CONCEPT DRIFT

AN OVERVIEW

PAULO RICARDO LISBOA DE ALMEIDA

SUPERVISED TRAINING – CLASSICAL APPROACH

- Get a set of labeled training data
  - The environment is static
  - The underlying distribution is assumed to be the same in the environment that the classifier will be employed
- Generate a classifier using the training data
  - The training dataset is assumed to have all information necessary to learn the “concept”
- Employ the classifier to label unknown instances
  - The classifier is representative in all environments
  - The classifier can be employed for an undefined period of time

THE CONCEPT DRIFT PROBLEM

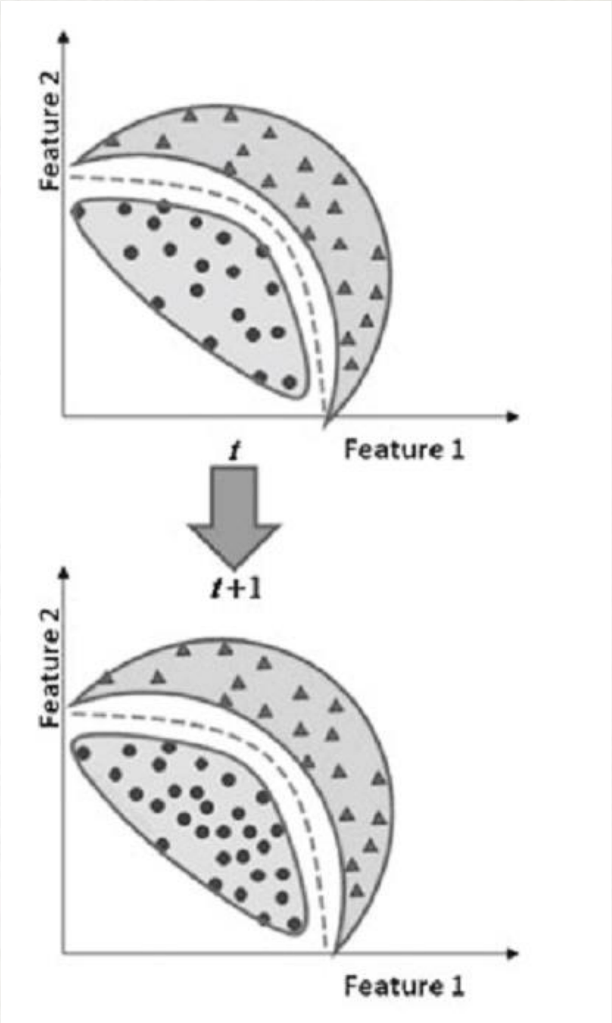
- The distributions stationarity is unrealistic for many real world problems
  - The problem evolves and changes over time, making the classifier obsolete
  - Examples: spam detection, fraud detection, climate analysis,…
- The boundaries generated by the classifier does not match with the current environment
  - A Concept Drift has occurred
  - Some measure need to be taken to adapt to the new concept
    - Retrain the classifier, train a new classifier

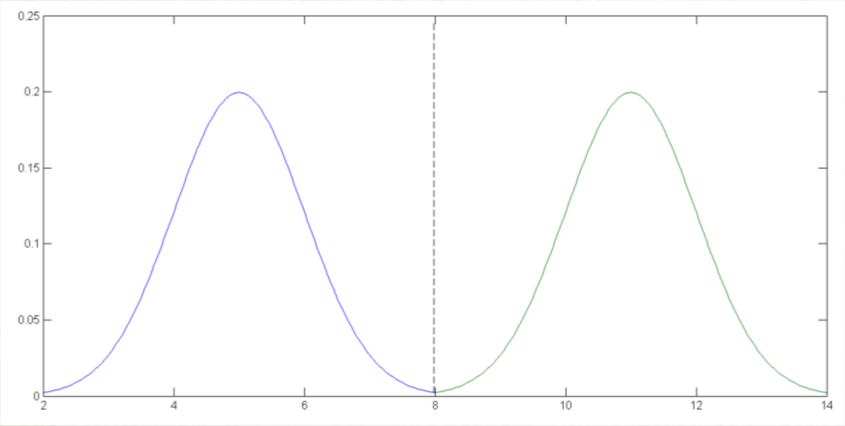
TYPES OF DRIFT

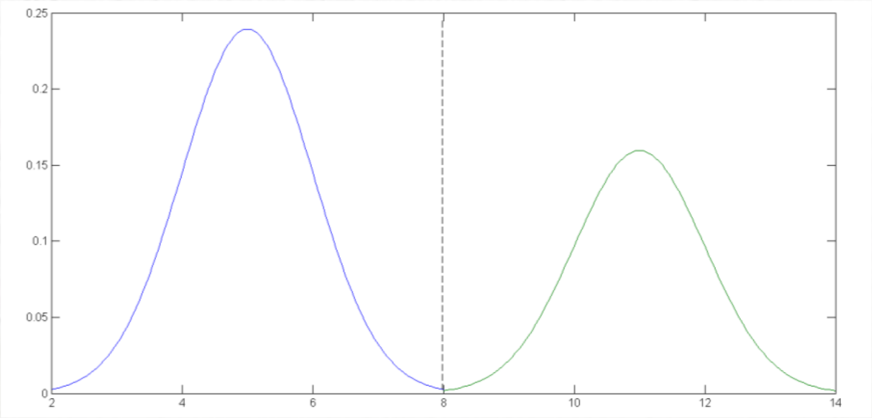
- Consider
  - A set of features x.
  - A target (class) variable y.
  - The a priori probabilities $P(x, y)$
  - The a posteriori probabilities $P(y, x)$

VIRTUAL DRIFT

- Also known as population drift
- Change in Class Priors
  - Change in the class distributions
  - $P_t(y) \neq P_{t+1}(y)$
  - Cost-sensitive learning

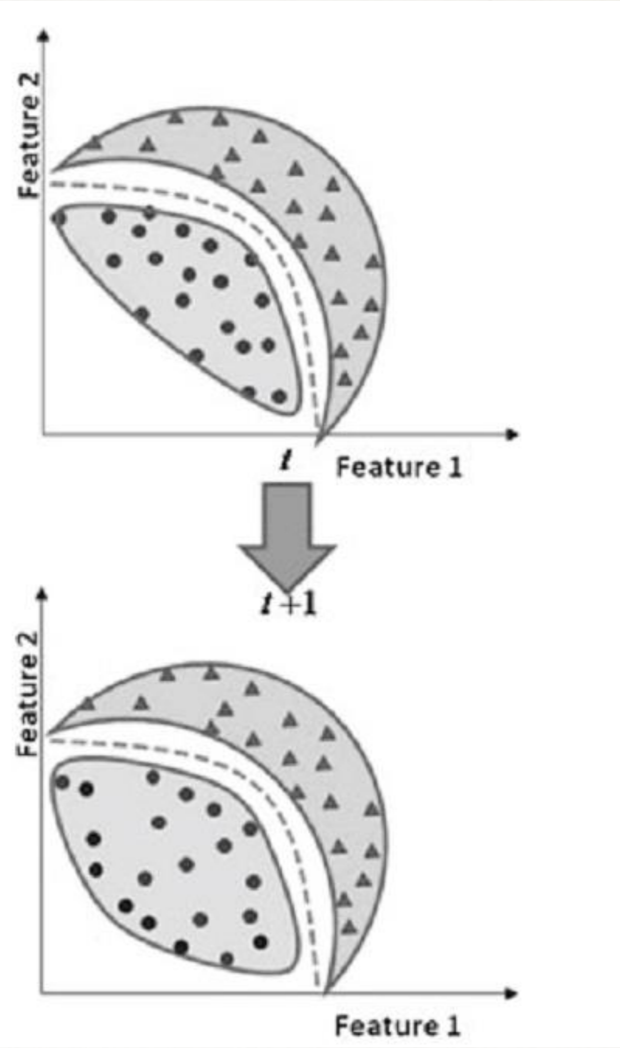


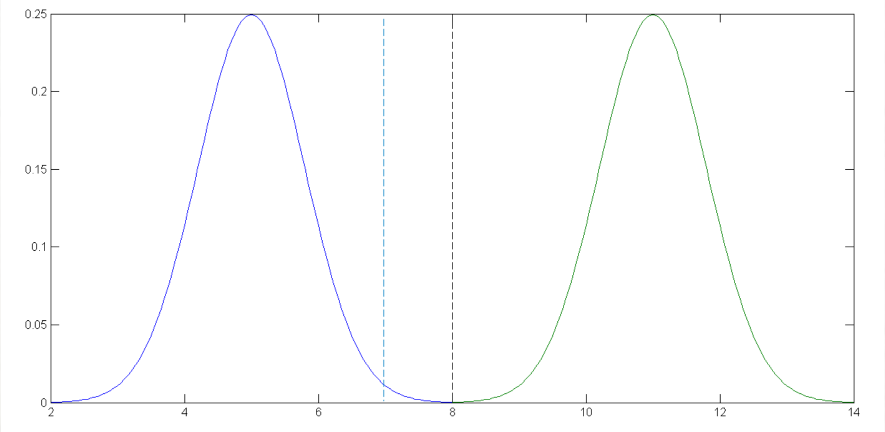


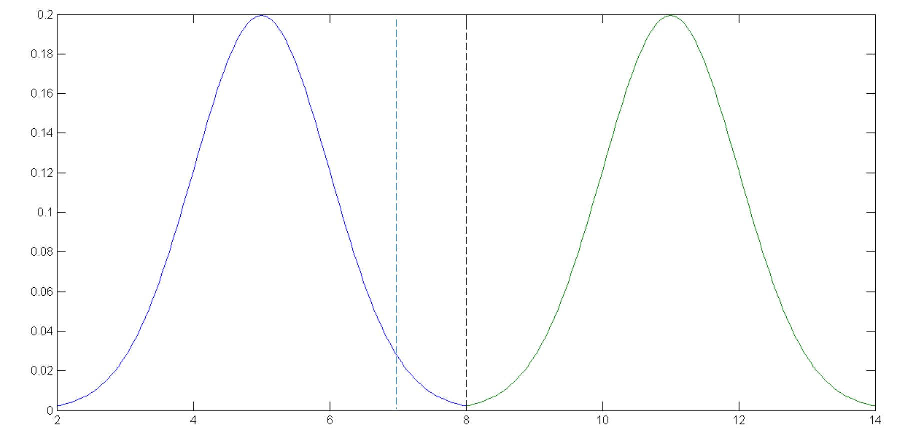


VIRTUAL DRIFT

- Change in the distribution
  - $P_t(y|x) = P_{t+1}(y|x)$
  - $P_t(x) \neq P_{t+1}(x)$
  - The boundary is “extended”, but still valid







VIRTUAL DRIFT – AN EXAMPLE

- Consider the problem of classifying parking spaces
  - Supervised training using images
- After training the classifier is employed
  - A virtual concept drift may happen if
    - The camera angle changes
    - The parking lot changes
    - Illumination changes

