Toward a Semi-Supervised Non-Intrusive Load Monitoring System for Event-based Energy Disaggregation

GlobalSIP 2015, 14th Dec.

Karim Said Barsim and Bin Yang

Institute of Signal Processing and System Theory

University of Stuttgart, Stuttgart, Germany

{karim.barsim,bin.yang}@iss.uni-Stuttgart.de

Introduction

- Semi-Supervised Learning (SSL): *leverage both labeled and unlabeled data.*

- Motivations
  - Labeled data are: scarce … costly… time-consuming
  - Unlabeled data are: plentiful … cheap … rapidly growing

- Advantages
  - Improved performance with reduced labeling efforts.
  - NILM systems learning over time.

- Requirements / Limitations:
  - Cluster assumption: *decision boundaries through sparse regions only !*
  - Manifold assumption: *same labels are close in geometry !*
  - Lazy training (transductive SSL models).

Supervised machine learning
- *Eager learning,*
- *Labeled data*

Semi-Supervised Learning (SSL)

Inductive learning
- *Eager learning,*
- *Labeled & unlabeled data*

Transductive learning
- *Lazy learning,*
- *Labeled & unlabeled data*

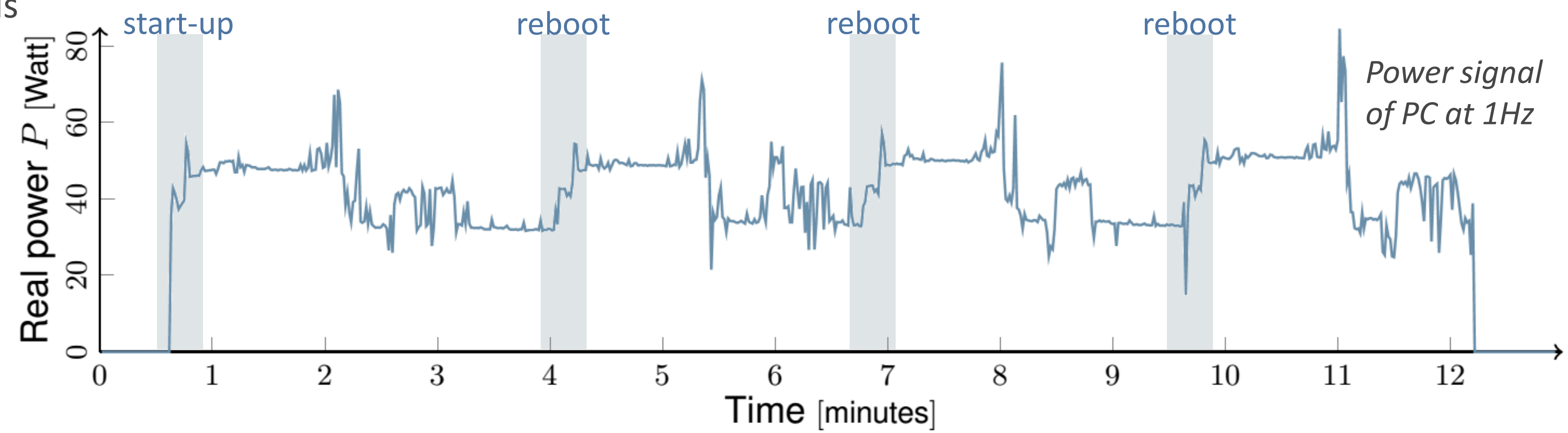
Constrained clustering
- *Lazy learning,*
- *Unlabeled data & labeling constraints*

Unsupervised machine learning
- *Lazy learning,*
- *Unlabeled data*

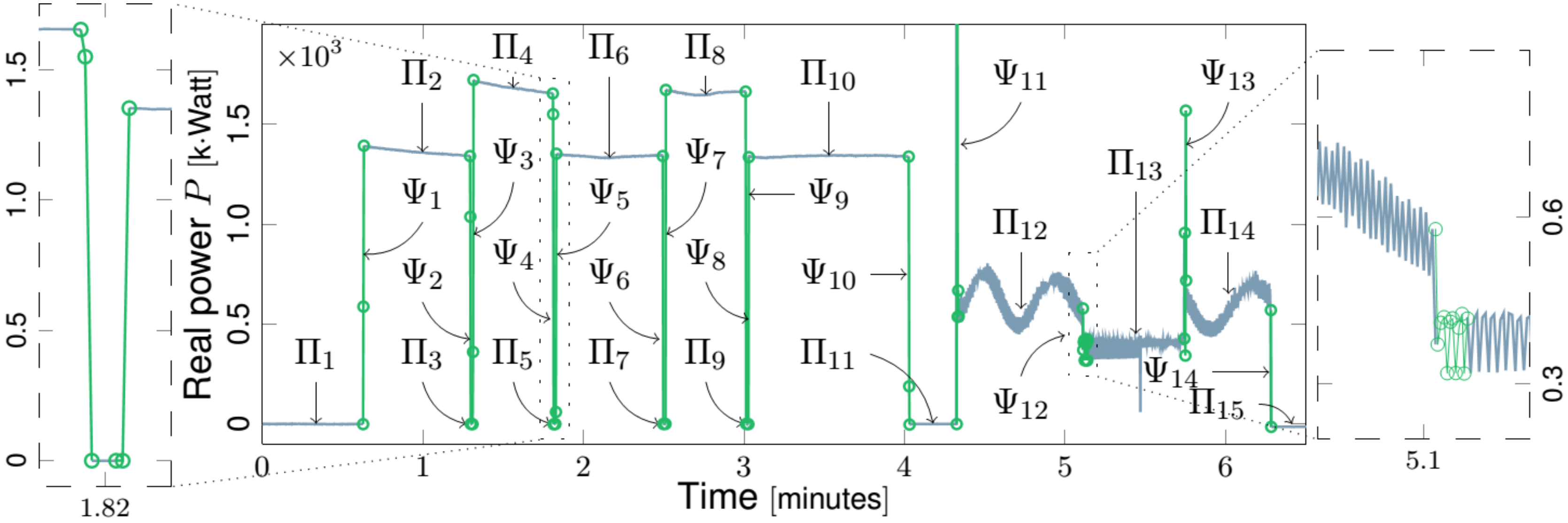
K. S. Barsim
Stuttgart University

Labeled & Unlabeled data

- Labeled data
  - Sub-metered loads (eventless)
  - Labeled events (event-based)



- Unlabeled data
  - Aggregate signals (eventless NILM)
  - Segmented signals (event-based NILM)

