

# Some supplements

## 1 Convexity

**Theorem 1.1.**

$$\mathcal{L}_{\mathcal{Y},\Omega}(\Theta, \mathbf{b}) = \sum_{\omega \in \Omega} \sum_{\ell \in [L]} \left\{ \mathbb{1}_{\{y_\omega = \ell\}} \log [f(\theta_\omega + b_\ell) - f(\theta_\omega + b_{\ell-1})] \right\}, \text{ where } f(x) = \frac{e^x}{1 + e^x}.$$

is a concave function to  $(\Theta, \mathbf{b})$

*Proof.* It is enough to show that  $\lambda(u, v) = \log [f(u) - f(v)]$  is a concave function to  $(u, v)$  where  $u > v$ . Because if  $\lambda(u, v)$  is a concave function and  $u, v$  are both linear functions of  $(\Theta, \mathbf{b})$  such that  $u = \mathbf{a}_1^T(\Theta, \mathbf{b})$ ,  $v = \mathbf{a}_2^T(\Theta, \mathbf{b})$ , then  $\lambda(u, v) = \lambda(\mathbf{a}_1^T(\Theta, \mathbf{b}), \mathbf{a}_2^T(\Theta, \mathbf{b}))$  is a concave function to  $(\Theta, \mathbf{b})$  by the definition of the convexity. With the fact that  $\mathcal{L}_{\mathcal{Y},\Omega}(\Theta, \mathbf{b})$  can be written as the form of summations of  $\lambda(u, v)$ , we can conclude that  $\mathcal{L}_{\mathcal{Y},\Omega}(\Theta, \mathbf{b})$  is a concave function to  $(\Theta, \mathbf{b})$  if we prove  $\lambda(u, v)$  is concave. Write  $\lambda(u, v) = \log [f(u) - f(v)]$  as  $\log \left[ \int \mathbb{1}_{(u,v)}(x) f'(x) dx \right]$ . Notice that  $\log \mathbb{1}_{(u,v)}(x)$  is concave to  $(u, v, x)$  and  $\log f'(x)$  is concave to  $x$  because  $f'(x) = \frac{e^x}{(1+e^x)^2}$ . Thus,  $\log \mathbb{1}_{(u,v)}(x) f'(x)$  is a concave function to  $(u, v, x)$ . Therefore, the  $\lambda(u, v)$  is a concave function because the following lemma says that the integral of a log concave function with respect to some of its arguments is a log concave function of its remaining argument.  $\square$

**Lemma 1** (Corollary 3.5 in [Brascamp & Lieb \(2002\)](#)). *Let  $F(x, y) : \mathbb{R}^{m+n} \rightarrow \mathbb{R}$  be an integrable function where  $x \in \mathbb{R}^m, y \in \mathbb{R}^n$ . Let*

$$G(x) = \int_{\mathbb{R}^n} F(x, y) dy.$$

*If  $F(x, y)$  is a log concave function to  $(x, y)$ , then  $G(x)$  is a log concave function.*

## 2 Property of an estimator

**Property 1.**  $y_\omega^{(\text{Median})}$  minimizes  $R(y) = \mathbb{E}_{\hat{\theta}_\omega, \hat{\mathbf{b}}} |y_\omega - y|$ .

*Proof.* We can write  $R(y) = \sum_{\ell \in [L]} |\ell - y| \hat{f}_\ell$  where  $\hat{f}_\ell = f_\ell(\hat{\theta}_\omega, \hat{\mathbf{b}})$ . Let us denote  $\mu = y_\omega^{(\text{Median})}$ . Then,

