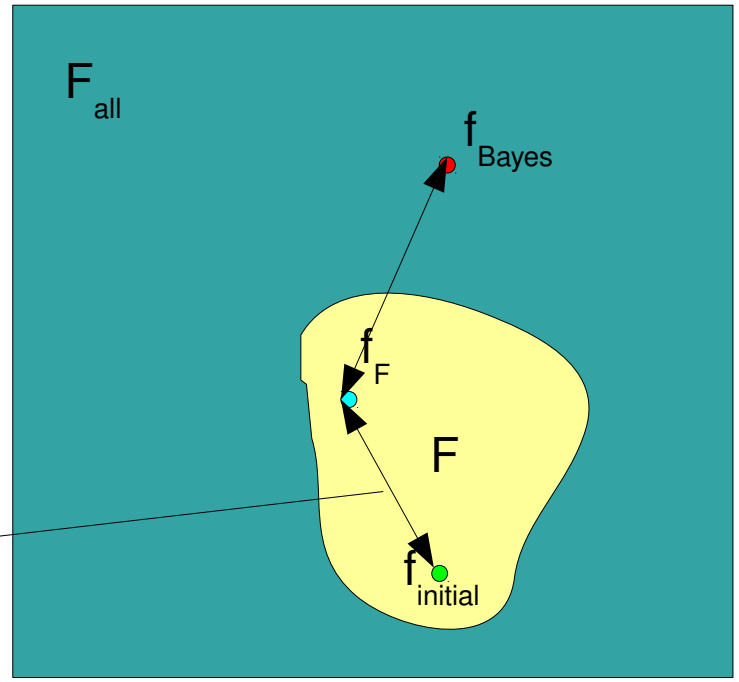
Statistical Learning Theory: Bias-Variance Dilemma

- The dichotomy associated to the Bias-Variance Dilemma



Statistical Learning Theory: Bias-Variance Dilemma

- The dichotomy associated to the Bias-Variance Dilemma



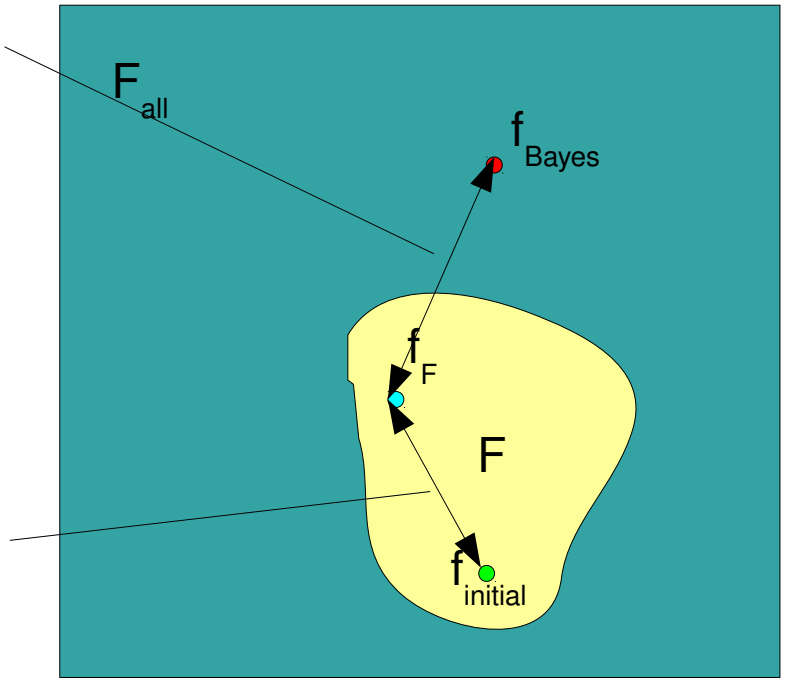
Estimation error:
how far our solution is
from the best classifier in F