Training

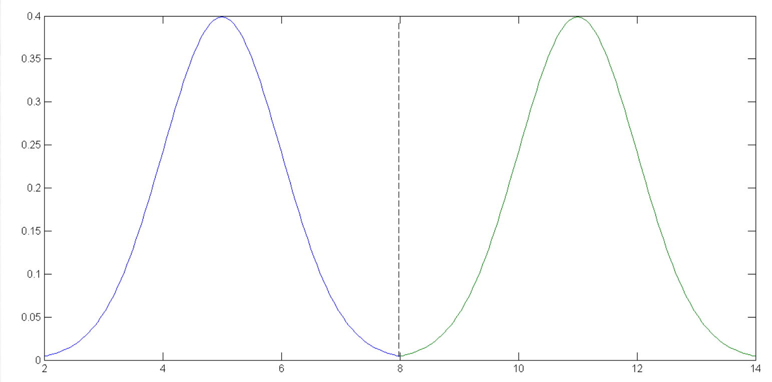
Testing

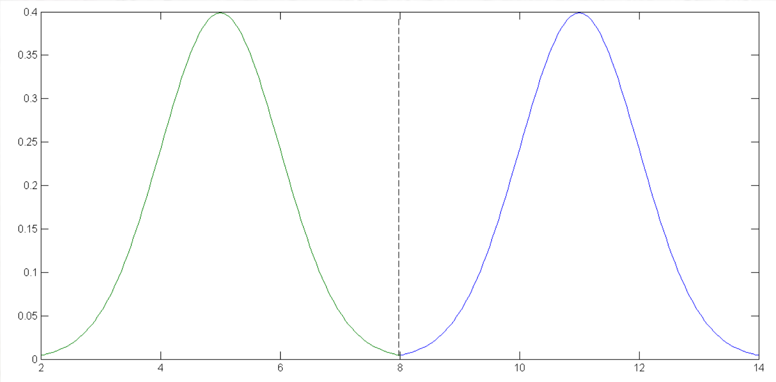
REAL CONCEPT DRIFT

- The relationship between the feature vectors and the target class changes
  - $P_t(y|x) \neq P_{t+1}(y|x)$
  - The boundary’s shape changes over time
  - Considered the most challenging drift
  - The classifier knowledge becomes irrelevant
    - The classifier could become a worse option than choosing at random

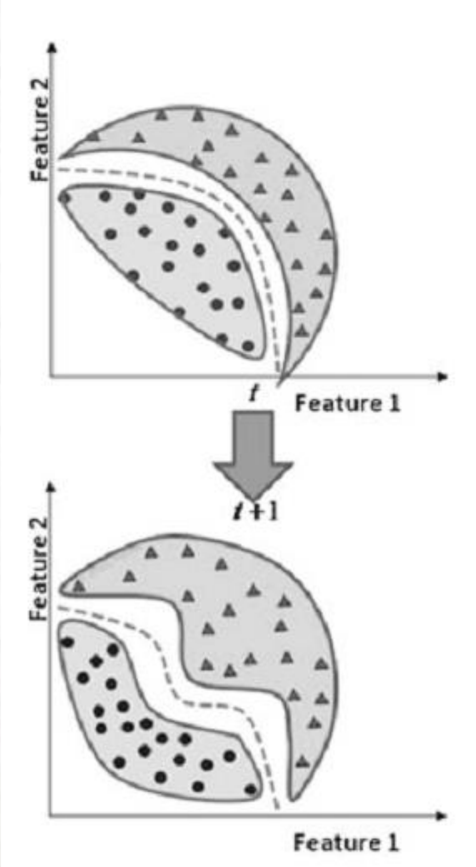
REAL CONCEPT DRIFT

Class swap





Gradual Drift



REAL CONCEPT DRIFT – AN EXAMPLE

- Consider the problem of classifying the user's e-mail
  - The e-mails would be classified as interesting or not for today
- The user's interests changes over time
  - The change is unpredictable
- Something that was interesting yesterday, is not today
  - Ex.: Yesterday the user was interested in the subject "computers", but today he is interested in "concept drift"

THE STABILITY PLASTICITY DILEMMA

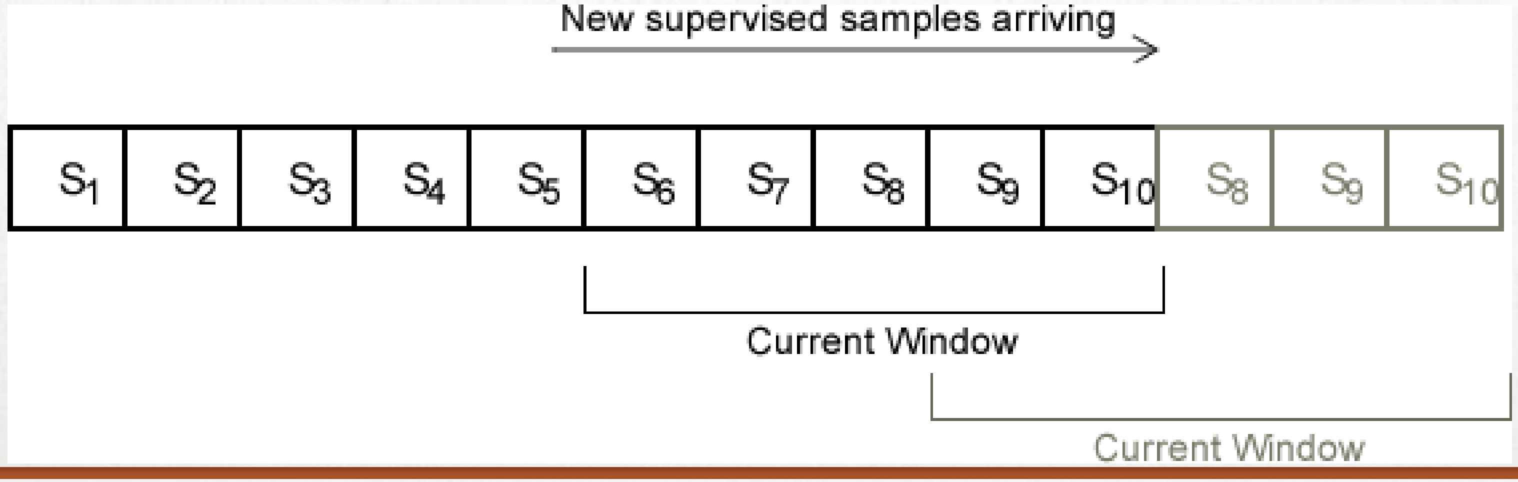
- How a learning system can be designed to be
  - Stable to irrelevant events (noise and outliers)
  - Able to learn new concepts (plasticity)

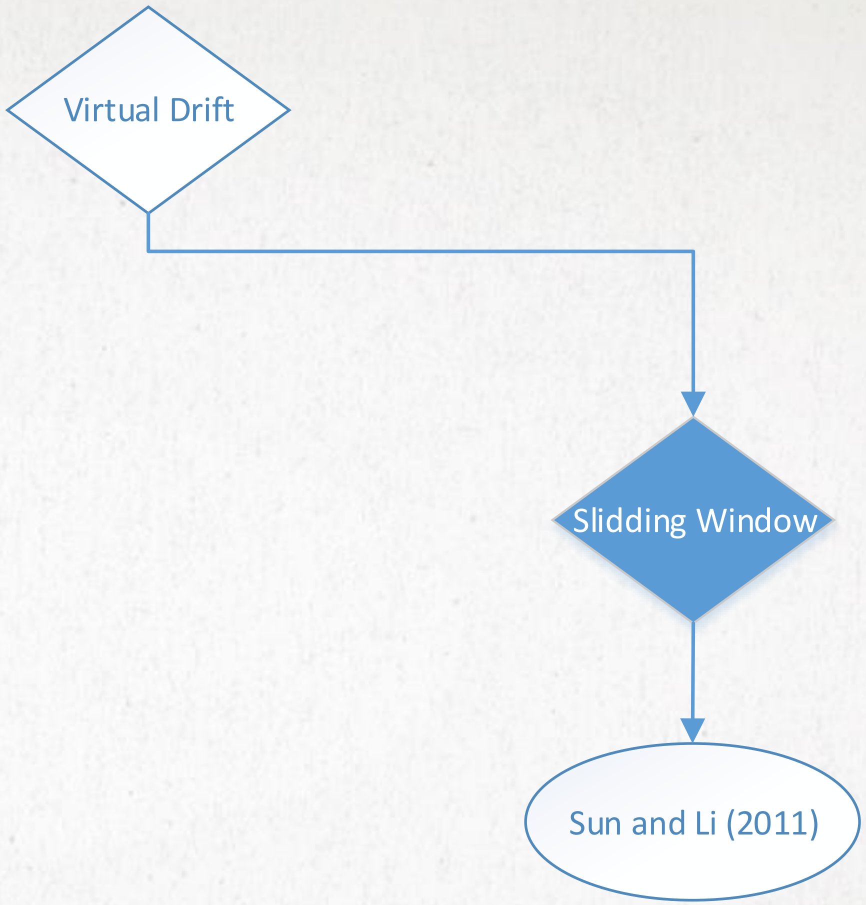
DEALING WITH CONCEPT DRIFT

- Authors considers that new supervised data will be available over time
  - Examples
    - The user will classify some e-mails when he changes his subject of interest (Just a few examples are available)
    - All real labels of the climate predictions for the next year will become available in the next year
- Forms the new supervised data arrives
  - In a stream.
    - Ex.: A new supervised instance for every hour
  - In batches containing several supervised instances.
    - Ex.: 1.000 supervised instances at every new month