$$R(y) = \begin{cases} [-(\hat{f}_1 \cdots + \hat{f}_L)]y + [(L\hat{f}_L + \cdots + 1\hat{f}_1)], & \text{if } y \in (-\infty, 1], \\ [\hat{f}_1 - (\hat{f}_2 \cdots + \hat{f}_L)]y + [(L\hat{f}_L + \cdots + 2\hat{f}_2) - 1\hat{f}_1], & \text{if } y \in (1, 2], \\ \vdots & \vdots \\ [(\hat{f}_1 + \cdots + \hat{f}_{\mu-1}) - (\hat{f}_\mu \cdots + \hat{f}_L)]y + [(L\hat{f}_L + \cdots + \mu\hat{f}_\mu) - ((\mu-1)\hat{f}_{\mu-1} + \cdots + 1\hat{f}_1)], & \text{if } y \in (\mu-1, \mu] \\ [(\hat{f}_1 + \cdots + \hat{f}_\mu) - (\hat{f}_{\mu+1} \cdots + \hat{f}_L)]y + [(L\hat{f}_L + \cdots + (\mu+1)\hat{f}_{\mu+1}) - (\mu\hat{f}_\mu + \cdots + 1\hat{f}_1)], & \text{if } y \in (\mu, \mu+1] \\ \vdots & \vdots \\ [(\hat{f}_1 + \cdots + \hat{f}_{L-1}) - \hat{f}_L]y + [(L\hat{f}_L - ((L-1)\hat{f}_{L-1} + \cdots + 1\hat{f}_1)], & \text{if } y \in (L-1, L] \\ [(\hat{f}_1 + \cdots + \hat{f}_L)]y + [(L\hat{f}_L + \cdots + 1\hat{f}_1)], & \text{if } y \in (L, \infty]. \end{cases}$$

Based on the above formula, we can find out that  $R(y)$  is minimized when  $y = \mu$  because  $(\hat{f}_1 + \cdots + \hat{f}_{\mu-1}) - (\hat{f}_\mu \cdots + \hat{f}_L) < 0$  and  $(\hat{f}_1 + \cdots + \hat{f}_\mu) - (\hat{f}_{\mu+1} \cdots + \hat{f}_L) > 0$  by the definition of  $\mu$ .  $\square$

### 3 Theorems in the case when the cut points is unknown

When the cut points  $\mathbf{b}$  is unknown, we estimate  $(\hat{\Theta}, \hat{\mathbf{b}})$  by alternately fixing each factor and proceed until each factor does not change from each update. We have the following relationship between  $\hat{\Theta}$  and  $\hat{\mathbf{b}}$

$$\hat{\Theta} = \arg \max_{\Theta \in \mathcal{P}} \mathcal{L}_{\mathcal{Y}, \Omega}(\Theta | \hat{\mathbf{b}}) \quad \text{and} \quad \hat{\mathbf{b}} = \arg \max_{\mathbf{b} \in \mathcal{B}} \mathcal{L}_{\mathcal{Y}, \Omega}(\mathbf{b} | \hat{\Theta}).$$

We assess the estimation accuracy using the mean squared error (MSE):

$$\text{MSE}(\hat{\Theta}, \hat{\mathbf{b}}) = \frac{1}{\prod_k d_k} \|(\hat{\Theta}, \hat{\mathbf{b}}) - (\Theta^{\text{true}}, \mathbf{b}^{\text{true}})\|_F.$$

We show that  $\frac{1}{\prod_k d_k} \|(\hat{\Theta}, \hat{\mathbf{b}}) - (\Theta^{\text{true}}, \mathbf{b}^{\text{true}})\|_F$  is bounded by  $o(\frac{1}{\prod_k d_k})$  with high probability so that

$$\frac{1}{\prod_k d_k} \|(\hat{\Theta}, \hat{\mathbf{b}}) - (\Theta^{\text{true}}, \mathbf{b}^{\text{true}})\|_F \leq \frac{1}{\prod_k d_k} \|\hat{\Theta} - \Theta^{\text{true}}\|_F + o(\frac{1}{\prod_k d_k}).$$

Therefore, we can utilize Theorem 4.1 to establish the upper bound for the proposed  $(\hat{\Theta}, \hat{\mathbf{b}})$ . We define a few key quantities that will be used in the proof.

$$U_\alpha = \max_{\ell \in [L-1], |\theta| \leq \alpha} \max \left( \frac{\dot{f}(\theta + b_\ell)}{f(\theta + b_\ell) - f(\theta + b_{\ell-1})}, \frac{\dot{f}(\theta + b_\ell)}{f(\theta + b_{\ell+1}) - f(\theta + b_\ell)} \right)$$

$$L_\alpha = \min_{\ell \in [L-1], |\theta| \leq \alpha} \min \left( -\frac{\ddot{f}(\theta + b_\ell)[f(\theta + b_\ell) - f(\theta + b_{\ell-1})] - \dot{f}(\theta + b_\ell)[\dot{f}(\theta + b_\ell) - \dot{f}(\theta + b_{\ell-1})]}{[f(\theta + b_\ell) - f(\theta + b_{\ell-1})]^2}, \right. \\ \left. \frac{\ddot{f}(\theta + b_\ell)[f(\theta + b_{\ell+1}) - f(\theta + b_\ell)] - \dot{f}(\theta + b_\ell)[\dot{f}(\theta + b_{\ell+1}) - \dot{f}(\theta + b_\ell)]}{[f(\theta + b_{\ell+1}) - f(\theta + b_\ell)]^2} \right)$$