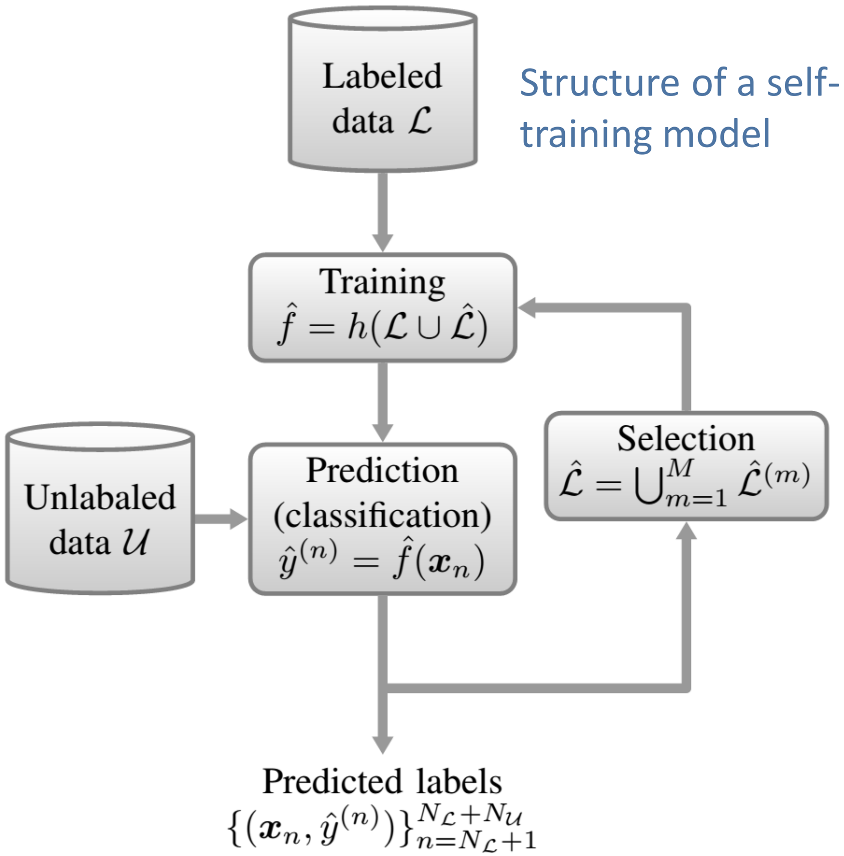


K. S. Barsim
Stuttgart University

Task and tools

- Is semi-supervised learning suitable for NILM systems ?

- An SSL model: *self-training*
  - Advantages
    - Simple SSL model
    - A wrapper model
    - Does not require unsupervised components
  - Requirements
    - A learning algorithm h or a seed classifier f^0
    - Confidence-rated predictions
  - Limitations
    - Separable data/classes



Structure of a self-training model

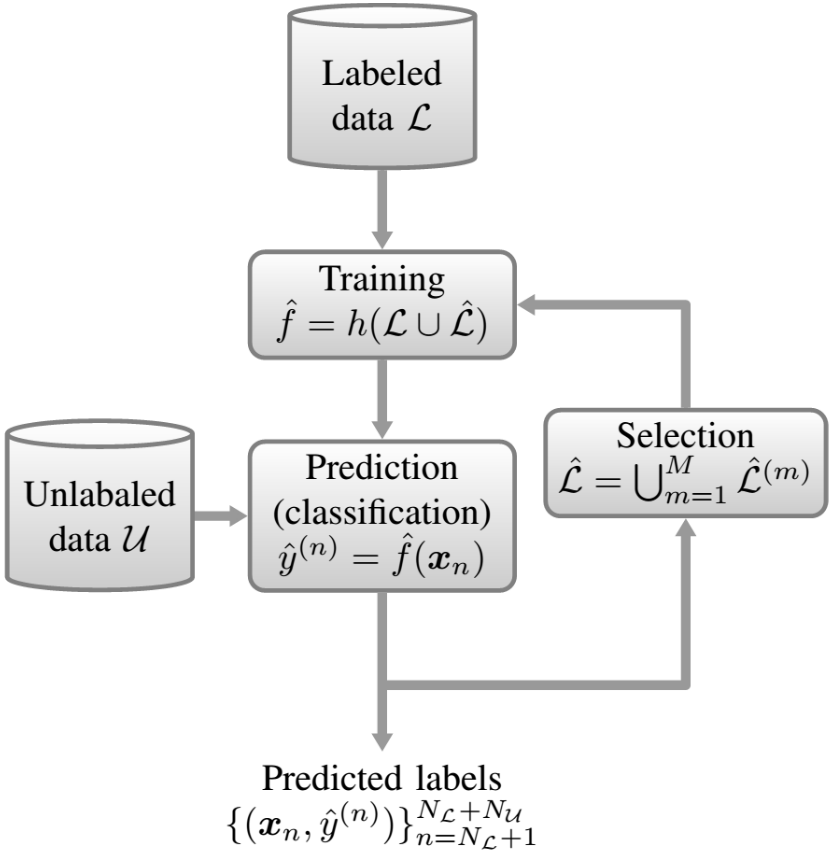
- NILM test dataset: BLUED[1]
  - Suitable for event-based NILM

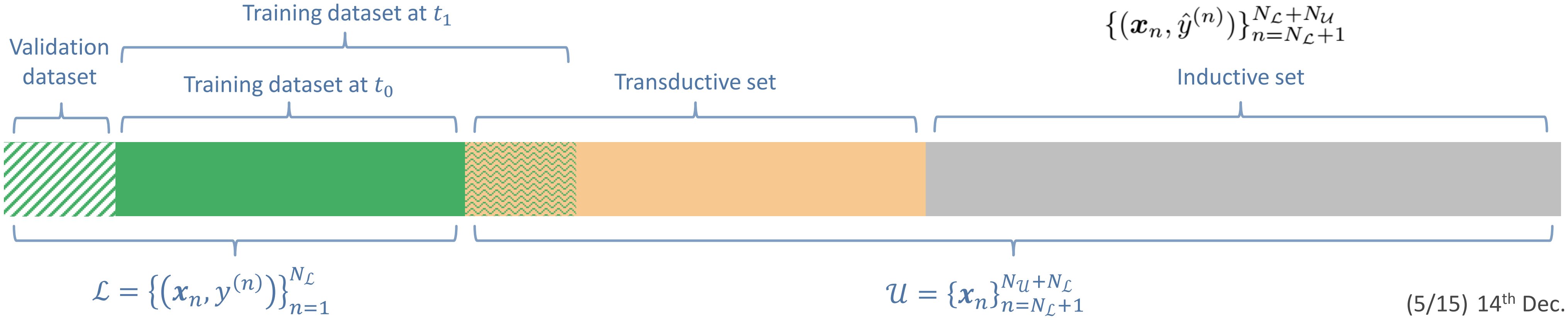
K. S. Barsim
Stuttgart University

[1] K. Anderson, A. Ocneanu, D. Benitez, D. Carlson, A. Rowe, and M. Berges, “BLUED: a fully labeled public dataset for Event-Based Non-Intrusive load monitoring research,” in Proceedings of the 2nd KDD Workshop on Data Mining Applications in Sustainability (SustKDD), Beijing, China, Aug. 2012

