Supplementary material for: Multiresolution dictionary learning for conditional distributions

Anonymous Author(s)
Affiliation
Address
email

1 Full conditionals

Introduce the latent variable $S_i \in \{1, \ldots, k\}$, for $i = 1, \ldots, n$, denoting the multiscale level used by the ith subject. Assuming data are normalized prior to analysis, we let $\mu \sim \mathcal{N}(0, I)$ and $\sigma = \mathcal{IG}(a, b)$ for the means and variances of the dictionary densities. Let n_{B_j} be the number of observations allocated to node B_j . Each Gibbs sampler iteration can be summarized in the following steps.

1. Update S_i by sampling from the multinomial full conditional with
$$\Pr(S_i = j \mid -) = \frac{\pi_{B_j(x_i)} f_{B_j(x_i)}(y_i)}{\sum_{h=1}^{k} \pi_{B_h(x_i)} f_{B_h(x_i)}(y_i)}$$
2. Update stick-breaking random variable $V_{B_j(x_i)}$, for $j = 1, \ldots, k$ and $i = 1, \ldots, n$, from Beta(β_p, α_p) with $\beta_p = 1 + n_{B_j}$ and $\alpha_p = \alpha + \sum_{B_h(x_i) \in de\{B_j(x_i)\}} n_{B_h(x_i)}$.
3. Update $(\mu_{B_j(x_i)}, \sigma_{B_j(x_i)})$ by sampling from
$$\mu_{B_j} \sim \mathcal{N}\left(\bar{y}_{B_j} n_{B_j} / \sigma_{B_j}, (1 + n_{B_j}/\sigma_{B_j})^{-1}\right)$$
$$\sigma_{B_j} \sim \mathcal{IG}\left(a_\sigma, b + 0.5 \sum_{\{i: S_i = j, x_i \in B_j\}} \left(y_i - \mu_{B_j}\right)^2\right)$$
with $a_\sigma = a + n_{B_j}/2$, $\bar{y}_{B_j}$ being the average of the observation $\{y_i\}$ allocated to node B_j.

2 Predictions

Consider the case we want to predict the response y_{n+1} for a future subject based on the predictors x_{n+1} and $(y_1, \ldots, y_n)$. For each tree level, the new vector of predictors x_n is allocated to subsets having closer centers with respect some metric. We will consider the euclidean metric. Then, for a new observation the predictive density is defined as

$$p(y_{n+1}|x_{n+1}, y_1, \ldots, y_n) = \int f\left(y_{n+1}|x_{n+1}, \Omega\right) dp\left(\Omega|y_1, \ldots, y_n\right) \tag{1}$$

with $f\left(y_{n+1}|x_{n+1}, \Omega\right)$ defined as in (1) and Ω being the set of all parameters involved, i.e. weights, location and scale parameters. In order to make inference on the predictive density of y_{n+1}, at the sth Gibbs sampler iteration, we will first sample parameters involved in **??** from its posterior, i.e. $\Omega^{(s)} \sim p\left(\Omega|y_1, \ldots, y_n\right)$ and then we will sample $y_{n+1}^{(s)}$ from $p\left(y_{n+1}|x_{n+1}, \Omega^{(s)}\right)$. Let us assume the number of iterations is S an a burn-in of b is considered. Then, given the sequence $\left(y_{n+1}^{(b+1)}, \ldots, y_{n+1}^{(S)}\right)$, summaries of the predictive density such as mean, variance and quantiles can be computed.

Table 1: Linear manifold example 1: Mean and standard deviations of squared errors under multi-scale stick-breaking (MSB), CART and Lasso for sample size 50 and 100 for different simulation scenarios.

			$r=5$			$r=10$		
p	n		MSB	CART	LASSO	MSB	CART	LASSO
$1e+04$	50	MSE	0.18	0.31	0.25	0.22	0.58	0.22
		STD	0.32	0.30	0.42	0.24	0.54	0.30
		TIME	3	2	1	3	3	1
$1e+04$	100	MSE	0.18	0.27	0.26	0.20	0.41	0.52
		STD	0.26	0.42	0.46	0.23	0.46	0.78
		TIME	5	5	2	5	5	1
$1e+05$	50	MSE	0.35	0.45	0.89	0.16	0.33	0.20
		STD	0.53	0.77	1.04	0.21	0.46	0.31
		TIME	3	25	2	3	27	2
$1e+05$	100	MSE	0.43	0.88	0.52	0.17	0.50	0.31
		STD	0.59	1.29	0.70	0.24	0.75	0.49
		TIME	7	50	5	7	51	5
$5e+05$	50	MSE	0.11	0.16	0.15	0.83	2.26	0.92
		STD	0.15	0.24	0.19	1.01	2.60	3.69
		TIME	5	90	11	5	121	10
$5e+05$	100	MSE	0.003	0.17	0.08	0.13	1.37	1.06
		STD	0.16	0.23	0.13	1.12	1.81	1.50
		TIME	10	214	43	8	227	42
$7e+05$	50	MSE	1.70	1.48	1.47	0.66	1.65	1.07
		STD	2.18	2.47	1.63	0.87	1.49	0.95
		TIME	6	121	12	7	151	13
$5e+05$	100	MSE	0.69	1.36	0.82	0.78	1.52	1.43
		STD	0.94	1.47	1.28	1.03	1.34	2.11
		TIME	13	321	41	12	325	44

Table 2: Linear manifold example 2: Mean and standard deviations of squared errors under multi-scale stick-breaking (MSB), CART and Lasso for different sample sizes

			$r=2$			$r=5$		
p	n		MSB	CART	LASSO	MSB	CART	LASSO
$10e+03$	100	MSE	1.54	1.78	2.37	0.84	1.25	1.62
		STD	1.70	1.72	0.89	1.38	1.35	1.47
$50e+03$	100	MSE	0.76	0.97	1.77	0.88	1.53	1.43
		STD	1.04	1.21	3.13	1.00	1.59	2.73
$10e+04$	100	MSE	0.77	1.01	1.61	0.67	0.46	0.97
		STD	0.94	1.13	1.85	0.82	0.61	1.16
$20e+04$	100	MSE	0.86	0.90	1.41	0.74	1.09	0.78
		STD	1.30	1.35	1.41	0.95	1.98	0.95

3 Synthetic examples

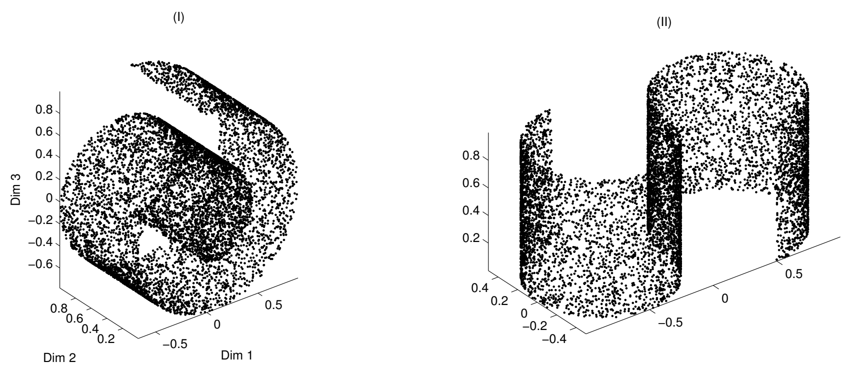

Figure 1: Non-linear manifolds: Swissroll (I) and S-Manifold (II) embedded in $\mathcal{R}^3$