Supplement to Tree-Structured Infinite Sparse Factor Model

XianXing Zhang
Electrical & Computer Engineering Department, Duke University, Durham, NC, USA
xianxing.zhang@duke.edu

David B. Dunson
Department of Statistical Sciences, Duke University, Durham, NC, USA
dunson@stat.duke.edu

Lawrence Carin
Electrical & Computer Engineering Department, Duke University, Durham, NC, USA
lcarin@duke.edu

1 Proof of Theorem 1

To prove $\Omega_b$ converges a.s. for any $b \in T$ we need to show that matrix $\Delta_b = \sum_{n \in b} d_n d_n^T$ converges a.s. for each element $\sum_{n \in b} d_{r,s} n^T n$.

Lemma 1. (Lévy’s theorem) Suppose $\{X_n\}_{n \geq 1}$ is an independent sequence of random variables, $\sum_{n=1}^{\infty} X_n$ converges a.s. iff $\sum_{n=1}^{\infty} X_n$ converges i.p.

Lemma 2. (Cauchy criterion) $\{X_n\}_{n \geq 1}$ converges i.p. iff $X_n + k - X_k \to 0$ i.p. as $n,k \to \infty$

Since each branch $b$ is modeled as a factor model, for all loadings on $b$ we denote $m_{k,p} = \sum_{l=k}^{\infty} d_{r,s} n^T n$. By Cauchy-Schwartz inequality:

$$\left( \sum_{l=k}^{\infty} d_{r,s} n^T n \right)^2 \leq \sum_{l=k}^{\infty} d_{r,l}^2 \sum_{l=k}^{\infty} d_{l,s}^2 \leq \max_{1 \leq p \leq P} m_{k,p}^2.$$ Combining this result with Lemma 1,2, to prove Theorem 1 it’s sufficient to show that

$$\lim_{k \to \infty} p(\max_{1 \leq p \leq P} m_{k,p} < \epsilon) = 1.$$ where the equality in the first line follows from the fact that $m_{k,p}$ are conditionally i.i.d given $\gamma_p$ and the subsequent two inequalities use Jensen’s and Chebyshev’s inequality respectively. Now based on equation (3) in the paper: $E(E(m_{k,1} | \gamma_1)) = \sum_{l=k}^{\infty} 3ba^{l-1} = \frac{3a}{1-a}a^{k-1}$, where $b = E(\zeta_1^{-1})$, $a = E(\zeta_2^{-1}) < 1$ if $c_2 > 2$. At last we arrive at the sufficiency equation and thus Theorem 1 is proved:

$$\lim_{k \to \infty} p(\max_{1 \leq p \leq P} m_{k,p} < \epsilon) \geq \lim_{k \to \infty} \left( 1 - \frac{3b}{(1-a)\epsilon}a^{k-1} \right)^P = 1.$$
Figure 1: The full tree structure inferred from faces data where each node is plotted as the average of all images that were assigned to that node. Leaves at branches with different depth are placed on same horizontal level for purpose of interpretation.
Figure 2: Tree-structured hierarchy (with top six layers) embedded with dictionaries learned from 100,000 patches of size 16×16 pixels.