Self-training 1

- Labeled dataset: $\mathcal{L} = \{(\boldsymbol{x}_n, y^{(n)})\}_{n=1}^{N_\mathcal{L}}$
- Unlabeled dataset: $\mathcal{U} = \{\boldsymbol{x}_n\}_{n=N_\mathcal{L}+1}^{N_\mathcal{U}+N_\mathcal{L}}$
- Training: $\hat{f}^{(t)} = h(\mathcal{L} \cup \hat{\mathcal{L}})$
- Prediction: $\hat{y}^{(n)} = \hat{f}^{(t)}(\boldsymbol{x}_n), \qquad \boldsymbol{x}_n \in \mathcal{U}$
- Selection: $\hat{\mathcal{L}} = \mathrm{Sel}\left(\{(\boldsymbol{x}_n, \hat{y}^{(n)})\}_{n=}^{N_a+N_b}\right)$

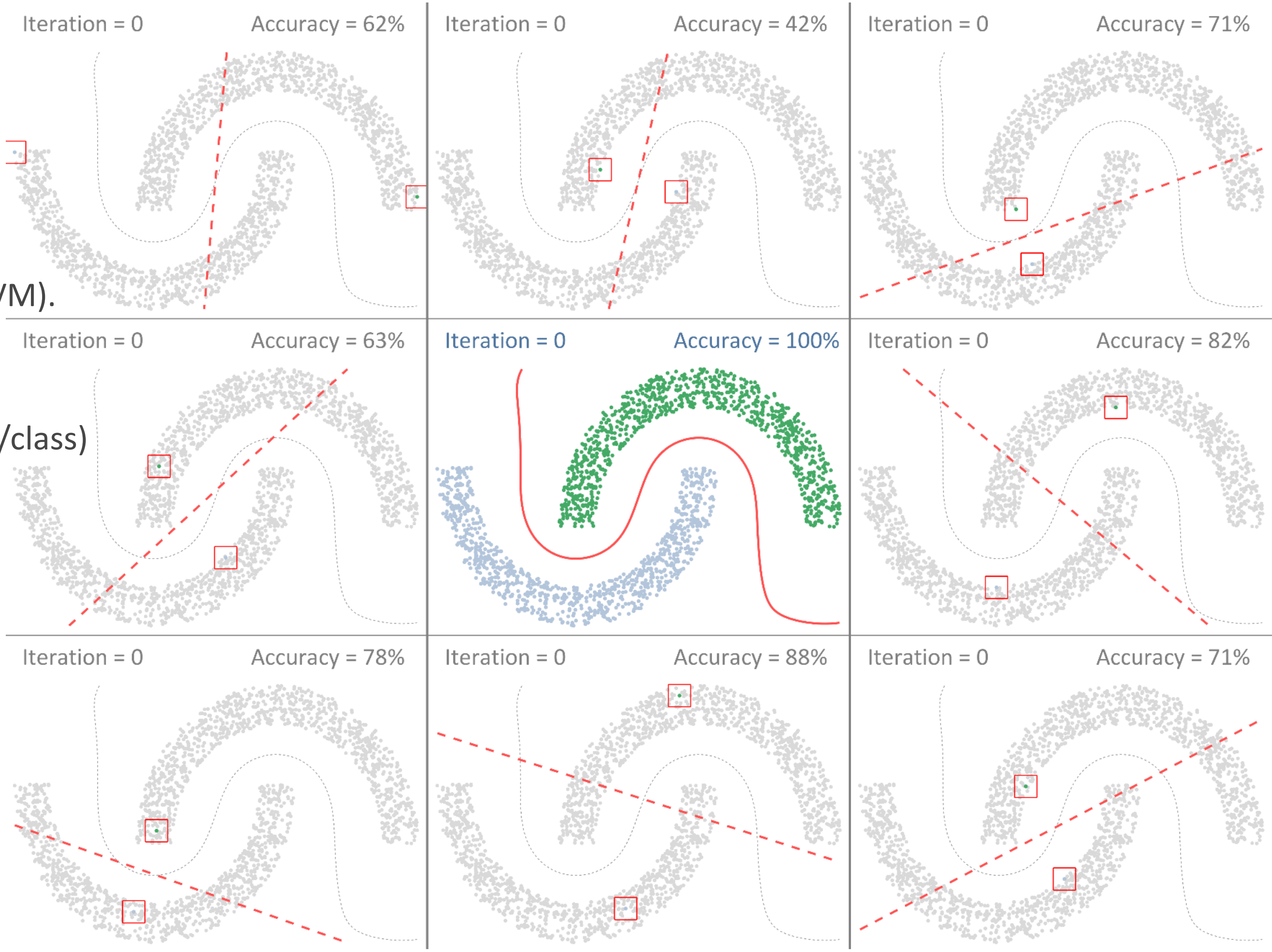




K. S. Barsim
Stuttgart University

Self-training 2: the double crescent problem

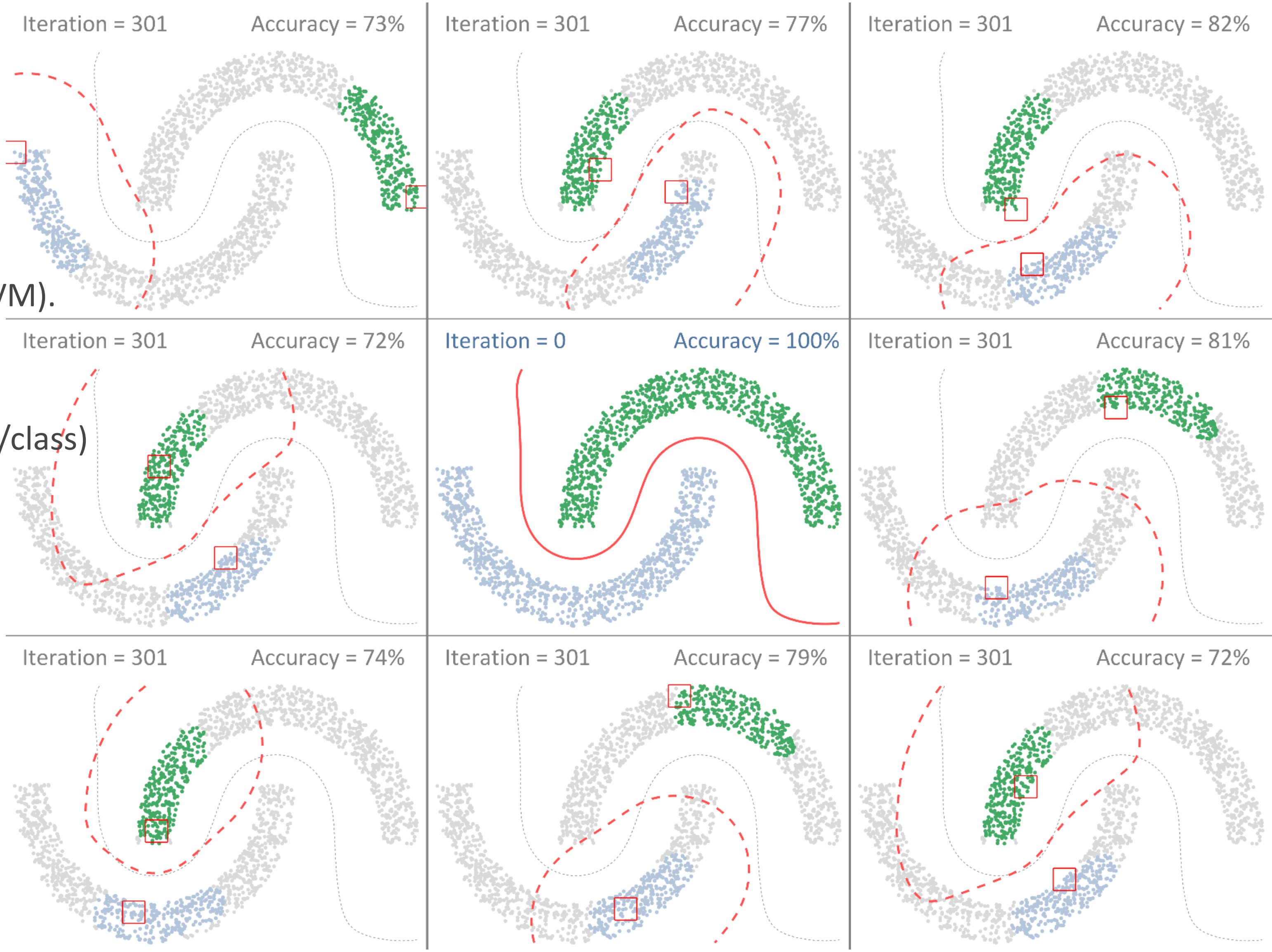
