

Additional File 3

Model Settings used in Experiments with PHS and AML Data Sets

June 23, 2014

1 Model Settings used in Experiments with PHS and AML Data Sets

Parameter settings used for ASPIRE, DPGMM, FLAME, FlowPeaks, and HDPGMM in experiments with PHS and AML data sets are shown in Table 1 below.

Table 1: Model Settings used in Experiments with PHS and AML Data Sets.

ASPIRE	$m=d+2$, $\kappa_0=0.05$, $s=150\log(d+1)/d$ $\alpha = 1$, $\gamma = 1$ $\kappa_i \in \{0.1, 0.25, 0.5, 1\}$ (actual value selected to maximize Gibbs likelihood) μ_0 =mean of all data points $\Sigma_0=I/s$ burn-in sweeps = 750 post burn-in sweeps = 250
DPGMM	$m=d+2$, $\kappa_0=0.05$, $s=150\log(d+1)/d$ $\alpha = 1$ μ_0 =mean of all data points $\Sigma_0=I/s$ burn-in sweeps = 750 post burn-in sweeps = 250
FLAME	minimal number of clusters to be fit = 3 maximal number of clusters to be fit = 8 density = skew t estimate mode = no method for choosing optimal number of clusters = scale-free weighted ratio
flowPeaks	tolerance for merging clusters = 0.1 multiplier of the variance matrix $S_0 = 1$ multiplier of the variance matrix $S = 1.5$
HDPGMM	truncation level = 128 burn-in sweeps = 10,000 post burn-in sweeps = 100