Infinite von Mises-Fisher Mixture Modeling of Whole Brain fMRI Data

Cluster analysis of functional magnetic resonance imaging (fMRI) data is often performed using gaussian mixture models, but when the time series are standardized such that the data reside on a hypersphere, this modeling assumption is questionable. The consequences of ignoring the underlying spherical manifold are rarely analyzed, in part due to the computational challenges imposed by directional statistics. In this letter, we discuss a Bayesian von Mises-Fisher (vMF) mixture model for data on the unit hypersphere and present an efficient inference procedure based on collapsed Markov chain Monte Carlo sampling. Comparing the vMF and gaussian mixture models on synthetic data, we demonstrate that the vMF model has a slight advantage inferring the true underlying clustering when compared to gaussian-based models on data generated from both a mixture of vMFs and a mixture of gaussians subsequently normalized. Thus, when performing model selection, the two models are not in agreement. Analyzing multisubject whole brain resting-state fMRI data from healthy adult subjects, we find that the vMF mixture model is considerably more reliable than the gaussian mixture model when comparing solutions across models trained on different groups of subjects, and again we find that the two models disagree on the optimal number of components. The analysis indicates that the fMRI data support more than a thousand clusters, and we confirm this is not a result of overfitting by demonstrating better prediction on data from held-out subjects. Our results highlight the utility of using directional statistics to model standardized fMRI data and demonstrate that whole brain segmentation of fMRI data requires a very large number of functional units in order to adequately account for the discernible statistical patterns in the data.