

Work accomplished so far:

I have explored some of the Dirichlet process (DP), hierarchical Dirichlet process (HDP) and hierarchical Dirichlet process for hidden Markov model (HDP-HMM) literature [1, 2, 3, 4]. To assist my understanding of the DP, I wrote a generative model of a mixture of 4 Gaussians and I wrote a basic Stan model for inference over the parameters (with a stick breaking prior but a truncated number of components). The implementation marginalizes over the cluster assignments (z_i) for compatibility with NUTS and uses the posterior distributions of the parameters to make MLE cluster assignments. The Stan implementation is exceedingly slow but the implementation was useful for my understanding.

I have also implemented a blocked Gibbs sampler for the sticky HDP-HMM, detailed in Algorithm 2 of [1] and discussed further in [3]. I ran a test on a small generated time series where sequential data are generated from three Gaussian modes. The sticky HDP-HMM recovers the mode assignments and regime means as is presented in [1]. I used a truncation level of $L = 15$ modes. I ran the Gibbs sampler for 100 iterations and I used the final sample as a point estimate for the posterior. Due to the problems of label switching, it is not yet clear to me how to get a mean estimate of the cluster assignments from many samples (the sampler appears to switch among some subset of ‘active’ labels). It was interesting to investigate the effect of the ‘sticky’ parameter that encourages self transitions in the HMM. Figure 1 shows that the HDP-HMM without the sticky parameter ($\kappa = 0$) finds a noisier mode sequence with more mode transitions and more possible modes than the HDP-HMM with a high sticky parameter ($\kappa = 100$).

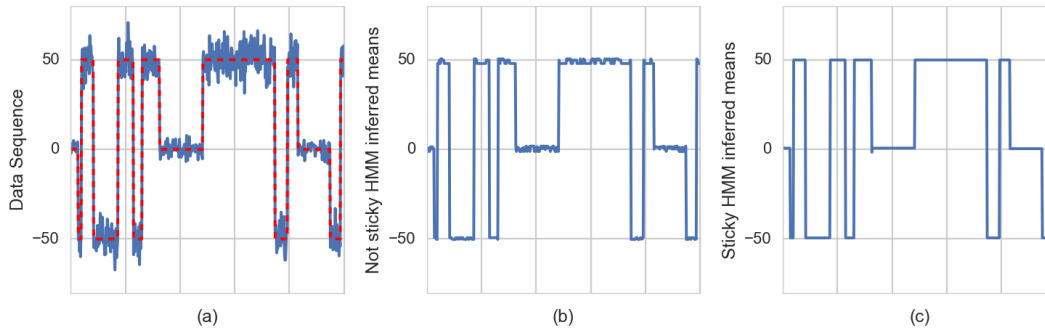


Figure 1: (a) generated data sequence - there are 3 true modes. Right two plots show the inferred means and mode assignments with (b) $\kappa = 0$ (model infers 15 modes) and (c) $\kappa = 100$ (model infers 5 modes).

Plan for rest of project:

The rest of the project aims to implement the Gibbs samplers for the HDP-SLDS and HDP-AR-HMM models from [2]. I will begin by implementing and reproducing the tests on generated data for the HDP-AR-HMM model. If time permits, I will expand this implementation to that for the HDP-SLDS.

1. Implement the forward and backward recursions for the state space model in the AR/SLDS models (I have already written HMM filters for the implementation above).
2. Write the Gibbs sampler that is discussed in [2] for the HDP-AR-HMM model.
3. Expand upon (2) to develop the Gibbs sampler for the HDP-SLDS model (if time permits).
4. Reproduce the tests from [2] (specifically I'd like to reproduce Figure 2) on a similar generated data set.

References

- [1] E. Fox, E. Sudderth, M. Jordan, and A. Willsky. Developing a tempered hdp-hmm for systems with state persistence. *MIT Laboratory for Information and Decision Systems, Tech. Rep.*, 2007.
- [2] E. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky. Bayesian nonparametric inference of switching dynamic linear models. *IEEE Transactions on Signal Processing*, 59(4):1569–1585, 2011.
- [3] E. B. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky. An hdp-hmm for systems with state persistence. In *Proceedings of the 25th international conference on Machine learning*, pages 312–319. ACM, 2008.
- [4] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei. Sharing clusters among related groups: Hierarchical dirichlet processes. In *Advances in neural information processing systems*, pages 1385–1392, 2005.