- Classification problem:
  - The double moon (2-class).
  - 1000 samples/class.
- Learner/Classifier:
  - Support Vector Machine (SVM).
  - Gaussian kernel $e^{-\|x_i - x_j\|^2}$
  - Minimal labelling (1 sample/class)
- Selection:
  - Farthest from boundary.



K. S. Barsim
Stuttgart University

Self-training 2: the double crescent problem

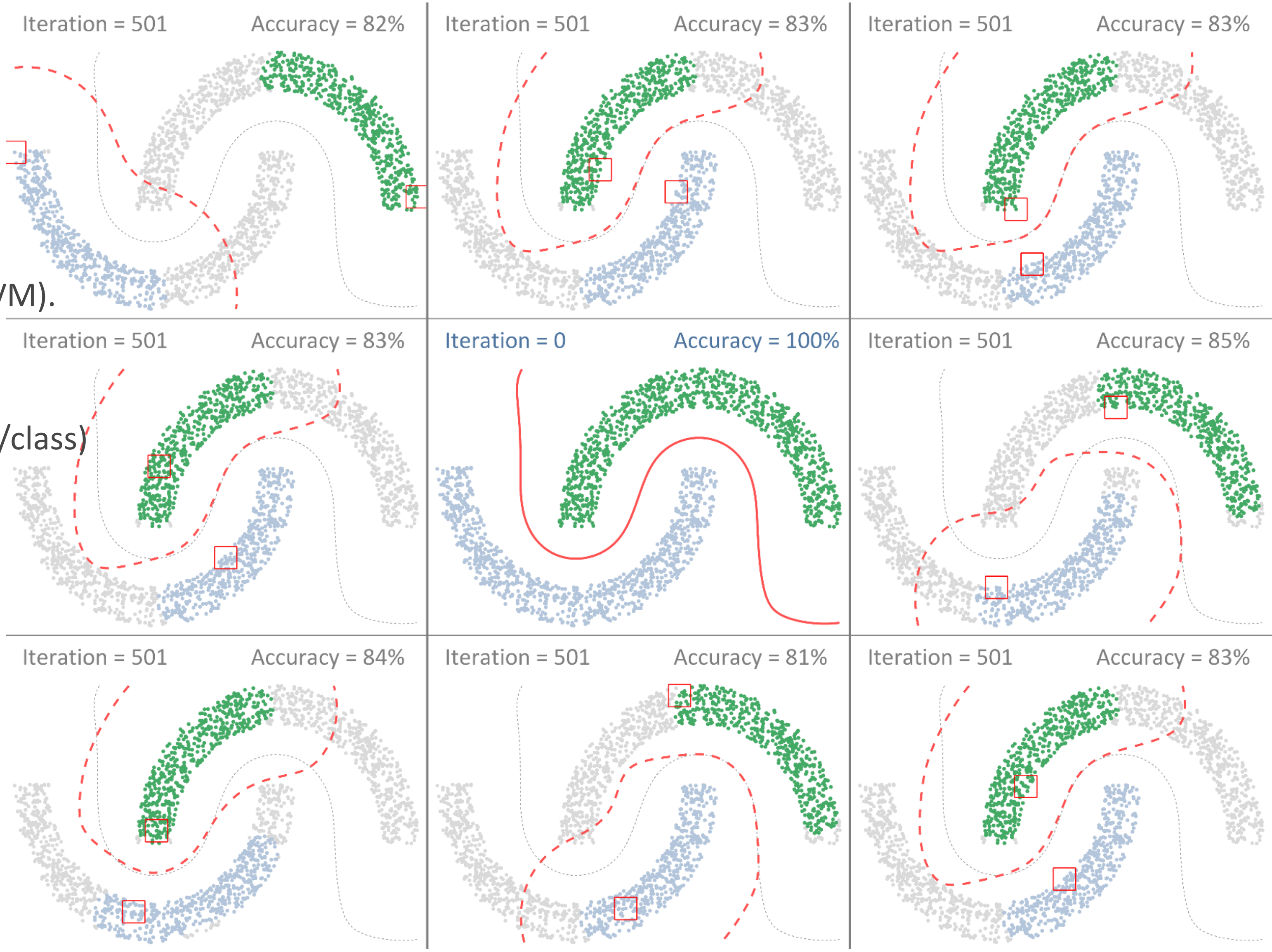
- Classification problem:
  - The double moon (2-class).
  - 1000 samples/class.
- Learner/Classifier:
  - Support Vector Machine (SVM).
  - Gaussian kernel $e^{-\|x_i - x_j\|^2}$
  - Minimal labelling (1 sample/class)
- Selection:
  - Farthest from boundary.
- 300 Iterations: > 70%