DEALING WITH CONCEPT DRIFT

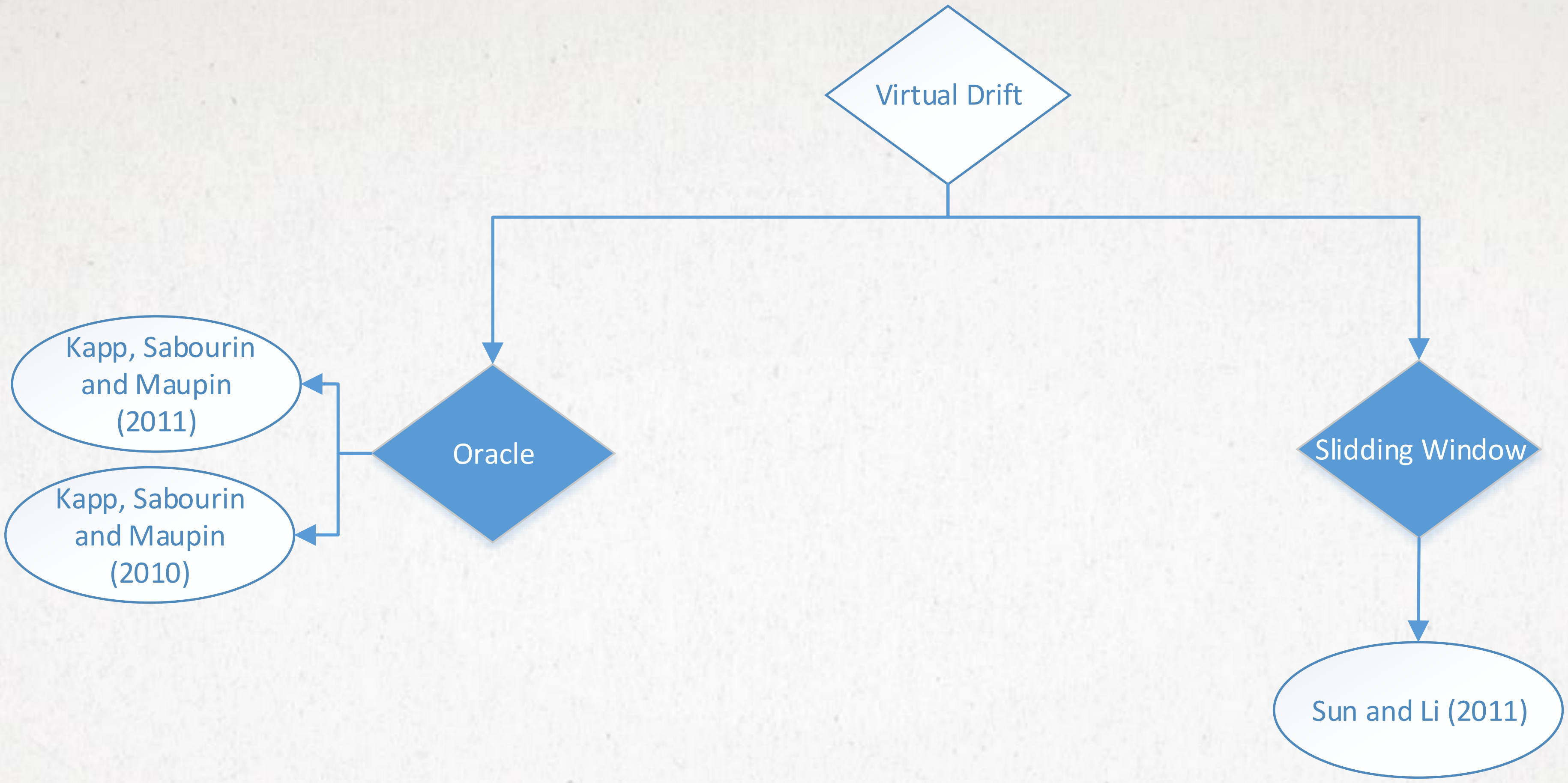
- Windowing
  - Passive method
  - Train the classifier using the N latest supervised instances
    - Greater values of N gives a better stability
    - Smaller values of N gives a better plasticity
  - How to define the window size?




Virtual Drift
Slidding Window
Sun and Li (2011)

DEALING WITH CONCEPT DRIFT

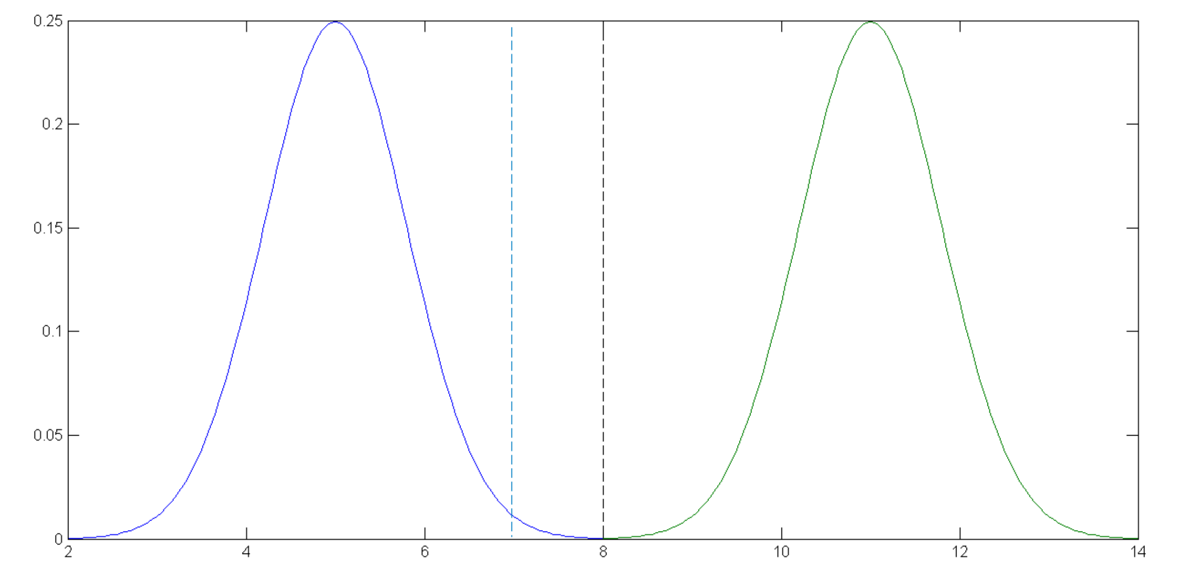
- Active Methods
  - Methods that actively tries to detect a drift and recover from it
  - Oracle
- False alarms versus undetected/delayed drifts

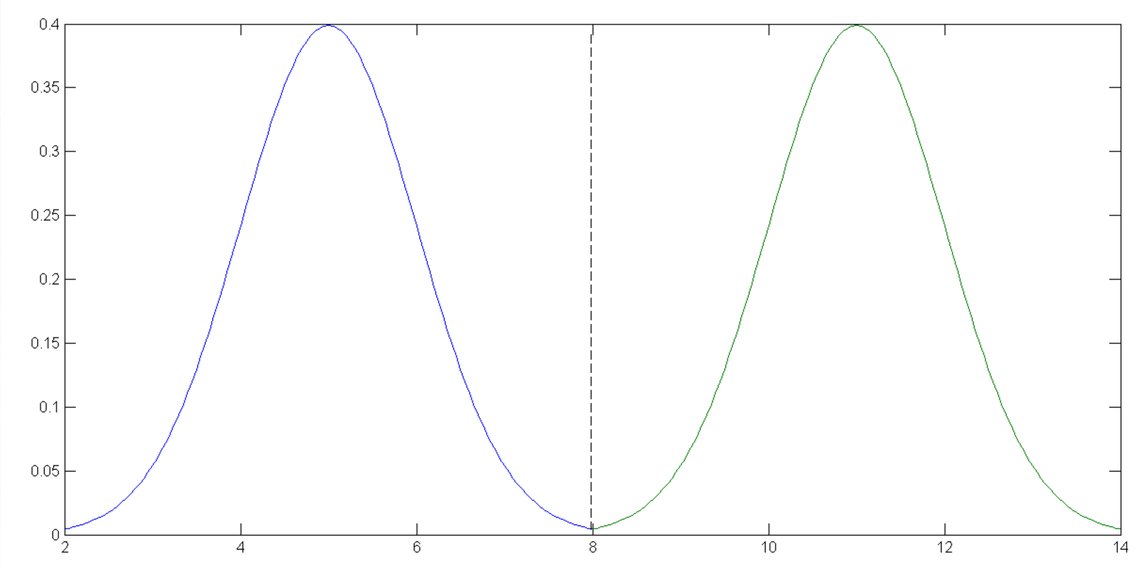

Virtual Drift
Oracle
Slidding Window
Kapp, Sabourin and Maupin (2011)
Kapp, Sabourin and Maupin (2010)
Sun and Li (2011)

DEALING WITH CONCEPT DRIFT

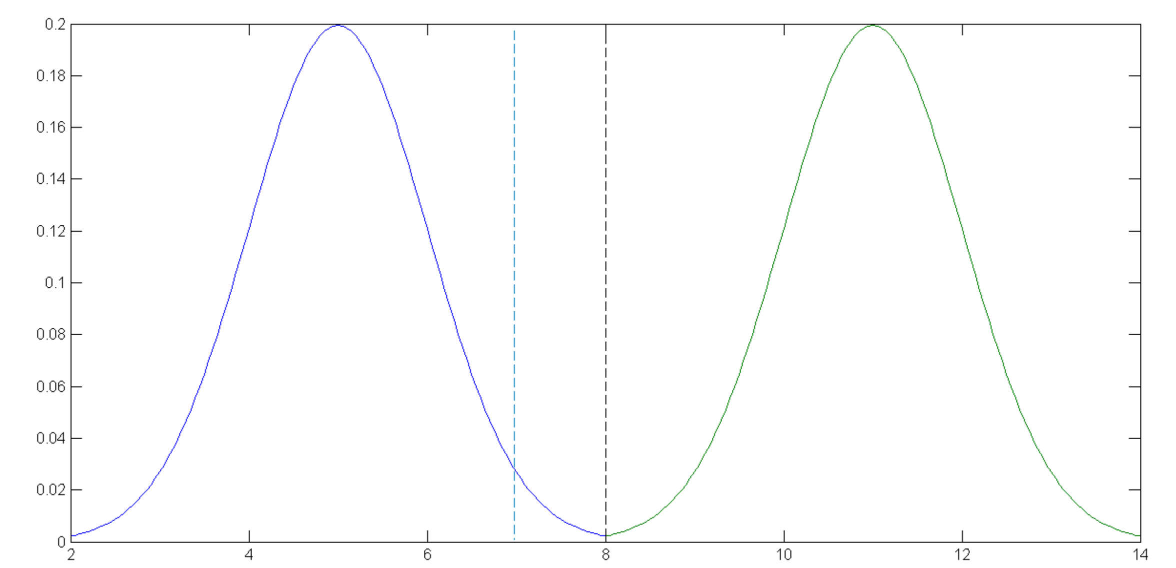
- Analyze the distribution of the new batch to detect drifts
  - Does not requires labeled instances
    - Labeled instances are required to build a new classifier if a drift is detected
- Some drifts may pass undetected using this technique
- Suitable for virtual drifts

