The authors study online supervised learning under the empirical zero-one loss and introduce a novel classification algorithm with strong theoretical guarantees. The proposed method is a highly dynamical self-organizing decision tree structure, which adaptively partitions the feature space into small regions and combines (takes the union of) the local simple classification models specialized in those regions. The authors' approach sequentially and directly minimizes the cumulative loss by jointly learning the optimal feature space partitioning and the corresponding individual partition-region classifiers. They mitigate overtraining issues by using basic linear classifiers at each region while providing a superior modeling power through hierarchical and data adaptive models. The computational complexity of the introduced algorithm scales linearly with the dimensionality of the feature space and the depth of the tree. Their algorithm can be applied to any streaming data without requiring a training phase or a priori information, hence processing data on-the-fly and then discarding it. Therefore, the introduced algorithm is especially suitable for the applications requiring sequential data processing at large scales/high rates. The authors present a comprehensive experimental study in stationary and nonstationary environments. In these experiments, their algorithm is compared with the state-of-the-art methods over the well-known benchmark datasets and shown to be computationally highly superior. The proposed algorithm significantly outperforms the competing methods in the stationary settings and demonstrates remarkable adaptation capabilities to nonstationarity in the presence of drifting concepts and abrupt/sudden concept changes.