Statistical Learning Theory: Bias-Variance Dilemma

- The dichotomy associated to the Bias-Variance Dilemma

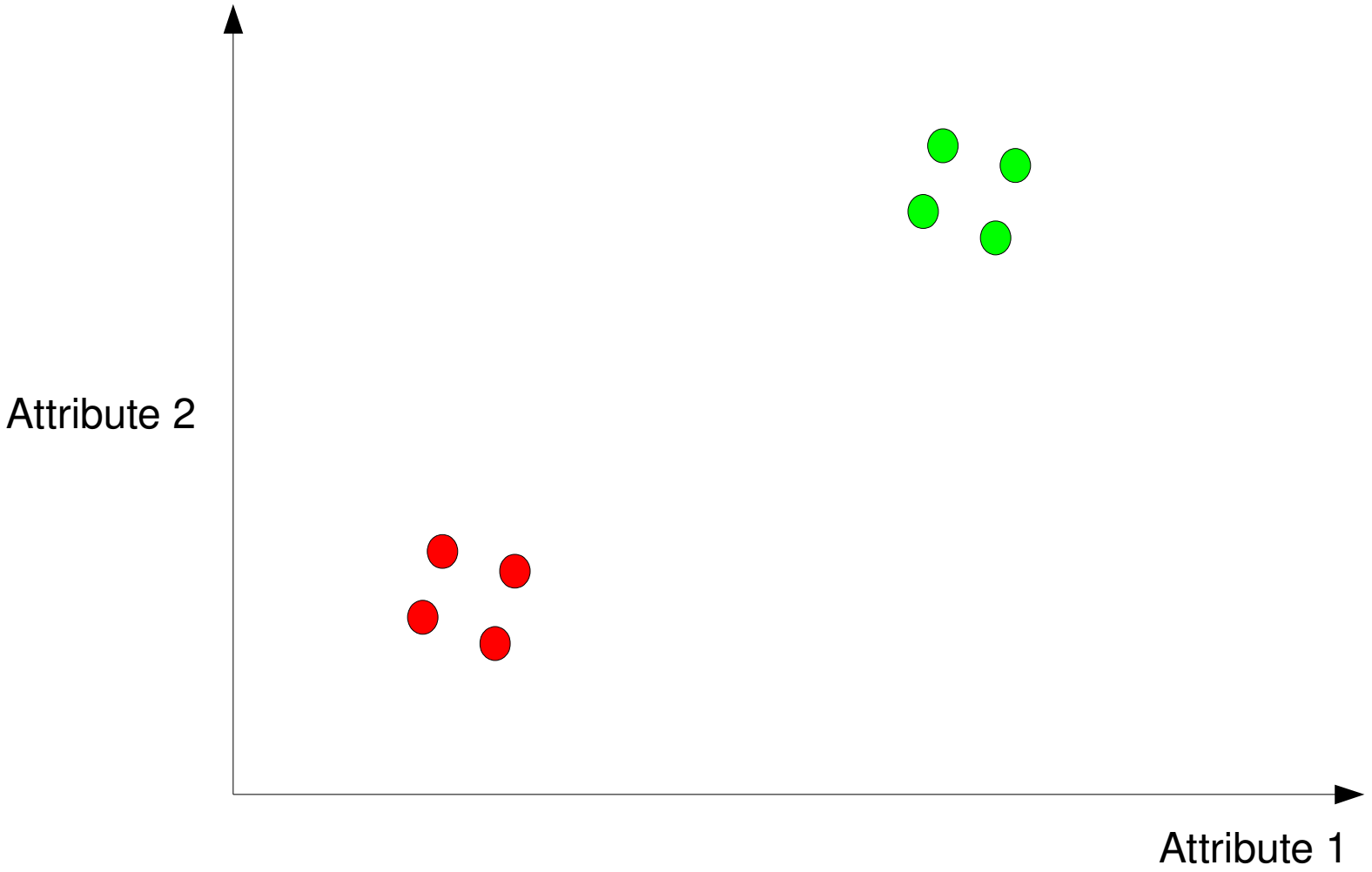
Approximation error:
how far the best
solution in F is
from the best classifier at all

