Supplementary Material
"Bayesian Subject-Specific Bi-Level Feature Selection Model"

Daniel R. Baer

1 Subject-Specific Group-Level Feature Selection Model

As motivated by Xu and Ghosh [1] and Liquet et al. [2], we present here a Bayesian subject-specific group-level feature selection model.

1.1 Model hierarchy

For $k = 1, \ldots, K$ let

- $vec\left(\underset{J\times n}{\mathbf{Y}^T}\right) | \underset{nJ\times Jp}{\mathbf{X}}, \underset{pJ\times 1}{\beta}, \sigma^2, \underset{nJ\times nq}{\mathbf{Z}}, \underset{nq\times 1}{\mathbf{b}} \sim$ $N_{Jn}\left(\underset{nJ\times Jp}{\mathbf{X}}\underset{pJ\times 1}{\beta} + \underset{nJ\times nq}{\mathbf{Z}}\underset{nq\times 1}{\mathbf{b}}, \left(\sigma^2 \underset{n\times n}{\mathbf{I}_n} \otimes \underset{J\times J}{\mathbf{I}_J}\right)\right)$,
- $vec\left(\underset{p_k\times J}{\mathbf{B}_k}\right) | \underset{J\times J}{\mathbf{\Sigma}}, \tau_k^2, \pi_{0k} \overset{ind.}{\sim} \pi_{0k} N_{p_k J}\left(\underset{p_k J\times 1}{\mathbf{0}}, \tau_k^2 \underset{J\times J}{\mathbf{\Sigma}} \otimes \underset{p_k\times p_k}{\mathbf{I}_{p_k}}\right) + (1-\pi_{0k})\,\delta_0\left(vec\left(\underset{p_k\times J}{\mathbf{B}_k}\right)\right)$,
- $\tau_k^2|\lambda_k^2 \overset{ind.}{\sim} Gamma\left(shape = \frac{p_k J+1}{2}, rate = \frac{\lambda_k^2}{2}\right)$,
- $\underset{J\times J}{\mathbf{\Sigma}}|d, \underset{J\times J}{\mathbf{Q}} \sim Inverse\ Wishart\left(df = d, scale = \underset{J\times J}{\mathbf{Q}}\right)$,
- $\pi_{0k}|\theta_\beta \overset{iid}{\sim} Bernoulli\,(\theta_\beta)$,
- $\theta_\beta|a,b \sim Beta\,(a,b)$,
- $\sigma^2|\alpha,\gamma \overset{ind.}{\sim} Inverse\ Gamma\,(shape = \alpha, scale = \gamma)$,
- $\lambda_k^2|r,\delta \overset{ind.}{\sim} Gamma\,(shape = r, rate = \delta)$,
- $\underset{nq\times 1}{\mathbf{b}} | \underset{q\times q}{\mathbf{G}} \sim N_{nq}\left(\underset{nq\times 1}{\mathbf{0}}, \underset{n\times n}{\mathbf{I}_n} \otimes \underset{q\times q}{\mathbf{G}}\right)$, and
- $\underset{q\times q}{\mathbf{G}}|\nu_o, \underset{q\times q}{\mathbf{C_0}} \sim Inverse\ Wishart\left(df = \nu_o, scale = \underset{q\times q}{\mathbf{C_0}}\right)$.

1.2 Gibbs sampler

Let

- $\underset{Jp_k \times Jp_k}{\mathbf{\Sigma}_k} \equiv \left[\tau_k^2 \underset{Jp_k \times nJ}{\mathbf{X}_k^T} \underset{nJ \times Jp_k}{\mathbf{X}_k} + \sigma^2 \left(\underset{J \times J}{\mathbf{\Sigma}} \otimes \underset{p_k \times p_k}{\mathbf{I}_{p_k}}\right)^{-1}\right]$,
- $\underset{Jn \times 1}{\mathbf{z}_k} \equiv vec\left(\underset{J \times n}{\mathbf{Y}^T}\right) - \left(\underset{nJ \times Jp_{\neg k}}{\mathbf{X}_{\neg k}} \, vec\left(\underset{p_{\neg k} \times J}{\mathbf{B}_{\neg k}}\right) + \underset{nJ \times nq}{\mathbf{Z}} \underset{nq \times 1}{\mathbf{b}}\right)$,
- $\underset{Jn \times 1}{\mathbf{r}} \equiv vec\left(\underset{J \times n}{\mathbf{Y}^T}\right) - \left(\underset{nJ \times Jp}{\mathbf{X}} \underset{pJ \times 1}{\beta} + \underset{nJ \times nq}{\mathbf{Z}} \underset{nq \times 1}{\mathbf{b}}\right)$,
- $\underset{nq \times nq}{\mathbf{\Sigma}_{\mathbf{b}}} \equiv \left[\underset{nq \times nJ}{\mathbf{Z}^T} \underset{nJ \times nq}{\mathbf{Z}} + \sigma^2 \left(\underset{n \times n}{\mathbf{I}_n} \otimes \underset{q \times q}{\mathbf{G}}\right)^{-1}\right]$,
- $\underset{Jn \times 1}{\mathbf{r}_{\mathbf{b}}} \equiv vec\left(\underset{J \times n}{\mathbf{Y}^T}\right) - \underset{nJ \times Jp}{\mathbf{X}} \underset{pJ \times 1}{\beta}$,
- $f_k \equiv \underset{1 \times Jn}{\mathbf{z}_k^T} \underset{nJ \times Jp_k}{\mathbf{X}_k} \underset{Jp_k \times Jp_k}{\mathbf{\Sigma}_k^{-1}} \underset{Jp_k \times nJ}{\mathbf{X}_k^T} \underset{Jn \times 1}{\mathbf{z}_k}$, and
- $\eta_k^2 \equiv \frac{1}{\tau_k^2}$.