We can see  $U_\alpha > 0$  and  $L_\alpha > 0$  from the assumption. Especially when  $f(x) = \frac{1}{1+e^{-x}}$ , we have

$$L_\alpha = \min_{\ell \in [L-1], |\theta| \leq \alpha} \min \left( \frac{e^{\theta+b_\ell}(e^{\theta+b_\ell} - e^{\theta+b_{\ell-1}})^2}{(e^{\theta+b_\ell} + 1)^4(e^{\theta+b_{\ell-1}} + 1)^2}, \frac{e^{\theta+b_\ell}(e^{\theta+b_{\ell+1}} - e^{\theta+b_\ell})^2}{(e^{\theta+b_{\ell+1}} + 1)^2(e^{\theta+b_\ell} + 1)^4} \right).$$

**Theorem 3.1.** *From the same setting in Theorem 4.1 except the knowledge of  $\mathbf{b}$ , with very high probability, the estimator satisfies*

$$\frac{1}{\prod_k d_k} \|(\hat{\Theta}, \hat{\mathbf{b}}) - (\hat{\Theta}, \mathbf{b}^{\text{true}})\|_F = o(\frac{1}{\prod_k d_k}).$$

*Proof.* It follows from the expression of  $\mathcal{L}_{\mathcal{Y}, \Omega}(\Theta, \mathbf{b})$  that

$$\frac{\partial \mathcal{L}_{\mathcal{Y}}}{\partial \beta_\ell} = \sum_{\omega \in \Omega} \left[ \mathbb{1}_{\{y_\omega = \ell\}} \frac{\dot{f}(\theta_\omega + b_\ell)}{f(\theta_\omega + b_\ell) - f(\theta_\omega + b_{\ell-1})} - \mathbb{1}_{\{y_\omega = \ell+1\}} \frac{\dot{f}(\theta_\omega + b_\ell)}{f(\theta_\omega + b_{\ell+1}) - f(\theta_\omega + b_\ell)} \right],$$

$$\frac{\partial^2 \mathcal{L}_{\mathcal{Y}}}{\partial b_\ell^2} = \sum_{\omega \in \Omega} \left[ \mathbb{1}_{\{y_\omega = \ell\}} \frac{\ddot{f}(\theta_\omega + b_\ell)[f(\theta_\omega + b_\ell) - f(\theta_\omega + b_{\ell-1})] - \dot{f}(\theta_\omega + b_\ell)^2}{[f(\theta_\omega + b_\ell) - f(\theta_\omega + b_{\ell-1})]^2} \right. \\ \left. - \mathbb{1}_{\{y_\omega = \ell+1\}} \frac{\ddot{f}(\theta_\omega + b_\ell)[f(\theta_\omega + b_{\ell+1}) - f(\theta_\omega + b_\ell)] + \dot{f}(\theta_\omega + b_\ell)^2}{[f(\theta_\omega + b_{\ell+1}) - f(\theta_\omega + b_\ell)]^2} \right] \text{ for } \ell \in [L-1],$$

$$\frac{\partial^2 \mathcal{L}_{\mathcal{Y}}}{\partial b_\ell \partial b_{\ell+1}} = \sum_{\omega \in \Omega} \mathbb{1}_{\{y_\omega = \ell+1\}} \frac{\dot{f}(\theta_\omega + b_\ell) \dot{f}(\theta_\omega + b_{\ell+1})}{[f(\theta_\omega + b_{\ell+1}) - f(\theta_\omega + b_\ell)]^2} \text{ for } \ell \in [L-2] \text{ and } \frac{\partial^2 \mathcal{L}_{\mathcal{Y}}}{\partial b_i \partial b_j} = 0 \text{ if } |i - j| > 1.$$

Therefore, all entries in  $\frac{1}{\prod_k d_k} \nabla_{\mathbf{b}} \mathcal{L}_{\mathcal{Y}}$  are upper bounded  $U > 0$ , and  $\frac{1}{\prod_k d_k} \nabla_{\mathbf{b}}^2 \mathcal{L}_{\mathcal{Y}}$  is a tridiagonal matrix.