Estimation error:
how far our solution is
from the best classifier in F



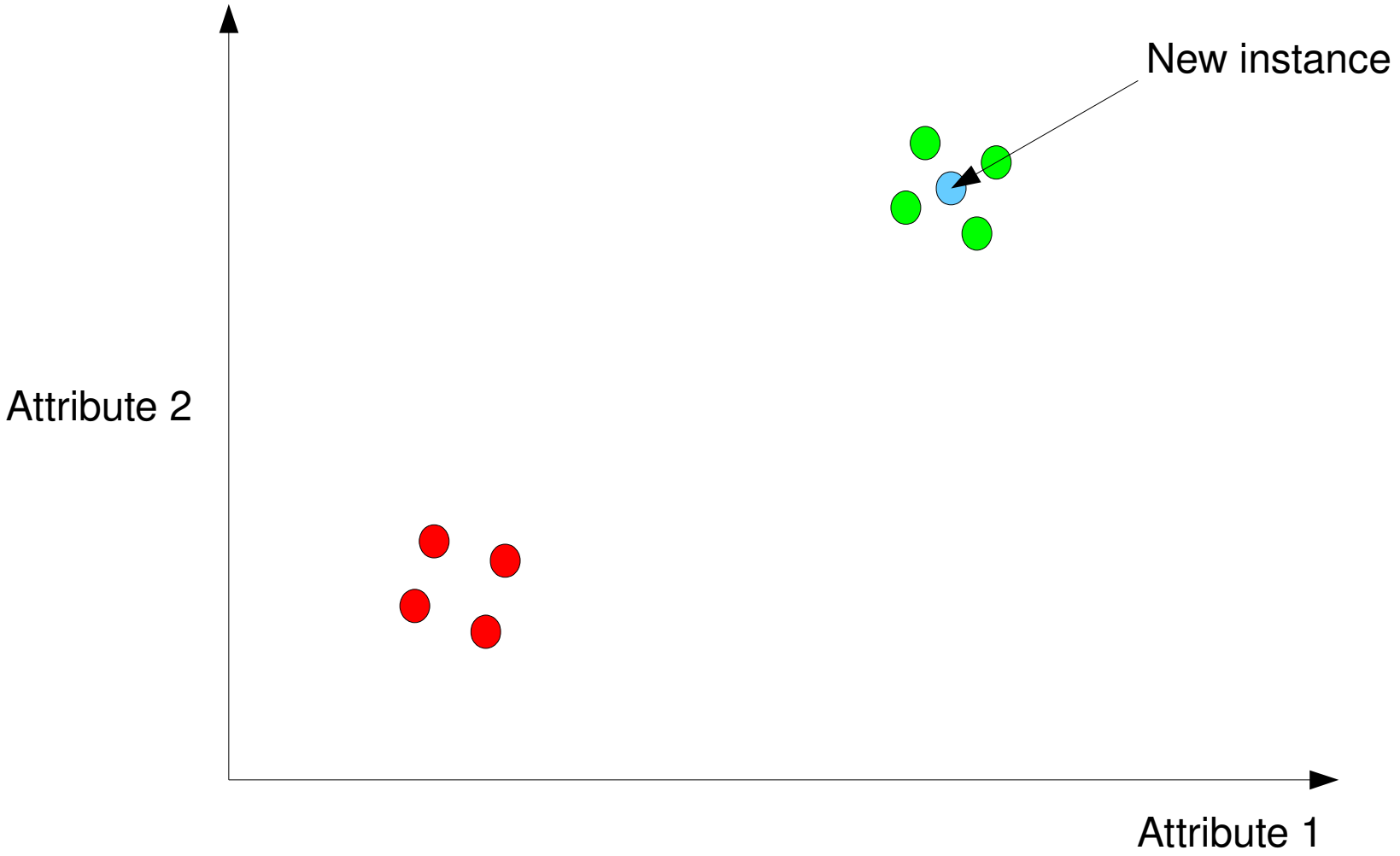
Distance-Weighted Nearest Neighbors