Then, the full-conditional posterior distributions for our subject-specific group-level feature selection model Gibbs sampler are given below:

- $vec\left(\underset{p_k \times J}{\mathbf{B}_k}\right) | rest \overset{ind.}{\sim} \pi_{0k} N_{p_k J}\left(\tau_k^2 \underset{Jp_k \times Jp_k}{\mathbf{\Sigma}_k^{-1}} \underset{Jp_k \times nJ}{\mathbf{X}_k^T} \underset{Jn \times 1}{\mathbf{z}_k}, \sigma^2 \tau_k^2 \underset{Jp_k \times Jp_k}{\mathbf{\Sigma}_k^{-1}}\right) + (1 - \pi_{0k}) \, \delta_0 \left(vec\left(\underset{p_k \times J}{\mathbf{B}_k}\right)\right)$,
- $\eta_k^2 | rest \overset{ind.}{\sim} \begin{cases} Inverse\ Gaussian\left(\left(\frac{\lambda_k^2}{tr\left(\underset{p_k \times J}{\mathbf{B}_k} \underset{J \times J}{\mathbf{\Sigma}^{-1}} \underset{J \times p_k}{\mathbf{B}_k^T}\right)}\right)^{\frac{1}{2}}, \lambda_k^2\right) & if\ \pi_{0k} = 1 \\ Inverse\ Gamma\left(shape = \frac{p_k J + 1}{2}, scale = \frac{\lambda_k^2}{2}\right) & if\ \pi_{0k} = 0 \end{cases}$,
- $\underset{J \times J}{\mathbf{\Sigma}} | rest \sim Inverse\ Wishart\left(df = d + \sum_{k=1}^{K} \pi_{0k} p_k, scale = \underset{J \times p}{\mathbf{B}^T} \, diag\left(\frac{1}{\tau_k^2} \underset{p_k \times p_k}{\mathbf{I}_{p_k}}\right) \underset{p \times J}{\mathbf{B}} + \underset{J \times J}{\mathbf{Q}}\right)$,

- $\pi_{0k}|rest \overset{ind.}{\sim} Bernoulli\left(\frac{\theta_\beta(\sigma^2)^{\frac{Jp_k}{2}} \det\left(\underset{J\times J}{\mathbf{\Sigma}} \otimes \underset{p_k\times p_k}{\mathbf{I}_{p_k}}\right)^{-\frac{1}{2}} \exp\left\{\frac{\tau_k^2}{2\sigma^2} f_k\right\} \det\left(\underset{Jp_k\times Jp_k}{\mathbf{\Sigma}_k^{-1}}\right)^{\frac{1}{2}}}{(1-\theta)+\theta_\beta(\sigma^2)^{\frac{Jp_k}{2}} \det\left(\underset{J\times J}{\mathbf{\Sigma}} \otimes \underset{p_k\times p_k}{\mathbf{I}_{p_k}}\right)^{-\frac{1}{2}} \exp\left\{\frac{\tau_k^2}{2\sigma^2} f_k\right\} \det\left(\underset{Jp_k\times Jp_k}{\mathbf{\Sigma}_k^{-1}}\right)^{\frac{1}{2}}}\right)$,

- $\theta_\beta|rest \sim Beta\left(a + \sum_{k=1}^{K} \pi_{0k}, b + K - \sum_{k=1}^{K} \pi_{0k}\right)$,

- $\sigma^2|rest \sim Inverse\ Gamma\left(shape = \frac{nJ}{2} + \alpha, scale = \frac{\underset{1\times Jn}{\mathbf{r}^T}\ \underset{Jn\times 1}{\mathbf{r}}}{2} + \gamma\right)$,

- $\lambda_k^2|rest \overset{ind.}{\sim} Gamma\left(shape = \frac{p_kJ+1}{2} + r, rate = \frac{\tau_k^2}{2} + \delta\right)$,

- $\underset{nq\times 1}{\mathbf{b}}|rest \sim N_{nq}\left(\underset{nq\times nq}{\mathbf{\Sigma}_{\mathbf{b}}^{-1}}\ \underset{nq\times nJ}{\mathbf{Z}^T}\ \underset{Jn\times 1}{\mathbf{r}_{\mathbf{b}}}, \sigma^2 \underset{nq\times nq}{\mathbf{\Sigma}_{\mathbf{b}}^{-1}}\right)$, and

- $\underset{q\times q}{\mathbf{G}}|rest \sim Inverse\ Wishart\left(df = n + \nu_o, scale = \sum_{i=1}^{n} \underset{q\times 1}{\mathbf{b}_i}\ \underset{1\times q}{\mathbf{b}_i^T} + \underset{q\times q}{\mathbf{C_0}}\right)$.

