

Hidden Markov and semi-Markov models: When and why are these models useful for classifying state in time series data?

Hidden Markov models (HMMs) and their extensions have proven to be powerful tools for classification of observations that stem from systems with temporal dependence as they take into account that observations close in time are likely generated from the same state (i.e. class). When information on the classes of the observations is available in advance, supervised methods can be applied. In this paper, we provide

details for the implementation of four models for classification in a supervised learning context: HMMs, hidden semi-Markov models (HSMMs), autoregressive-HMMs, and autoregressive-HSMMs. Using simulations, we study the classification performance under various degrees of model misspecification to characterize when it would be important to extend a basic HMM to an HSMM. As an application of these techniques we use the models to classify accelerometer data from Merino sheep to distinguish between four different behaviors of interest. In particular in the field of movement ecology, collection of fine-scale animal movement data over time to identify behavioral states has become ubiquitous, necessitating models that can account for the dependence

structure in the data. We demonstrate that when the aim is to conduct classification, various degrees of model misspecification of the proposed model may not impede good classification performance unless there is high overlap between the state-dependent distributions, that is, unless the observation distributions of the different states are difficult to differentiate.