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



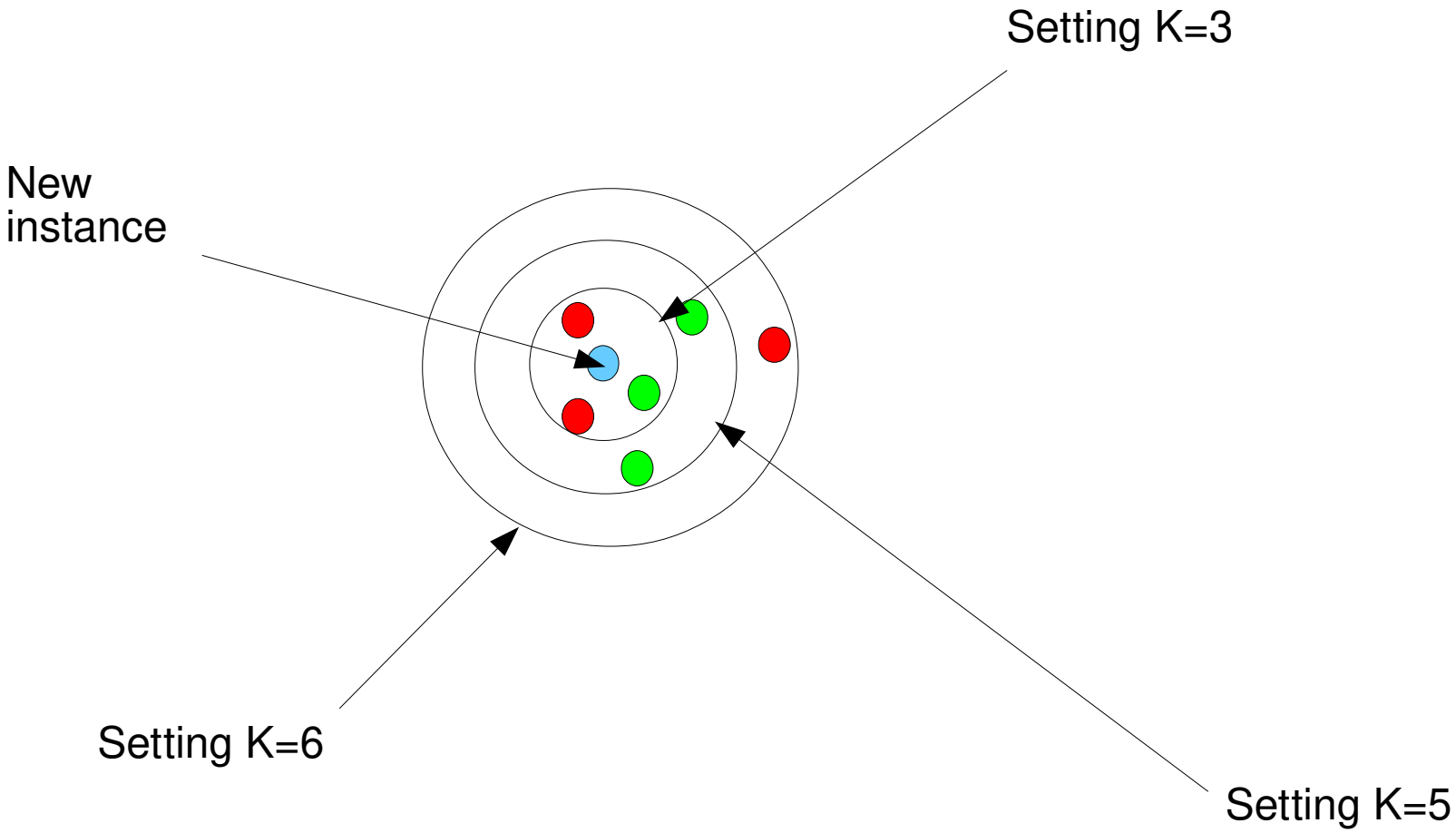
Distance-Weighted Nearest Neighbors

- 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:

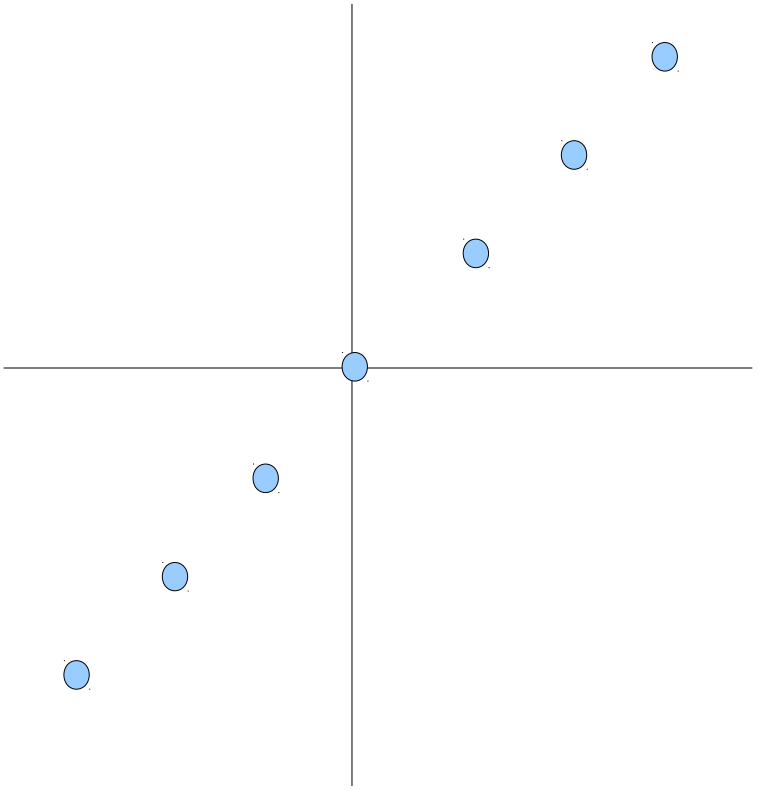
$$f(\mathbf{x}) = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i}$$

- Given the weight function:

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

Distance-Weighted Nearest Neighbors

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



Two main questions:

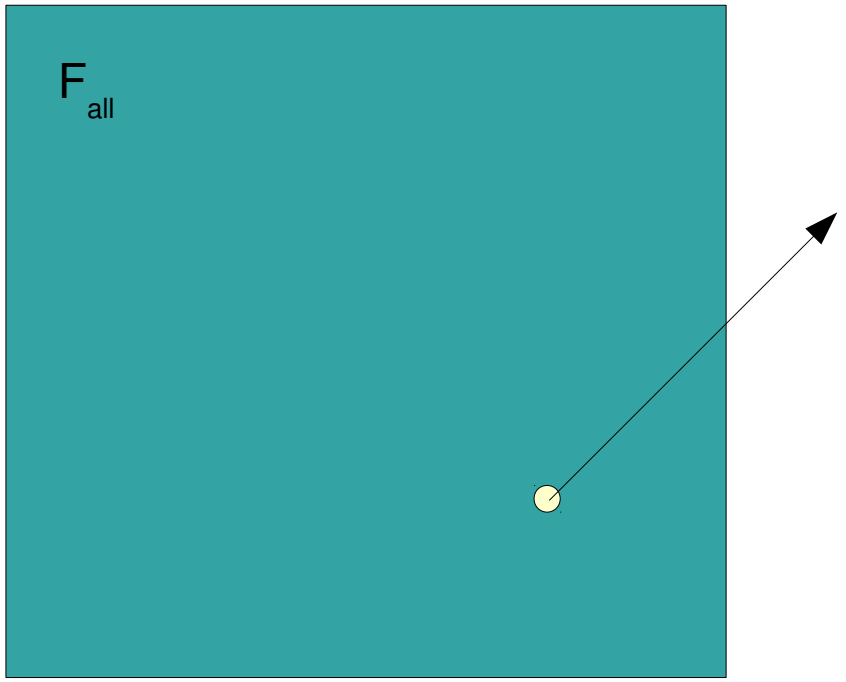
- What happpens if sigma is too big?