2 Alzheimer's Disease Neuroimaging Initiative (ADNI) Feature Data

Below is a table 1 detailing all of the $p = 44$ features we used for our applied analysis.

Bibliography

[1] Xiaofan Xu and Malay Ghosh. Bayesian Variable Selection and Estimation for Group Lasso. *Bayesian Analysis*, 10(4):909–936, 2015.

[2] Benoit Liquet, Kerrie Mengersen, Anthony Pettitt, and Matthew Sutton. Bayesian variable selection regression of multivariate responses for group data. *Bayesian Analysis*, 12(4):1039–1067, 2017.

Feature (l)	Feature Group (k)	p_k	Feature Description
ABETA	CSF biomarkers	2	Amyloid beta peptide
TAU	CSF biomarkers	2	Total tau protein
Age	Demographics	5	Age
Education	Demographics	5	Education level (in years)
Sex	Demographics	5	Sex (female vs. male)
Race	Demographics	5	Race (white vs. other)
Marriage	Demographics	5	Marital status (married vs. other)
MCI	Neurological diagnoses	2	MCI vs. cognitively normal diagnosis
Dementia	Neurological diagnoses	2	Dementia vs. cognitively normal diagnosis
APOE4 1	Genetic markers	2	1 vs. 0 APOE-ϵ4 alleles
APOE4 2	Genetic markers	2	2 vs. 0 APOE-ϵ4 alleles
Hippocampus	MRI measurements	7	Volumetric quantification of the hippocampus
Ventricles	MRI measurements	7	Volumetric quantification of the ventricles
Whole Brain	MRI measurements	7	Volumetric quantification of the whole brain
Entorhinal	MRI measurements	7	Volumetric quantification of the entorhinal cortex
Fusiform	MRI measurements	7	Volumetric quantification of the fusiform gyrus
MidTemp	MRI measurements	7	Volumetric quantification of the middle temporal gyrus
ICV	MRI measurements	7	Intracerebral volume
MOCA	NP assessments/ECog	24	Montreal Cognitive Assessment
CDRSB	NP assessments/ECog	24	Clinical Dementia Rating sum of boxes
ADAS11	NP assessments/ECog	24	Alzheimer's Disease Assessment Subscale (11 task version)
ADASQ4	NP assessments/ECog	24	Alzheimer's Disease Assessment Cognitive Delayed Recall Task
MMSE	NP assessments/ECog	24	Mini-Mental State Examination
RAVLT Learning	NP assessments/ECog	24	Rey's Auditory Verbal Learning Test Learning
RAVLT Forgetting	NP assessments/ECog	24	Rey's Auditory Verbal Learning Test Forgetting
RAVLT Percent Forgetting	NP assessments/ECog	24	Rey's Auditory Verbal Learning Test Percent Forgetting
FAQ	NP assessments/ECog	24	Functional Assessment Questionnaire
LDELTOTAL	NP assessments/ECog	24	Logical Memory – Delayed Recall
TRABSCOR	NP assessments/ECog	24	Trail Making B Digit Symbol Substitution Test of the Wechsler Adult Intelligence Scale–Revised
ECog Pt Memory	NP assessments/ECog	24	ECog Participant Memory
ECog Pt Language	NP assessments/ECog	24	ECog Participant Language
ECog Pt Visuospatial	NP assessments/ECog	24	ECog Participant Visuospatial Abilities
ECog Pt Planning	NP assessments/ECog	24	ECog Participant Planning
ECog Pt Organization	NP assessments/ECog	24	ECog Participant Organization
ECog Pt Divided Attention	NP assessments/ECog	24	ECog Participant Divided Attention
ECog Pt Total	NP assessments/ECog	24	Average ECog Participant Score
ECog SP Memory	NP assessments/ECog	24	ECog Study Partner Assessment of Memory
ECog SP Language	NP assessments/ECog	24	ECog Study Partner Assessment of Language
ECog SP Visuospatial	NP assessments/ECog	24	ECog Study Partner Assessment of Visuospatial Abilities
ECog SP Planning	NP assessments/ECog	24	ECog Study Partner Assessment of Planning
ECog SP Organization	NP assessments/ECog	24	ECog Study Partner Assessment of Organization
ECog SP Divided Attention	NP assessments/ECog	24	ECog Study Partner Assessment of Divided Attention
FDG	PET measurements	2	Average FDG PET of the angular, temporal, and posterior cingulate
AV45	PET measurements	2	Average Florbetapir F 18 PET of THE whole cerebellum

Table 1: Baseline feature data from the ADNI; we consider a total of $p = 44$ features partitioned into $K = 7$ feature groups. **Abbreviations**: **MCI**: mild cognitive impairment, **APOE**: Apolipoprotein E, **NP**: neuropsychological, **ECog**: Everyday Cognition, and **FDG**: Fluorodeoxyglucose.