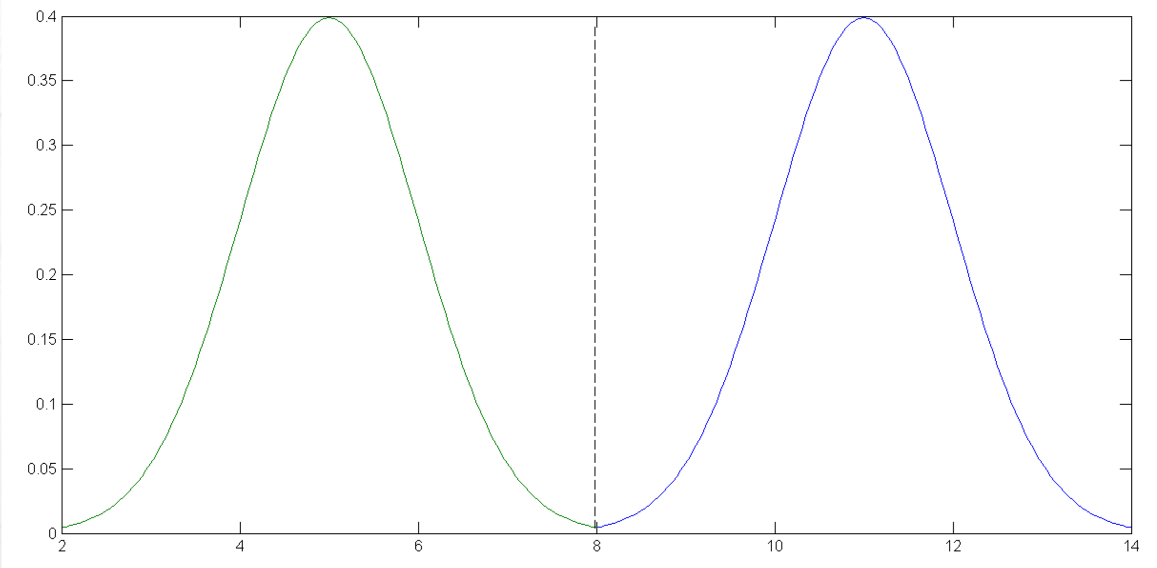
Before the Drift

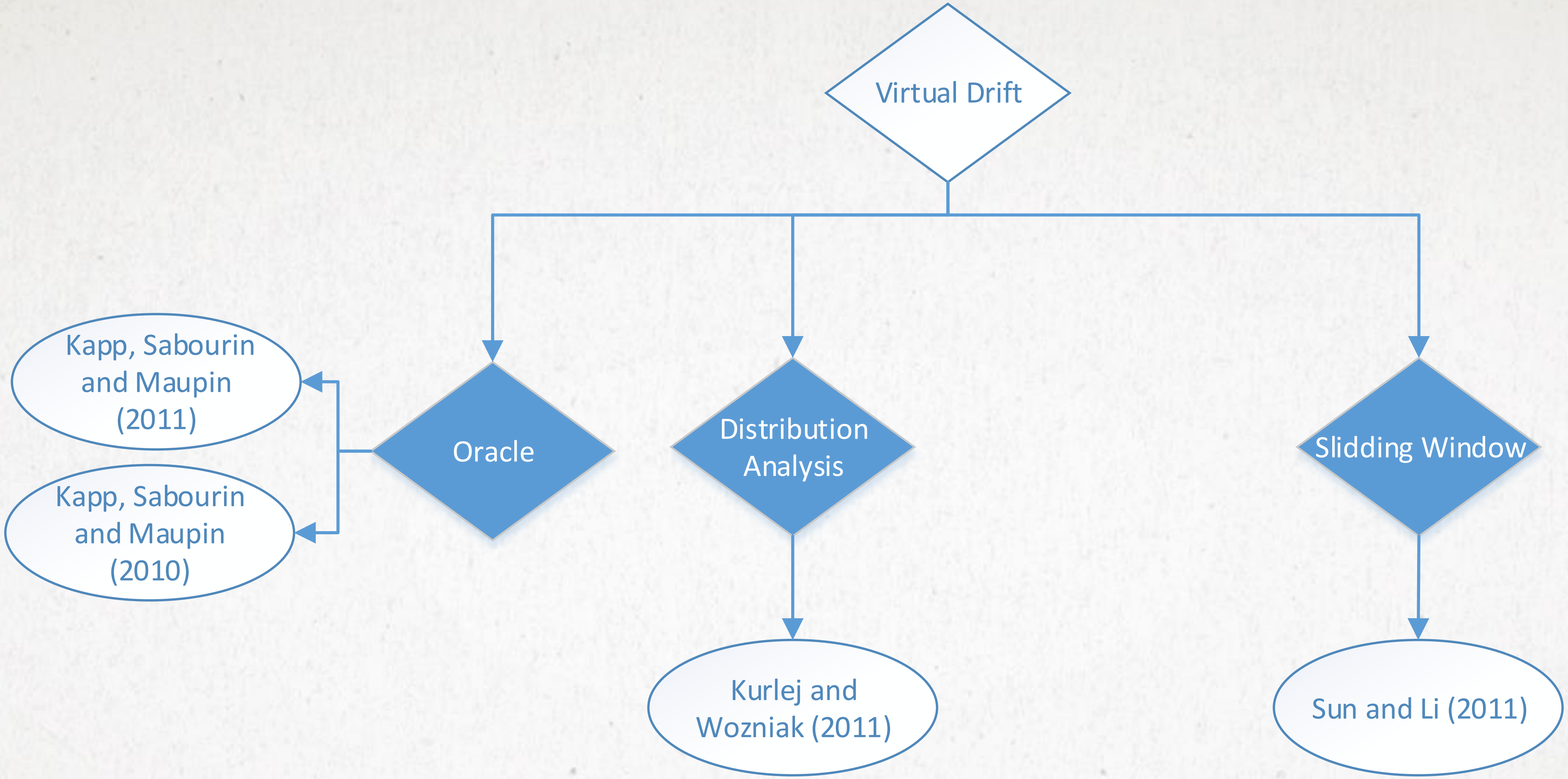




After the Drift

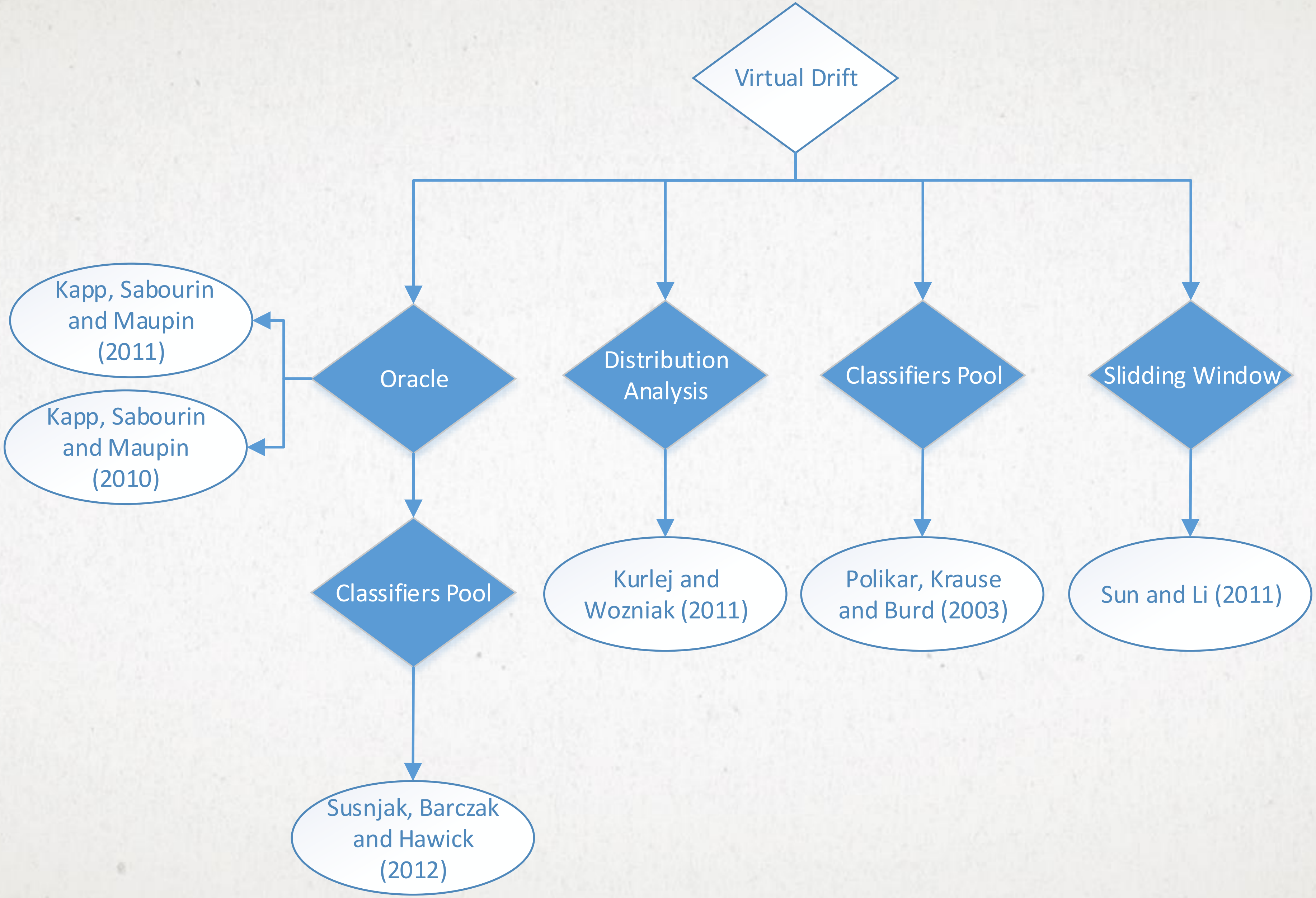


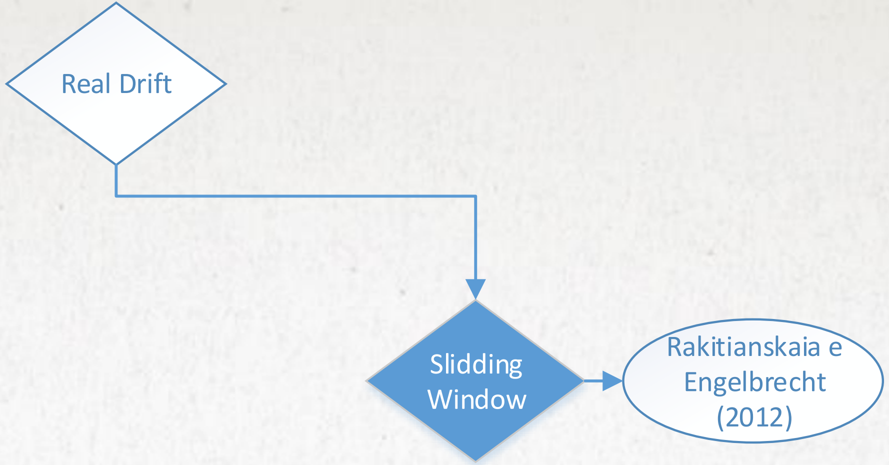


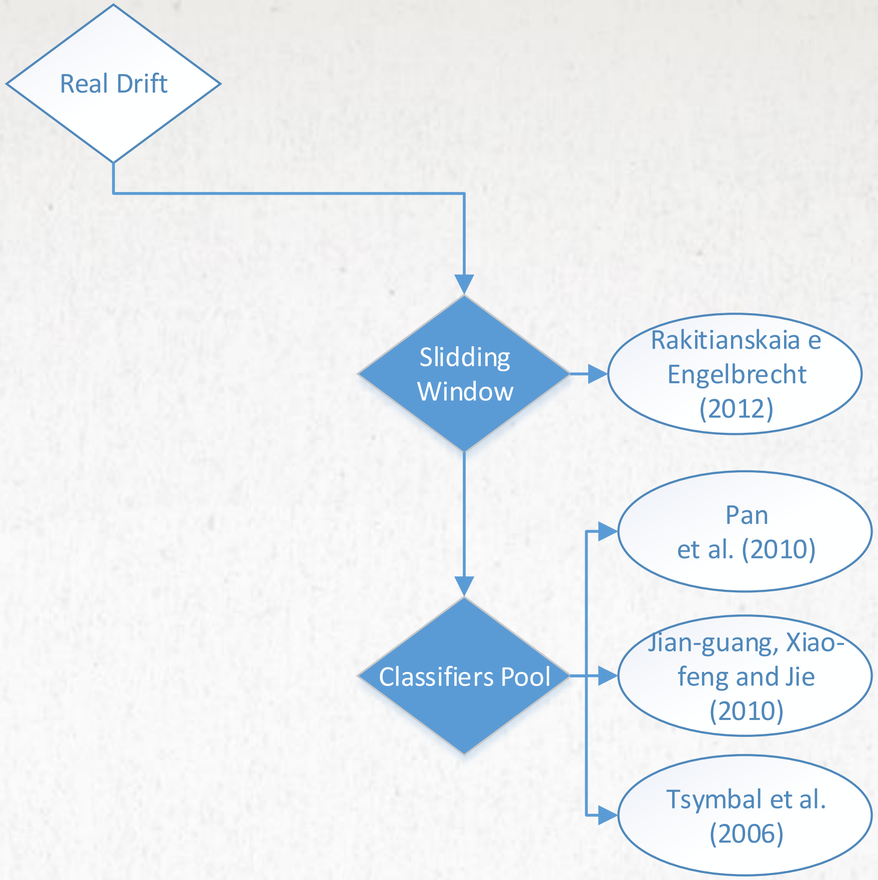

Virtual Drift
Oracle
Distribution Analysis
Sliding Window
Kapp, Sabourin and Maupin (2011)
Kapp, Sabourin and Maupin (2010)
Kurlej and Wozniak (2011)
Sun and Li (2011)

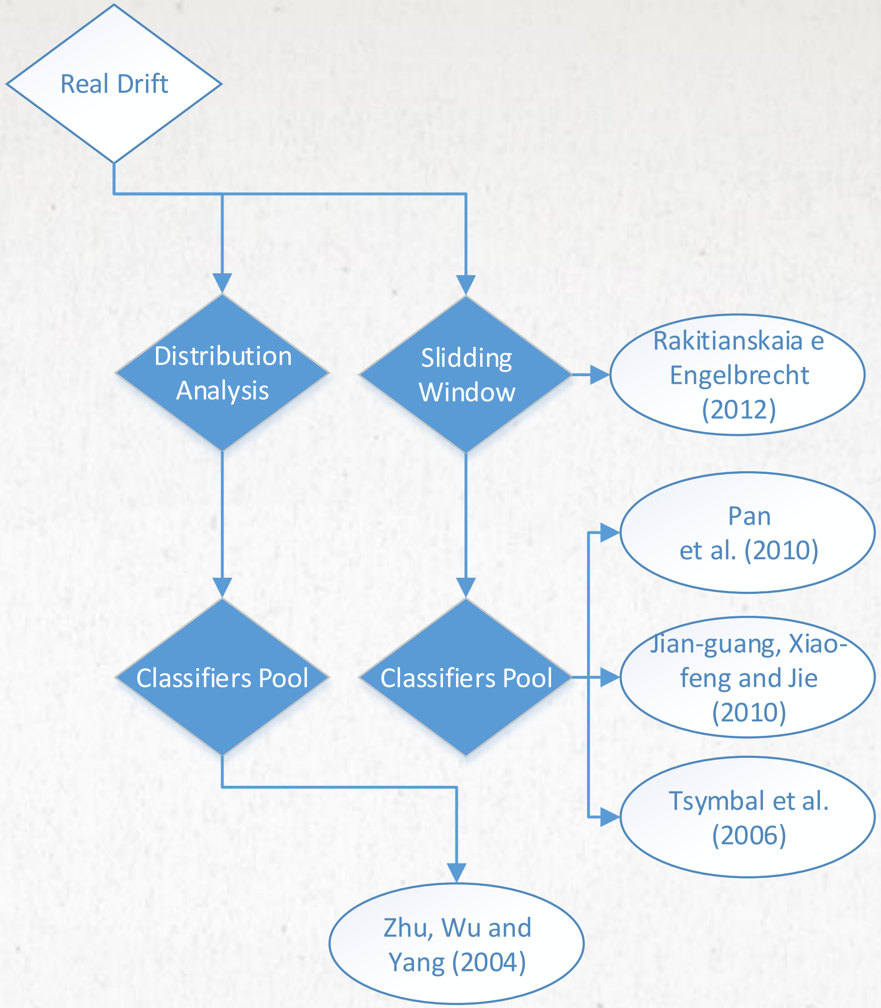
DEALING WITH CONCEPT DRIFT

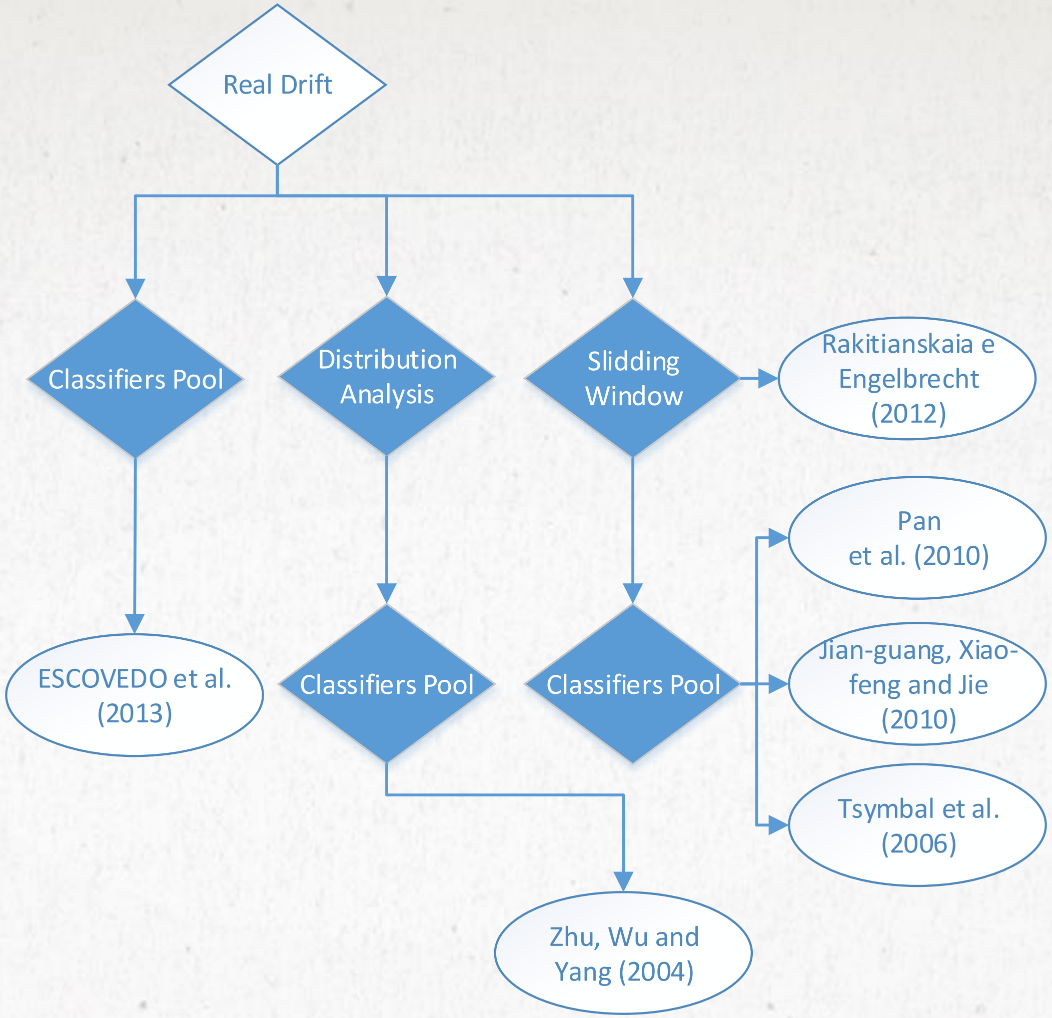
- Pool of Classifiers
  - Add a new classifier in the pool
    - Every time a drift is detected
    - For every new batch
  - Classifier weighting
    - Weights the classifier according to the performance in the latest batch
  - Pool management
    - Discard the worst performing classifiers
    - Discard classifiers with a performance bellow a threshold
  - Combining classifiers
    - Majority voting
    - Weighted majority voting


Virtual Drift
Oracle
Distribution Analysis
Classifiers Pool
Sliding Window
Kapp, Sabourin and Maupin (2011)
Kapp, Sabourin and Maupin (2010)
Classifiers Pool
Kurlej and Wozniak (2011)
Polikar, Krause and Burd (2003)
Sun and Li (2011)
Susnjak, Barczak and Hawick (2012)