- 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

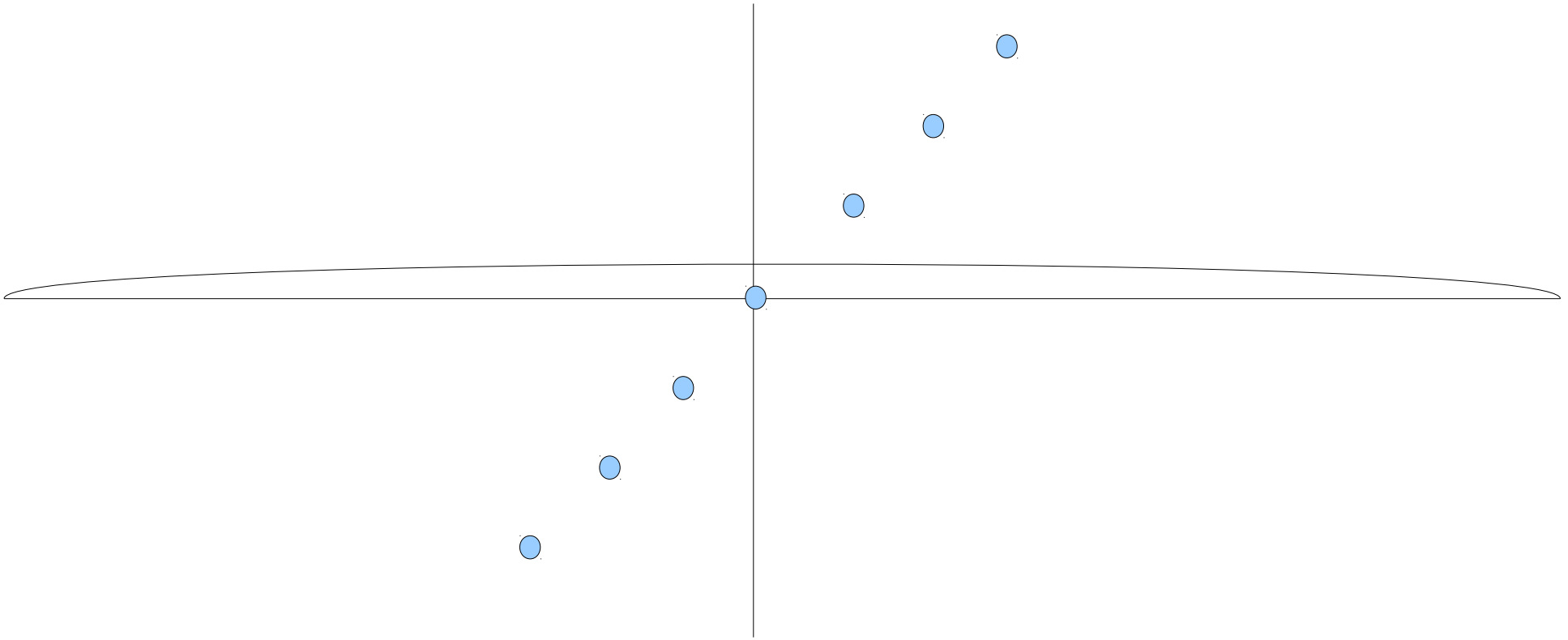


The space of admissible functions (bias) will contain a single function

In this case, the average function

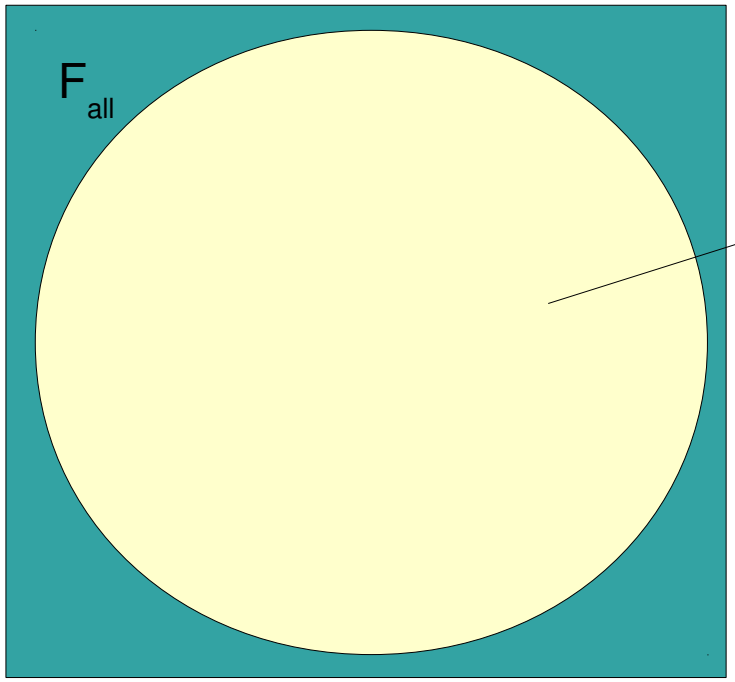
Distance-Weighted Nearest Neighbors

- When sigma tends to infinity



Distance-Weighted Nearest Neighbors