By the second-order Taylor's expansion of  $\mathcal{L}_{\mathcal{Y}}(\hat{\mathbf{b}}|\hat{\Theta})$  around  $\mathbf{b}^{\text{true}}$ , we obtain

$$\mathcal{L}_{\mathcal{Y}}(\hat{\mathbf{b}}|\hat{\Theta}) = \mathcal{L}_{\mathcal{Y}}(\mathbf{b}^{\text{true}}|\hat{\Theta}) + (\mathbf{b}^{\text{true}} - \hat{\mathbf{b}})^T \nabla_{\mathbf{b}} \mathcal{L}_{\mathcal{Y}}(\mathbf{b}^{\text{true}}) + (\mathbf{b}^{\text{true}} - \hat{\mathbf{b}})^T \nabla_{\mathbf{b}}^2 \mathcal{L}_{\mathcal{Y}}(\check{\mathbf{b}})(\mathbf{b}^{\text{true}} - \hat{\mathbf{b}}), \quad (1)$$

$\check{\mathbf{b}} = \gamma \mathbf{b}^{\text{true}} + (1 - \gamma) \hat{\mathbf{b}}$  for some  $\gamma \in [0, 1]$ , and  $\nabla_{\mathbf{b}}^2 \mathcal{L}_{\mathcal{Y}}(\check{\mathbf{b}})$  denotes the  $(L - 1)$ -by- $(L - 1)$  Hessian matrix evaluated at  $\check{\mathbf{b}}$ .

We first bound the linear term in (1). Note that, by Cauchy-Schwartz inequality,

$$(\mathbf{b}^{\text{true}} - \hat{\mathbf{b}})^T \nabla_{\mathbf{b}} \mathcal{L}_{\mathcal{Y}}(\mathbf{b}^{\text{true}}) \leq \|\mathbf{b}^{\text{true}} - \hat{\mathbf{b}}\| \|\nabla_{\mathbf{b}} \mathcal{L}_{\mathcal{Y}}(\mathbf{b}^{\text{true}})\| \leq \|\mathbf{b}^{\text{true}} - \hat{\mathbf{b}}\| U_{\alpha} \sqrt{L} \prod_k d_k.$$

The last inequality is followed by

$$\left. \frac{\partial \mathcal{L}_{\mathcal{Y}}}{\partial b_{\ell}} \right|_{\mathbf{b}=\mathbf{b}^{\text{true}}} \leq U_{\alpha} \quad \text{for all } \ell \in [L - 1].$$

We next bound the quadratic term in (1). Note that

$$\begin{aligned} & (\mathbf{b}^{\text{true}} - \hat{\mathbf{b}})^T \nabla_{\mathbf{b}}^2 \mathcal{L}_{\mathcal{Y}}(\check{\mathbf{b}})(\mathbf{b}^{\text{true}} - \hat{\mathbf{b}}) \\ &= \sum_{\ell \in [L-1]} \left( \left. \frac{\partial^2 \mathcal{L}_{\mathcal{Y}}}{\partial b_{\ell}^2} \right|_{\mathbf{b}=\check{\mathbf{b}}} \right) (\hat{b}_{\ell} - b_{\text{true},\ell})^2 + 2 \sum_{\ell \in [L-1]-\{1\}} \left( \left. \frac{\partial^2 \mathcal{L}_{\mathcal{Y}}}{\partial b_{\ell} \partial b_{\ell-1}} \right|_{\mathbf{b}=\check{\mathbf{b}}} \right) (\hat{b}_{\ell} - b_{\text{true},\ell})(\hat{b}_{\ell-1} - b_{\text{true},\ell-1}) \\ &\leq \sum_{\ell \in [L-1]} \left( \left. \frac{\partial^2 \mathcal{L}_{\mathcal{Y}}}{\partial b_{\ell}^2} \right|_{\mathbf{b}=\check{\mathbf{b}}} \right) (\hat{b}_{\ell} - b_{\text{true},\ell})^2 + \sum_{\ell \in [L-1]-\{1\}} \left( \left. \frac{\partial^2 \mathcal{L}_{\mathcal{Y}}}{\partial b_{\ell} \partial b_{\ell-1}} \right|_{\mathbf{b}=\check{\mathbf{b}}} \right) [(\hat{b}_{\ell} - b_{\text{true},\ell})^2 + (\hat{b}_{\ell-1} - b_{\text{true},\ell-1})^2] \\ &= \left( \left. \frac{\partial^2 \mathcal{L}_{\mathcal{Y}}}{\partial b_1^2} + \frac{\partial^2 \mathcal{L}_{\mathcal{Y}}}{\partial b_1 \partial b_2} \right|_{\mathbf{b}=\check{\mathbf{b}}} \right) (\hat{b}_1 - b_{\text{true},1})^2 + \left( \left. \frac{\partial^2 \mathcal{L}_{\mathcal{Y}}}{\partial b_{L-1}^2} + \frac{\partial^2 \mathcal{L}_{\mathcal{Y}}}{\partial b_{L-2} \partial b_{L-1}} \right|_{\mathbf{b}=\check{\mathbf{b}}} \right) (\hat{b}_{L-1} - b_{\text{true},L-1})^2 \\ &+ \sum_{\ell \in [L-2]-\{1\}} \left( \left. \frac{\partial^2 \mathcal{L}_{\mathcal{Y}}}{\partial b_{\ell}^2} + \frac{\partial^2 \mathcal{L}_{\mathcal{Y}}}{\partial b_{\ell} \partial b_{\ell-1}} + \frac{\partial^2 \mathcal{L}_{\mathcal{Y}}}{\partial b_{\ell+1} \partial b_{\ell}} \right|_{\mathbf{b}=\check{\mathbf{b}}} \right) (\hat{b}_{\ell} - b_{\text{true},\ell})^2 \\ &\leq -L_{\alpha} \prod_k d_k \sum_{\ell \in [L-1]} (\hat{b}_{\ell} - b_{\text{true},\ell})^2 \\ &= -L_{\alpha} \prod_k d_k \|\hat{\mathbf{b}} - \mathbf{b}^{\text{true}}\|_F^2, \end{aligned}$$