K. S. Barsim
Stuttgart University

Self-training 2: the double crescent problem

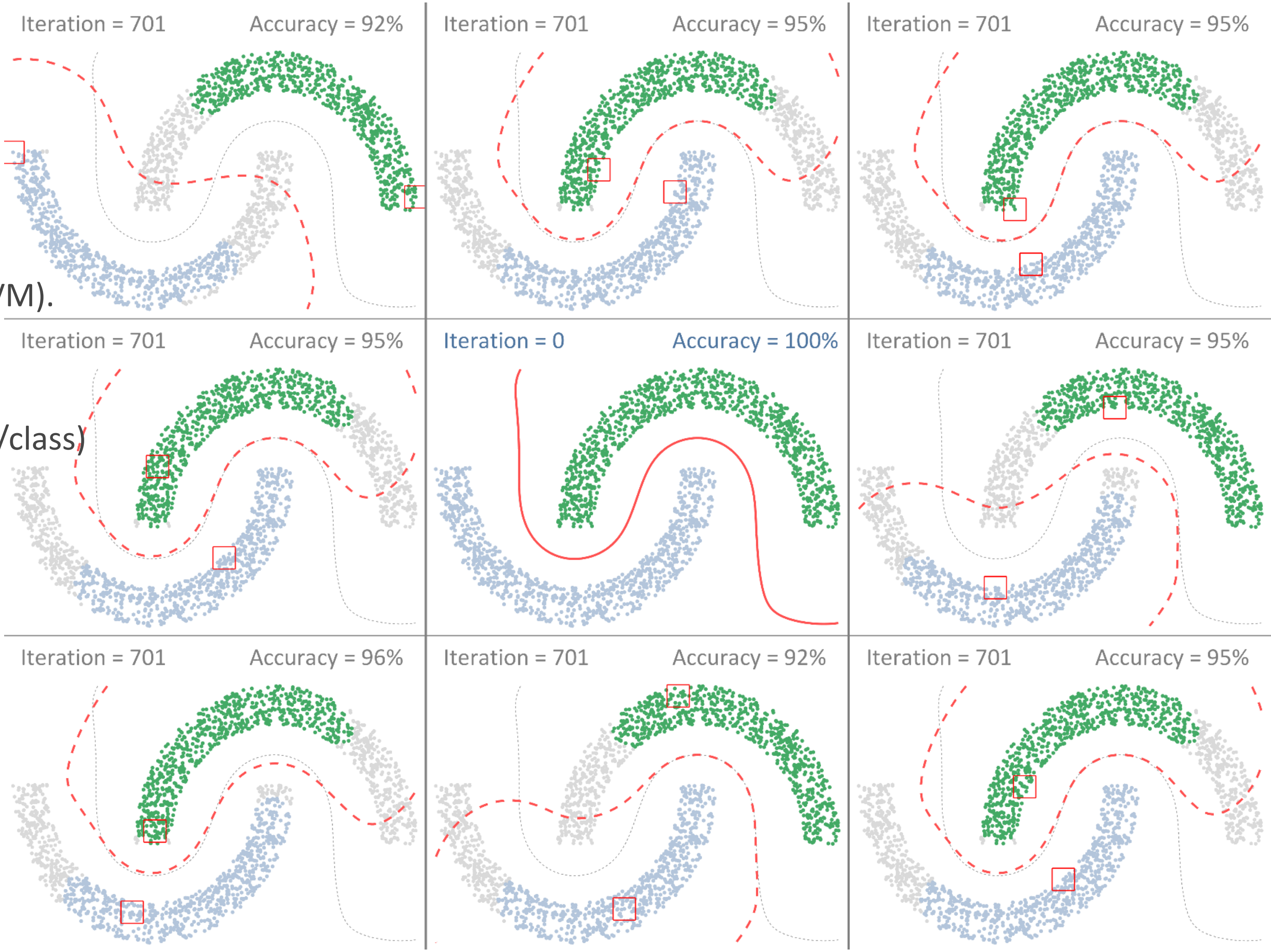
- Classification problem:
  - The double moon (2-class).
  - 1000 samples/class.
- Learner/Classifier:
  - Support Vector Machine (SVM).
  - Gaussian kernel $e^{-\|x_i - x_j\|^2}$
  - Minimal labelling (1 sample/class)
- Selection:
  - Farthest from boundary.
- 300 Iterations: $> 70\%$
- 500 Iterations: $> 80\%$