- When sigma tends to zero



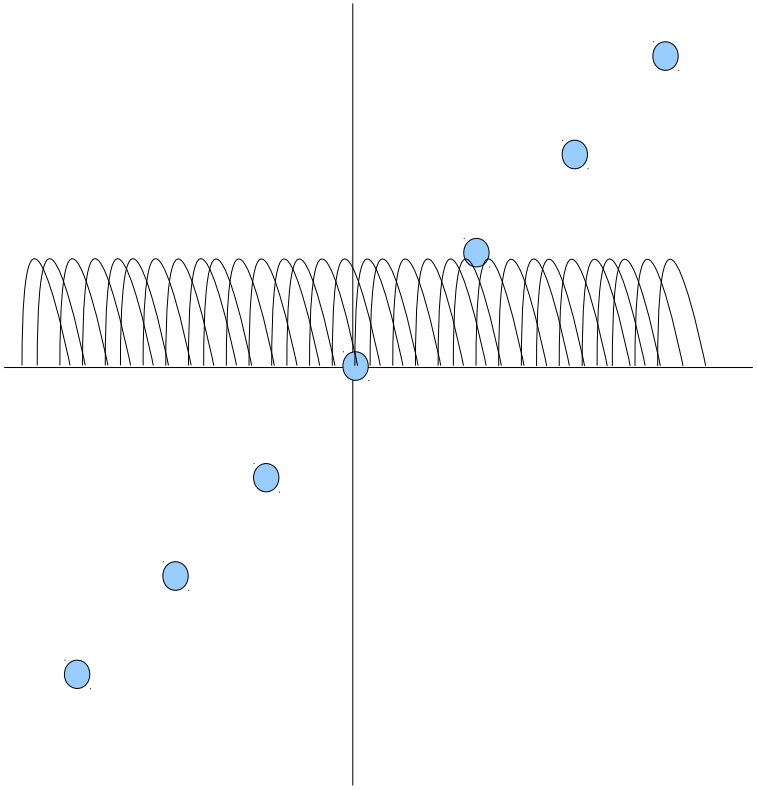
The space of admissible functions (bias) will tend to the whole space

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



References

- 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

- Book coming soon :)

References

Springer

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