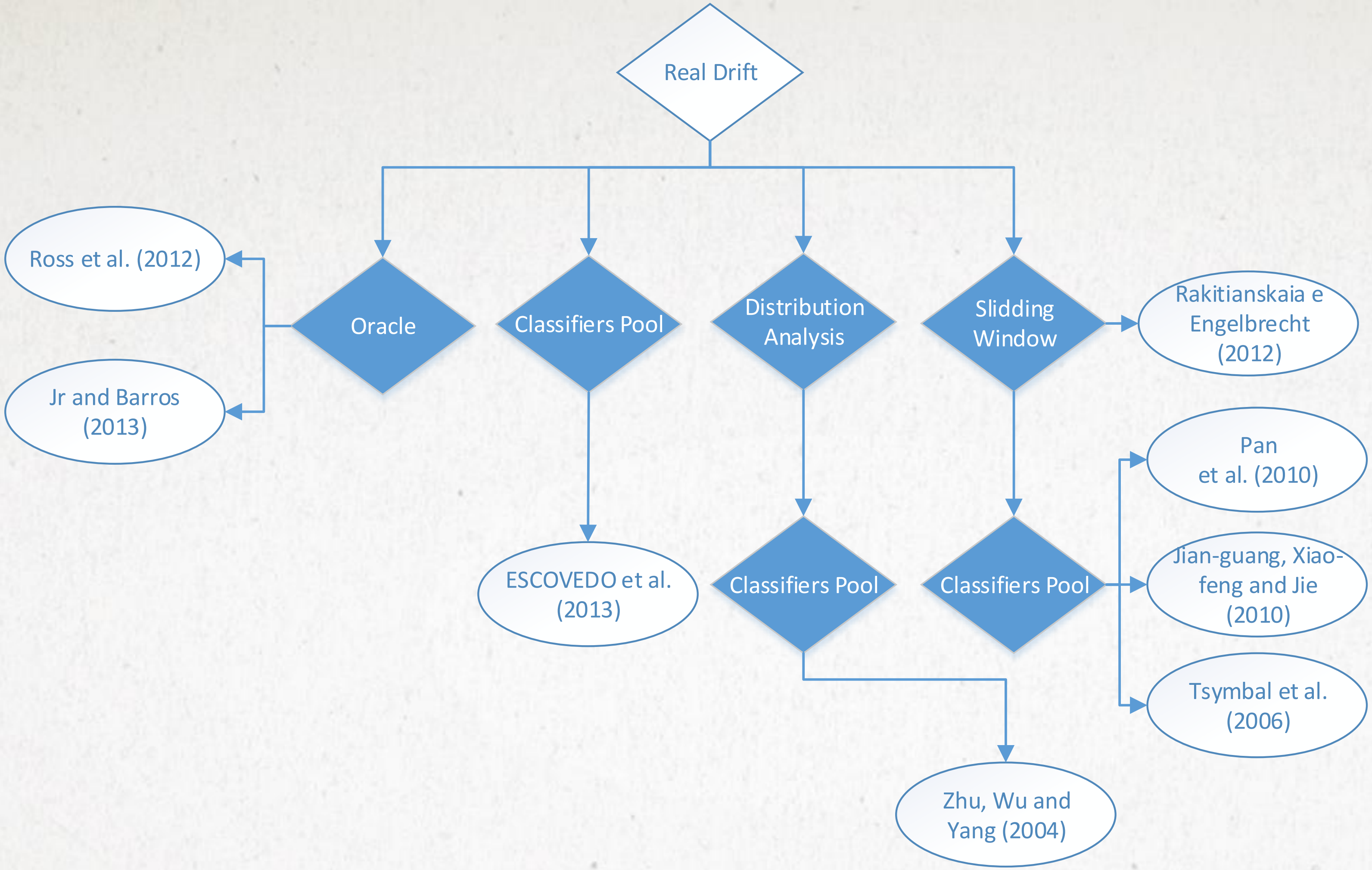

Real Drift
Sliding Window
Rakitianskaia e Engelbrecht (2012)

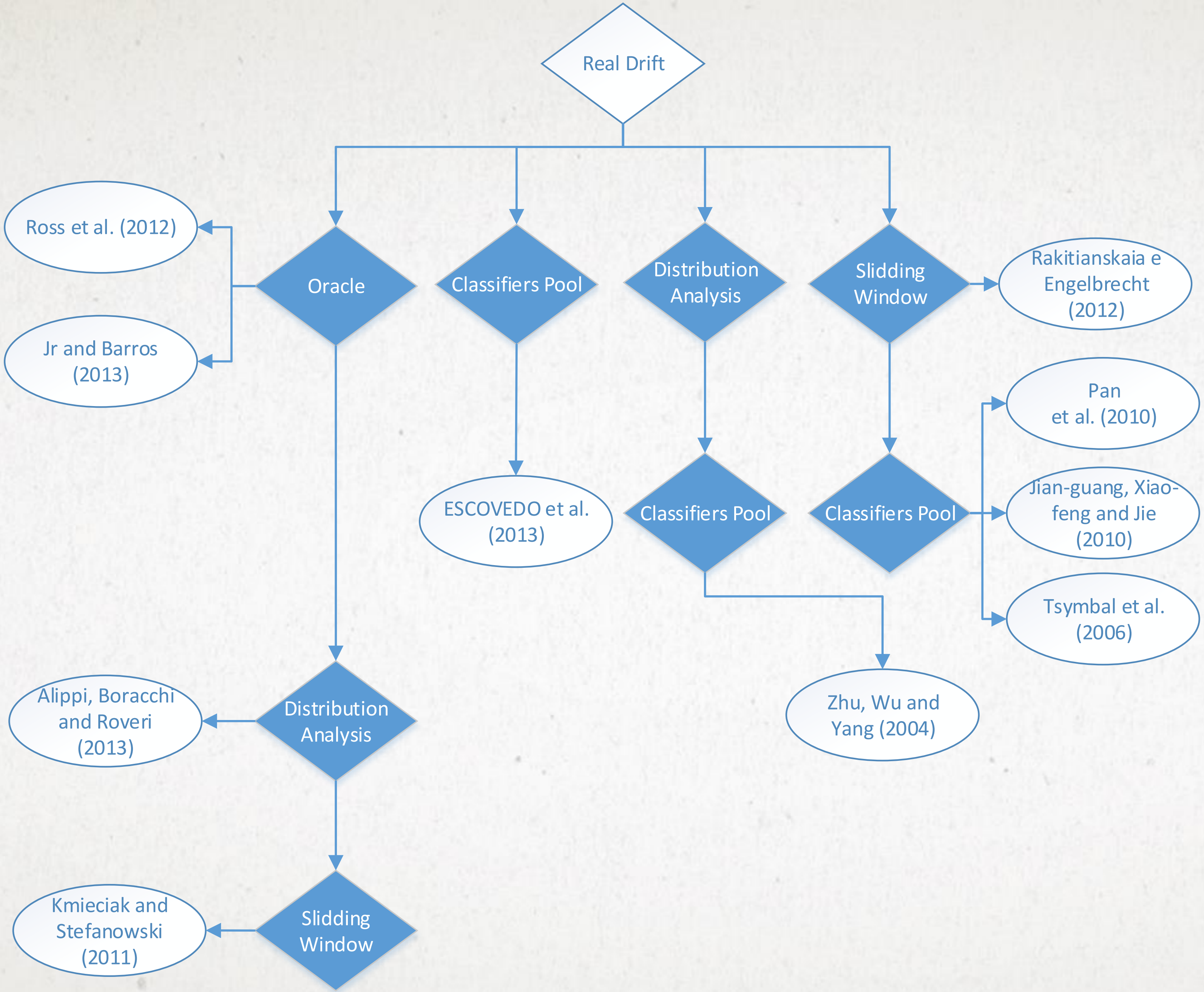

Real Drift
Slidding Window
Rakitianskaia e Engelbrecht (2012)
Classifiers Pool
Pan et al. (2010)
Jian-guang, Xiao-feng and Jie (2010)
Tsymbal et al. (2006)

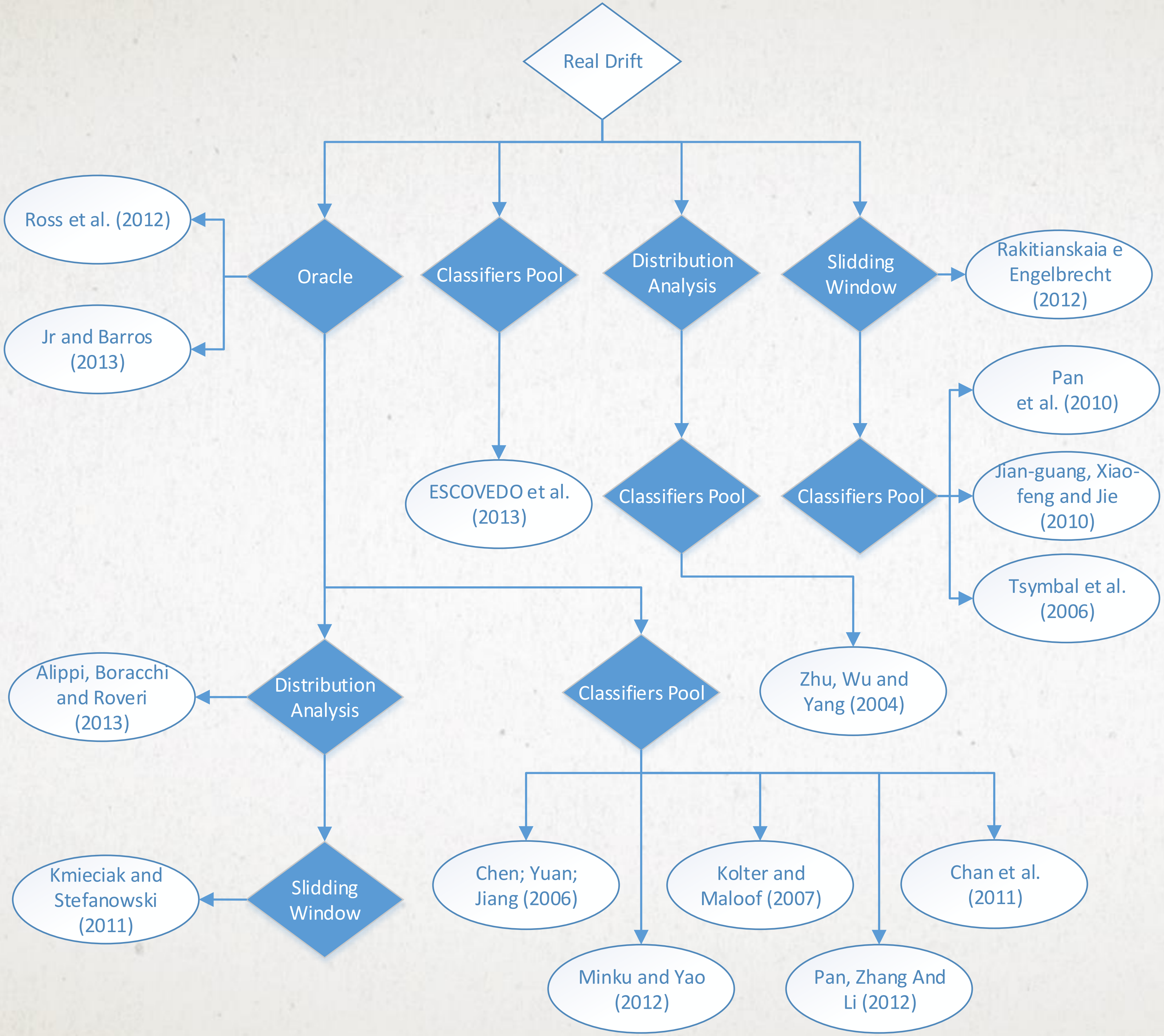

Real Drift
Distribution Analysis
Slidding Window
Rakitianskaia e Engelbrecht (2012)
Pan et al. (2010)
Classifiers Pool
Classifiers Pool
Jian-guang, Xiao-feng and Jie (2010)
Tsymbal et al. (2006)
Zhu, Wu and Yang (2004)