K. S. Barsim
Stuttgart University

Self-training 2: the double crescent problem

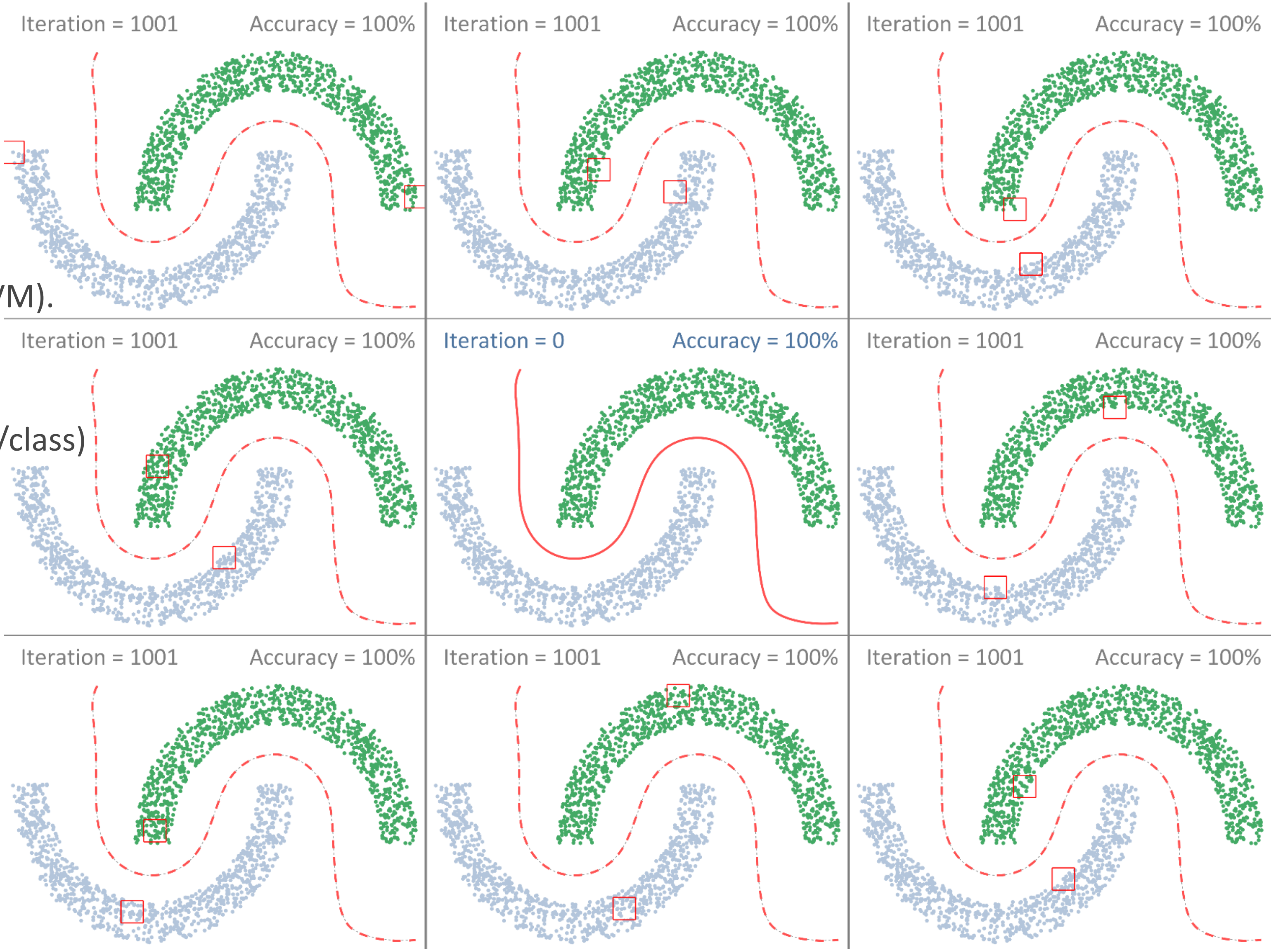
- Classification problem:
  - The double moon (2-class).
  - 1000 samples/class.
- Learner/Classifier:
  - Support Vector Machine (SVM).
  - Gaussian kernel $e^{-\|x_i - x_j\|^2}$
  - Minimal labelling (1 sample/class)
- Selection:
  - Farthest from boundary.
- 300 Iterations: $> 70\%$
- 500 Iterations: $> 80\%$
- 700 Iterations: $> 90\%$



K. S. Barsim
Stuttgart University

Self-training 2: the double crescent problem

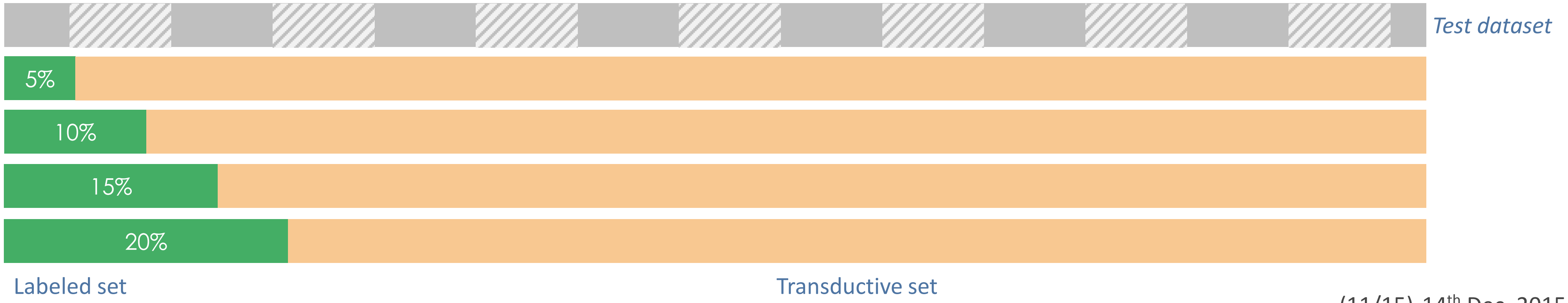
- Classification problem:
  - The double moon (2-class).
  - 1000 samples/class.
- Learner/Classifier:
  - Support Vector Machine (SVM).
  - Gaussian kernel $e^{-\|x_i - x_j\|^2}$
  - Minimal labelling (1 sample/class)
- Selection:
  - Farthest from boundary.
- 300 Iterations: $> 70\%$
- 500 Iterations: $> 80\%$
- 700 Iterations: $> 90\%$
- 1000 Iterations: *optimal !*



