

List of Figures

- 3.1 MOA graphical user interface.
- 3.2 MOA GUI running two different tasks.
- 3.3 Exercise 3.1, comparing the Naive Bayes Classifier and the Hoeffding tree.
- 3.4 Exercise 3.2, comparing three different evaluation methods on the same classifier.
- 3.5 A sigmoid function $f(t) = 1/(1 + e^{-4(t-p)/w})$.
- 3.6 Exercise 3.3, comparing the Hoeffding tree and Naive Bayes classifiers on a nonstationary stream.
- 3.7 Exercise 3.4, comparing three different classifiers.

- 4.1 The RESERVOIR SAMPLING sketch.
- 4.2 Morris's approximate counting algorithm.
- 4.3 The LINEAR COUNTING sketch.
- 4.4 Cohen's counting algorithm.
- 4.5 The basic Flajolet-Martin probabilistic counter.
- 4.6 The SPACESAVING sketch.
- 4.7 Example CM-sketch structure of width 7 and depth 4, corresponding to $\epsilon = 0.4$ and $\delta = 0.02$.
- 4.8 The CM-SKETCH algorithm.
- 4.9 The COUNTSKETCH algorithm.
- 4.10 Partitioning a stream of 29 bits into buckets, with $k = 2$. Most recent bits are to the right.
- 4.11 The EXPONENTIAL HISTOGRAMS sketch.

- 5.1 Managing change with adaptive estimators. Figure based on [30].
- 5.2 Managing change with explicit change detectors for model revision. Figure based on [30]
- 5.3 Managing change with model ensembles.

- 6.1 Evaluation on a stream of 1,000,000 instances, comparing holdout, interleaved test-then-train, and sequential with sliding window evaluation methods.
- 6.2 A dataset describing email features for deciding whether the email is spam, and a decision tree for it.
- 6.3 The Hoeffding Tree algorithm.
- 6.4 The CVFDT algorithm.
- 6.5 Gaussian approximation of two classes. Figure based on [199].
- 6.6 Active learning framework.
- 6.7 Variable uncertainty strategy, with a dynamic threshold.
- 6.8 MOA graphical user interface.

- 7.1 The Weighted Majority algorithm.
- 7.2 Online Bagging for M models. The indicator function $I(\text{condition})$ returns 1 if the condition is true, and 0 otherwise.

- 9.1 The k -means clustering algorithm, or Lloyd's algorithm.
- 9.2 The DEN-STREAM algorithm.
- 9.3 The MOA *Clustering* GUI tab.
- 9.4 Evolving clusters in the GUI *Clustering* tab.

- 10.1 Example of an itemset dataset.
- 10.2 Frequent, closed, and maximal itemsets with minimum absolute support 3, or relative support 0.5, for the dataset in figure 10.1.
- 10.3 A general stream pattern miner with batch updates and sliding window.
- 10.4 The WINGRAPHMINER algorithm and procedure CORESET.
- 10.5 The ADAGRAPHMINER algorithm.

- 11.1 The ADAMS flow editor.
- 11.2 The ADAMS flow example.
- 11.3 The MEKA GUI.
- 11.4 Integration of OpenML with MOA.
- 11.5 Parallelism hint in SAMOA.

- 12.1 The MOA Graphical User Interface.
- 12.2 Options to set up a task in MOA.
- 12.3 Option dialog for the RBF data generator. By storing and loading settings, streaming datasets can be shared for benchmarking, repeatability and comparison.
- 12.4 Visualization tab of the MOA clustering GUI.

- 13.1 Rendering of learning curves of two classifiers using `gnuplot`.

This is a section of [doi:10.7551/mitpress/10654.001.0001](https://doi.org/10.7551/mitpress/10654.001.0001)

Machine Learning for Data Streams

with Practical Examples in MOA

By: Albert Bifet, Ricard Gavaldà, Geoffrey Holmes, Bernhard Pfahringer

Citation:

Machine Learning for Data Streams: with Practical Examples in MOA

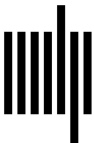
By: Albert Bifet, Ricard Gavaldà, Geoffrey Holmes, Bernhard Pfahringer

DOI: 10.7551/mitpress/10654.001.0001

ISBN (electronic): 9780262346047

Publisher: The MIT Press

Published: 2023



The MIT Press

© 2017 Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form or by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

This book was set in Times Roman and Mathtime Pro 2 by the authors.

Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data is available

ISBN: 978-0-262-03779-2

10 9 8 7 6 5 4 3 2 1