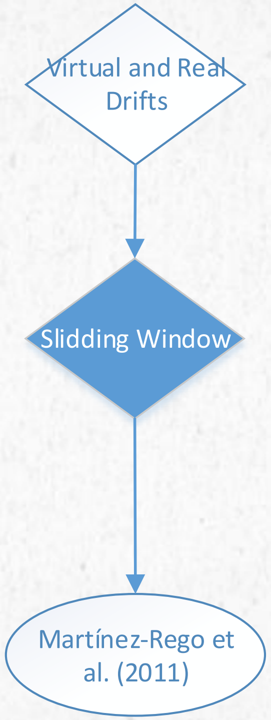
Real Drift
Classifiers Pool
Distribution Analysis
Sliddding Window
Rakitianskaia e Engelbrecht (2012)
ESCOVEDO et al. (2013)
Classifiers Pool
Classifiers Pool
Pan et al. (2010)
Jian-guang, Xiao-feng and Jie (2010)
Tsymbal et al. (2006)
Zhu, Wu and Yang (2004)

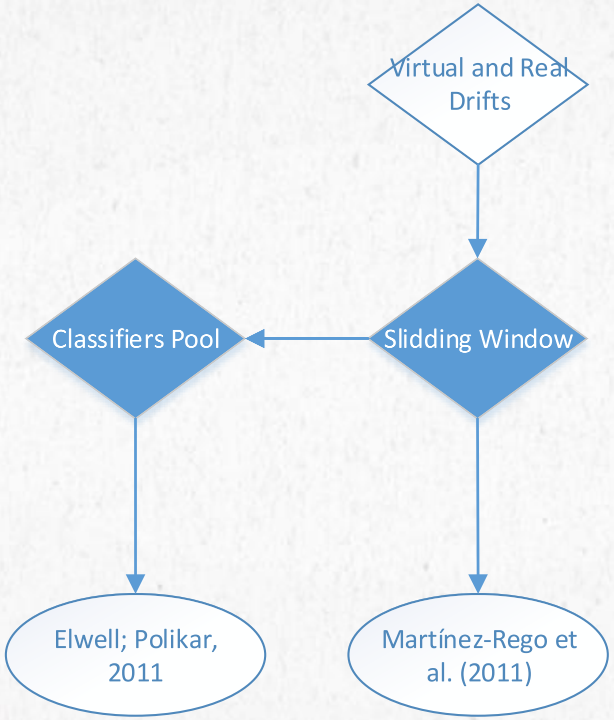

Real Drift
Oracle
Classifiers Pool
Distribution Analysis
Slidding Window
Ross et al. (2012)
Jr and Barros (2013)
ESCOVEDO et al. (2013)
Classifiers Pool
Classifiers Pool
Rakitianskaia e Engelbrecht (2012)
Pan et al. (2010)
Jian-guang, Xiao-feng and Jie (2010)
Tsymbal et al. (2006)
Zhu, Wu and Yang (2004)


Real Drift
Oracle
Classifiers Pool
Distribution Analysis
Sliddina Window
Ross et al. (2012)
Jr and Barros (2013)
Rakitianskaia e Engelbrecht (2012)
ESCOVEDO et al. (2013)
Classifiers Pool
Classifiers Pool
Pan et al. (2010)
Jian-guang, Xiao-feng and Jie (2010)
Tsymbal et al. (2006)
Zhu, Wu and Yang (2004)
Distribution Analysis
Alippi, Boracchi and Roveri (2013)
Sliddina Window
Kmieciak and Stefanowski (2011)



Real Drift
Ross et al. (2012)
Jr and Barros (2013)
Oracle
Classifiers Pool
Distribution Analysis
Sliding Window
Rakitianskaia e Engelbrecht (2012)
Pan et al. (2010)
Jian-guang, Xiao-feng and Jie (2010)
Tsymbal et al. (2006)
ESCOVEDO et al. (2013)
Classifiers Pool
Classifiers Pool
Zhu, Wu and Yang (2004)
Alippi, Boracchi and Roveri (2013)
Distribution Analysis
Classifiers Pool
Kmieciak and Stefanowski (2011)
Sliding Window
Chen; Yuan; Jiang (2006)
Kolter and Maloof (2007)
Chan et al. (2011)
Minku and Yao (2012)
Pan, Zhang And Li (2012)


Virtual and Real Drifts
Sliding Window
Martínez-Rego et al. (2011)


Virtual and Real Drifts
Classifiers Pool
Sliding Window
Elwell; Polikar, 2011
Martínez-Rego et al. (2011)

OUR APPROACH

- Use a time window to detect drifts
  - Compute the dissimilarities
    - When a drift happen, the dissimilarities values as expected to change

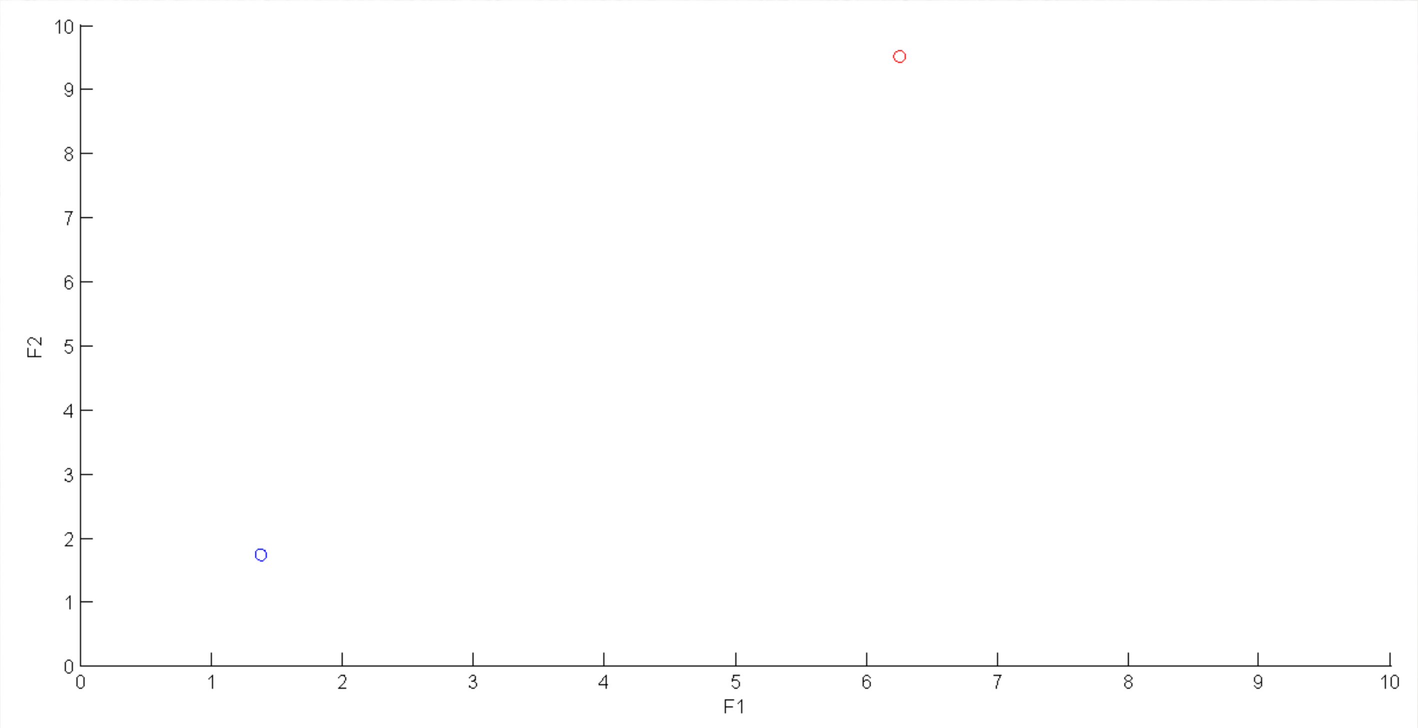
Before the drift

Mean dissimilarity		
	f1	f2
Same class	1.38	1.74
Different classes	6.25	9.5

Window

f1	f2	Class
5	12	2
3	12	2
4	13	2
13	5	1
7	13	2
11	5	1
6	12	2
11	1	1
4	12	2
11	1	1

New Supervised Samples



After the Drift

f1	f2	Class
5	12	2
3	12	2
4	13	2
13	5	1
7	13	2
11	5	1
6	12	2
11	1	1
4	12	2
11	1	1
4	12	1
4	13	1
14	3	2
12	2	2
3	12	1

Window

New Supervised Samples

New supervised batch arrived with a concept drift

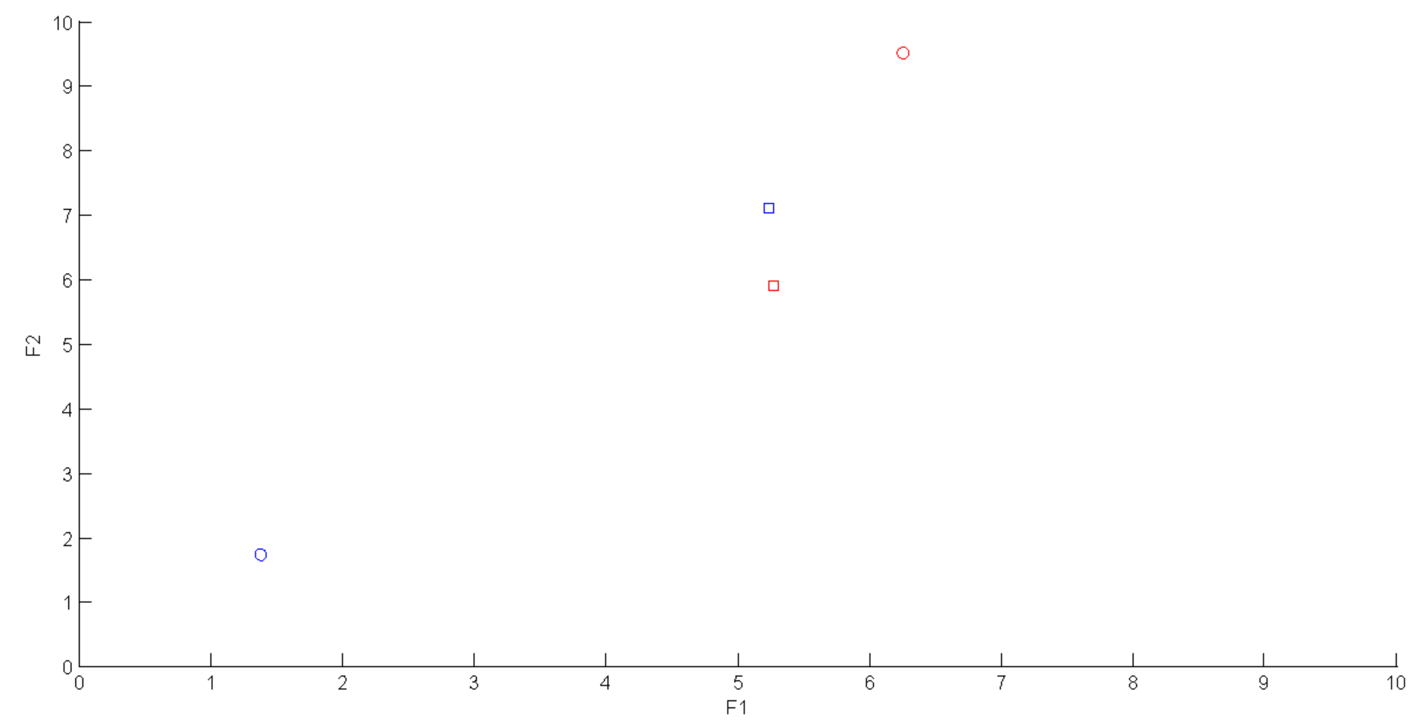
Before the Drift

Mean dissimilarity		
	f1	f2
Same class	1.38	1.74
Different classes	6.25	9.5

After the Drift

Mean dissimilarity		
	f1	f2
Same class	5.23	7.1
Different classes	5.27	5.9

Dissimilarity between same classes increased
Dissimilarity between different classes decreased



QUESTIONS?