K. S. Barsim
Stuttgart University

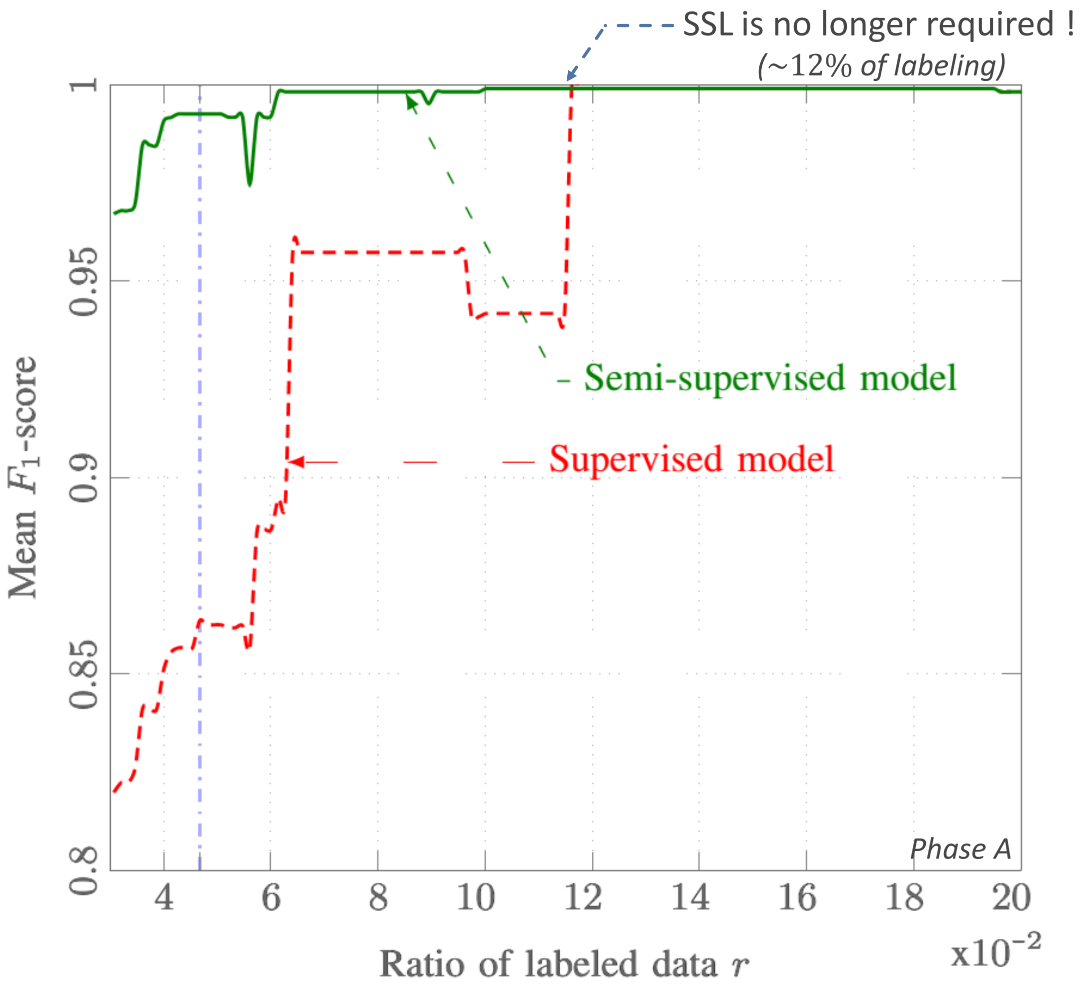
Transductive learning: how much labeling ?

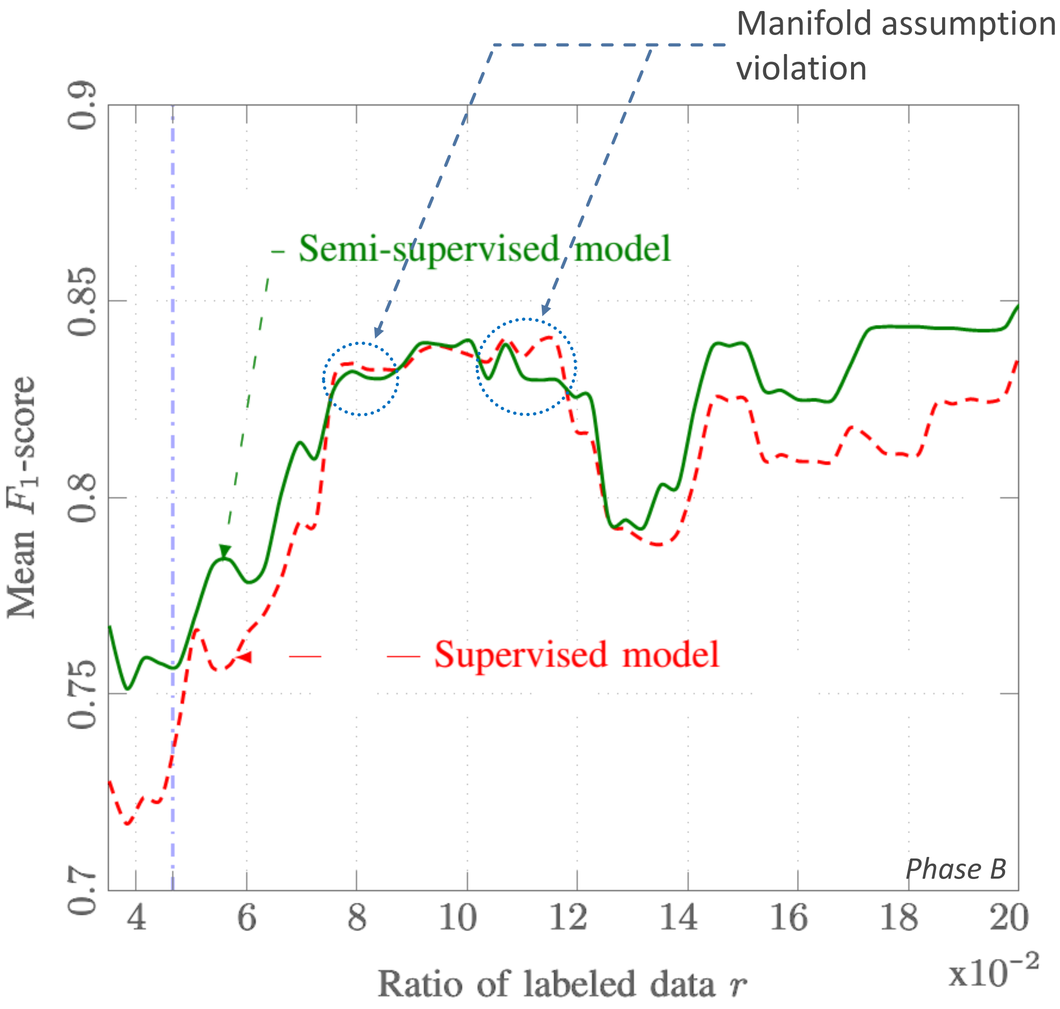
- Reduced labeling efforts: how much labeling is needed for near-optimal performance ?
- When should SSL replace purely supervised models ?
- **Object of classification:** *Event-based features ($[dP, dQ]^T$ feature vectors)*
- **Classifier:** *Support Vector Machine (SVM) with a linear kernel*
- **Selection:** *nearest to class mean (based on the labeled samples)*
- *1 sample per class per iteration, 3 iterations*
- **Dataset:** *BLUED dataset (refined)*
  - Phase A: 749 samples, 23 classes
  - Phase B: 1284 samples, 45 classes



Transductive learning: how much labeling ?