BACH, S.; MALOOF, M. Paired learners for concept drift. In: Data Mining, 2008. ICDM ’08. Eighth IEEE International Conference on. [S.l.: s.n.], 2008. p. 23–32. ISSN 1550-4786.

ALIPPI, C.; BORACCHI, G.; ROVERI, M. Just-in-time classifiers for recurrent concepts. Neural Networks and Learning Systems, IEEE Transactions on, v. 24, n. 4, p. 620–634, April 2013. ISSN 2162-237X.

ASUNCION, D. N. A. UCI Machine Learning Repository. 2007. Disponível em: <http://www.ics.uci.edu/$\sim$mlearn/{MLR}epository.ht>.

BACH, S.; MALOOF, M. Paired learners for concept drift. In: Data Mining, 2008. ICDM ’08. Eighth IEEE International Conference on. [S.l.: s.n.], 2008. p. 23–32. ISSN 1550-4786.

BAENA-GARCÍA, M. et al. Early drift detection method. 2006.

BIFET, A.; GAVALDà, R. Learning from time-changing data with adaptive windowing. In: In SIAM International Conference on Data Mining. [S.l.: s.n.], 2007.

BLUM, A. Empirical support for winnow and weighted-majority algorithms: Results on a calendar scheduling domain. Machine Learning, Kluwer Academic Publishers, v. 26, n. 1, p. 5–23, 1997. ISSN 0885-6125.

CHAN, P. et al. Dynamic base classifier pool for classifier selection in multiple classifier systems. In: Machine Learning and Cybernetics (ICMLC), 2011 International Conference on. [S.l.: s.n.], 2011. v. 3, p. 1093–1096. ISSN 2160-133X.

CHEN, H.; YUAN, S.; JIANG, K. Adaptive classifier selection based on two level hypothesis tests for incremental learning. In: YEUNG, D.-Y. et al. (Ed.). Structural, Syntactic, and Statistical Pattern Recognition. Springer Berlin Heidelberg, 2006, (Lecture Notes in Computer Science, v. 4109). p. 687–695. ISBN 978-3-540-37236-3.

DOMINGOS, P.; HULTEN, G. Mining high-speed data streams. In: Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM, 2000. (KDD ’00), p. 71–80. ISBN 1-58113-233-6.

ELWELL, R.; POLIKAR, R. Incremental learning of concept drift in nonstationary environments. Neural Networks, IEEE Transactions on, v. 22, n. 10, p. 1517–1531, Oct 2011. ISSN 1045-9227.

ESCOVEDO, T. et al. Learning under concept drift using a neuroevolutionary ensemble. International Journal of Computational Intelligence and Applications, v. 12, n. 04, p. 1340002, 2013.

FUNG, G. P. C. et al. Text classification without negative examples revisit. Knowledge and Data Engineering, IEEE Transactions on, v. 18, n. 1, p. 6–20, Jan 2006. ISSN 1041-4347.

GAMA, J. et al. Learning with drift detection. In: BAZZAN, A.; LABIDI, S. (Ed.). Advances in Artificial Intelligence – SBIA 2004. Springer Berlin Heidelberg, 2004, (Lecture Notes in Computer Science, v. 3171). p. 286–295. ISBN 978-3-540-23237-7.

HOENS, T.; POLIKAR, R.; CHAWLA, N. Learning from streaming data with concept drift and imbalance: an overview. Progress in Artificial Intelligence, Springer-Verlag, v. 1, n. 1, p. 89–101, 2012. ISSN 2192-6352.

JENHANI, I.; AMOR, N. B.; ELOUEDI, Z. Decision trees as possibilistic classifiers. International Journal of Approximate Reasoning, v. 48, n. 3, p. 784 – 807, 2008. ISSN 0888-613X. Special Section on Choquet Integration in honor of Gustave Choquet (1915–2006) and Special Section on Nonmonotonic and Uncertain Reasoning.

JIAN-GUANG, H.; XIAO-FENG, H.; JIE, S. Dynamic financial distress prediction modeling based on slip time window and multiple classifiers. In: Management Science and Engineering (ICMSE), 2010 International Conference on. [S.l.: s.n.], 2010. p. 148–155. ISSN 2155-1847.

JR, P. M. G.; BARROS, R. S. M. de. Rcd: A recurring concept drift framework. Pattern Recognition Letters, v. 34, n. 9, p. 1018 – 1025, 2013. ISSN 0167-8655.

KAPP, M.; SABOURIN, R.; MAUPIN, P. Adaptive incremental learning with an ensemble of support vector machines. In: Pattern Recognition (ICPR), 2010 20th International Conference on. [S.l.: s.n.], 2010. p. 4048–4051. ISSN 1051-4651.

KAPP, M. N.; SABOURIN, R.; MAUPIN, P. A dynamic optimization approach for adaptive incremental learning. International Journal of Intelligent Systems, Wiley Subscription Services, Inc., A Wiley Company, v. 26, n. 11, p. 1101–1124, 2011. ISSN 1098-111X.

KIFER, D.; BEN-DAVID, S.; GEHRKE, J. Detecting change in data streams. In: Proceedings of the Thirtieth International Conference on Very Large Data Bases - Volume 30. VLDB Endowment, 2004. (VLDB ’04), p. 180–191. ISBN 0-12-088469-0.

KMIECIAK, M. R.; STEFANOWSKI, J. Semi-supervised approach to handle sudden concept drift in enron data. Control and Cybernetics, Vol. 40, no 3, p. 667–695, 2011.

KO, A. H. et al. From dynamic classifier selection to dynamic ensemble selection. Pattern Recognition, v. 41, n. 5, p. 1718 – 1731, 2008. ISSN 0031-3203.

KOLTER, J. Z.; MALOOF, M. A. Dynamic weighted majority: An ensemble method for drifting concepts. Journal of Machine Learning Research, v. 8, p. 2755–2790, Dec 2007.

KUNCHEVA, L. I. Using Control Charts for Detecting Concept Change in Streaming Data. [S.l.], 2009.

KURLEJ, B.; WOZNIAK, M. Active learning approach to concept drift problem. Logic Journal of IGPL, 2011.

MARTíNEZ-REGO, D. et al. A robust incremental learning method for non-stationary environments. Neurocomputing, v. 74, n. 11, p. 1800 – 1808, 2011. ISSN 0925-2312. Adaptive Incremental Learning in Neural Networks Learning Algorithm and Mathematic Modelling Selected papers from the International Conference on Neural Information Processing 2009 (ICONIP 2009) {ICONIP} 2009.

MINKU, L.; YAO, X. Ddd: A new ensemble approach for dealing with concept drift. Knowledge and Data Engineering, IEEE Transactions on, v. 24, n. 4, p. 619–633, April 2012. ISSN 1041-4347.

MINKU, L. L.; WHITE, A. P.; YAO, X. The Impact of Diversity on On-line Ensemble Learning in the Presence of Concept Drift. 2009.

MORENO-TORRES, J. G. et al. A unifying view on dataset shift in classification. Pattern Recognition, v. 45, n. 1, p. 521 – 530, 2012. ISSN 0031-3203.

PAN, S. et al. Classifier ensemble for uncertain data stream classification. In: ZAKI, M. et al. (Ed.). Advances in Knowledge Discovery and Data Mining. Springer Berlin Heidelberg, 2010, (Lecture Notes in Computer Science, v. 6118). p. 488–495. ISBN 978-3-642-13656-6.

PAN, S.; ZHANG, Y.; LI, X. Dynamic classifier ensemble for positive unlabeled text stream classification. Knowledge and Information Systems, Springer-Verlag, v. 33, n. 2, p. 267–287, 2012. ISSN 0219-1377.

POLIKAR, R.; KRAUSE, S.; BURD, L. Ensemble of classifiers based incremental learning with dynamic voting weight update. In: Neural Networks, 2003. Proceedings of the International Joint Conference on. [S.l.: s.n.], 2003. v. 4, p. 2770–2775 vol.4. ISSN 1098-7576.

POLIKAR, R. et al. Learn++: an incremental learning algorithm for supervised neural networks. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, v. 31, n. 4, p. 497–508, Nov 2001. ISSN 1094-6977.

RAKITIANSKAIA, A.; ENGELBRECHT, A. Training feedforward neural networks with dynamic particle swarm optimisation. Swarm Intelligence, Springer US, v. 6, n. 3, p. 233–270, 2012. ISSN 1935-3812.

ROSS, G. J. et al. Exponentially weighted moving average charts for detecting concept drift. Pattern Recognition Letters, v. 33, n. 2, p. 191 – 198, 2012. ISSN 0167-8655.

SCHLIMMER, J.; GRANGER RICHARDH., J. Incremental learning from noisy data. Machine Learning, Kluwer Academic Publishers, v. 1, n. 3, p. 317–354, 1986. ISSN 0885-6125.

SUN, J.; LI, H. Dynamic financial distress prediction using instance selection for the disposal of concept drift. Expert Systems with Applications, v. 38, n. 3, p. 2566 – 2576, 2011. ISSN 0957-4174.

SUSNJAK, T.; BARCZAK, A.; HAWICK, K. Adaptive cascade of boosted ensembles for face detection in concept drift. Neural Computing and Applications, Springer-Verlag, v. 21, n. 4, p. 671–682, 2012. ISSN 0941-0643.

TSYMBAL, A. et al. Handling local concept drift with dynamic integration of classifiers: Domain of antibiotic resistance in nosocomial infections. In: Computer-Based Medical Systems, 2006. CBMS 2006. 19th IEEE International Symposium on. [S.l.: s.n.], 2006. p. 679–684. ISSN 1063-7125.

VIOLA, P.; JONES, M. Rapid object detection using a boosted cascade of simple features. In: Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. [S.l.: s.n.], 2001. v. 1, p. I–511–I–518 vol.1. ISSN 1063-6919.

ZHU, X.; WU, X.; YANG, Y. Dynamic classifier selection for effective mining from noisy data streams. In: Data Mining, 2004. ICDM ’04. Fourth IEEE International Conference on. [S.l.: s.n.], 2004. p. 305–312.

THANKS