- Reduced labeling efforts: how much labeling is needed for near-optimal performance ?
- When should SSL replace purely supervised models ?

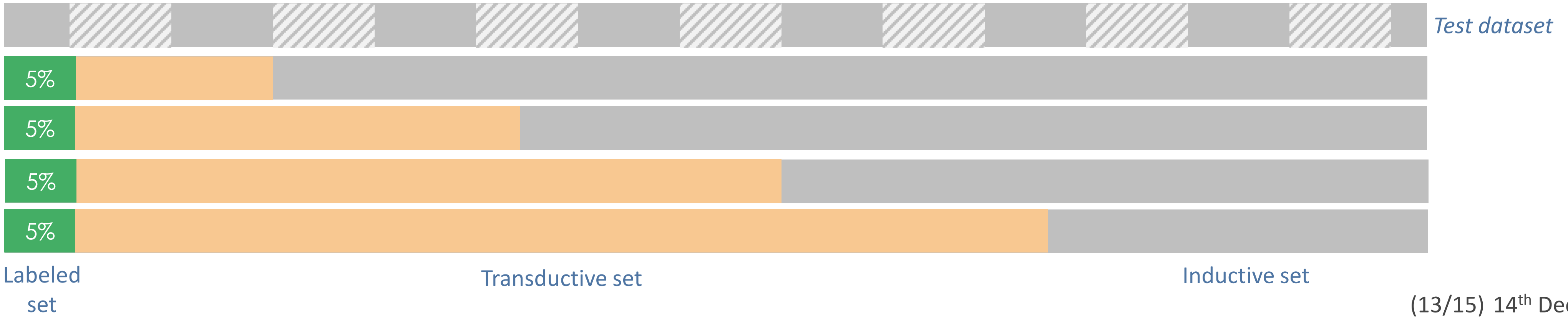




K. S. Barsim
Stuttgart University

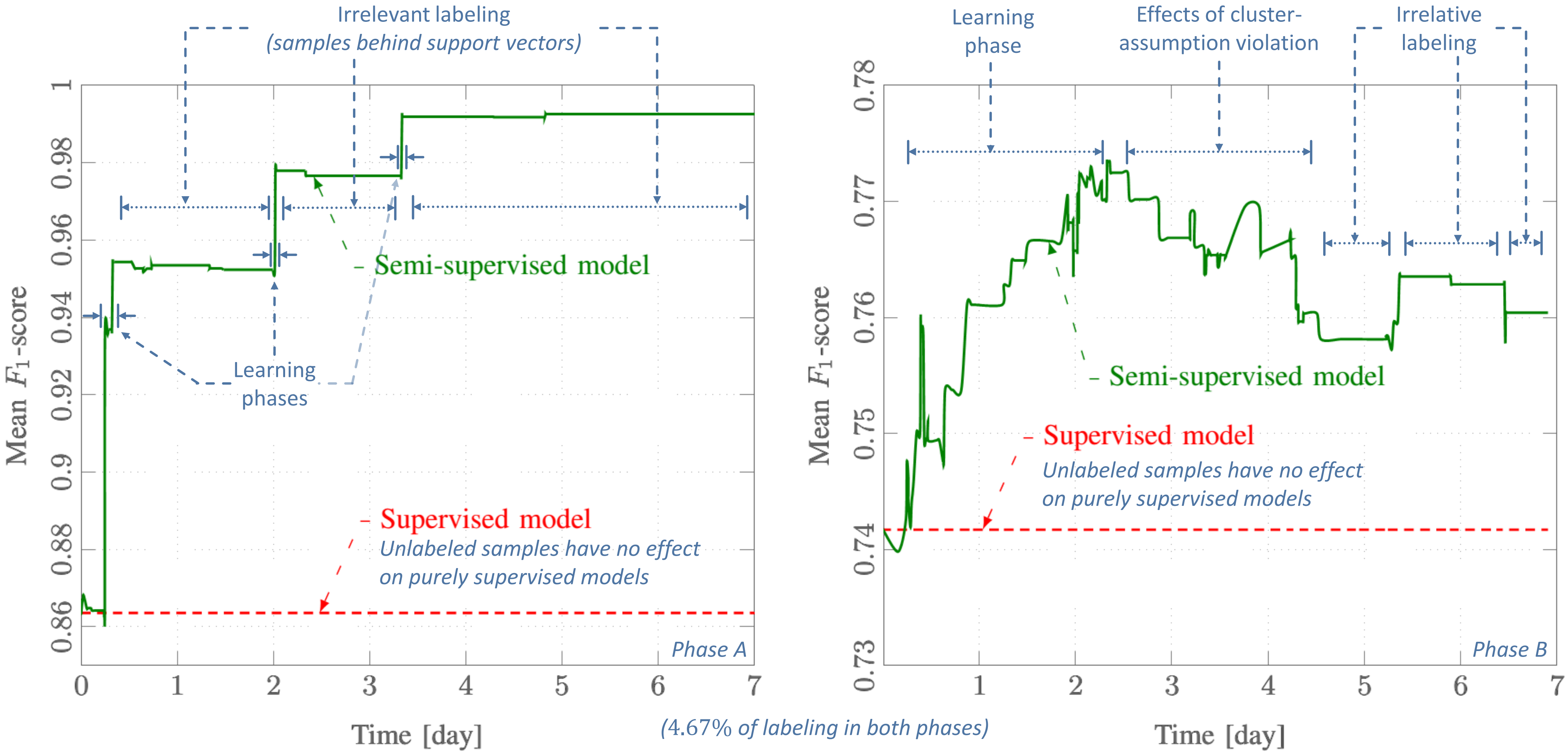
Inductive learning: learning over time ?

- Effect of increasing unlabeled dataset.
- Test dataset is fixed and includes inductive and transductive inference tests.
- **Object of classification:** *Event-based features ($[dP, dQ]^T$ feature vectors)*
- **Classifier:** *Support Vector Machine (SVM) with a linear kernel*
- **Selection:** *nearest to class mean (based on the labeled samples)*
- *1 sample per class per iteration, 3 iterations*
- **Dataset:** *BLUED dataset (refined)*
  - Phase A: 749 samples, 23 classes
  - Phase B: 1284 samples, 45 classes



Inductive learning: learning over time ?

- Effect of increasing unlabeled dataset.
- Test dataset is fixed and includes inductive and transductive inference tests.



(4.67% of labeling in both phases)

K. S. Barsim
Stuttgart University

Discussion

Thank you
for your attention

K. S. Barsim
Stuttgart University