Therefore, combining (1) and the above linear term and quadratic term results, we have that,

$$\mathcal{L}_{\mathcal{Y}}(\hat{\mathbf{b}}|\hat{\Theta}) \leq \mathcal{L}_{\mathcal{Y}}(\mathbf{b}^{\text{true}}|\hat{\Theta}) + U_{\alpha} \sqrt{L} \prod_k d_k \|\hat{\mathbf{b}} - \mathbf{b}^{\text{true}}\|_F - \frac{L_{\alpha}}{2} \prod_k d_k \|\hat{\mathbf{b}} - \mathbf{b}^{\text{true}}\|_F^2.$$

Since  $\hat{\mathbf{b}} = \arg \max_{\mathbf{b} \in \mathcal{B}} \mathcal{L}_{\mathcal{Y}}(\mathbf{b}|\hat{\Theta})$ ,  $\mathcal{L}_{\mathcal{Y}}(\hat{\mathbf{b}}|\hat{\Theta}) - \mathcal{L}_{\mathcal{Y}}(\mathbf{b}^{\text{true}}|\hat{\Theta}) \geq 0$ , which gives

$$U_{\alpha} \sqrt{L} \|\hat{\mathbf{b}} - \mathbf{b}^{\text{true}}\|_F - \frac{L_{\alpha}}{2} \|\hat{\mathbf{b}} - \mathbf{b}^{\text{true}}\|_F^2 \geq 0.$$

Henceforth,

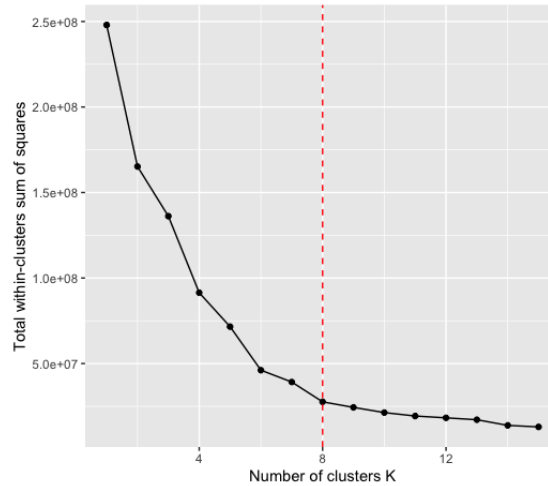
$$\|\hat{\mathbf{b}} - \mathbf{b}^{\text{true}}\|_F \leq \frac{2U_{\alpha} \sqrt{L}}{L_{\alpha}}.$$

Finally, this complete the proof.

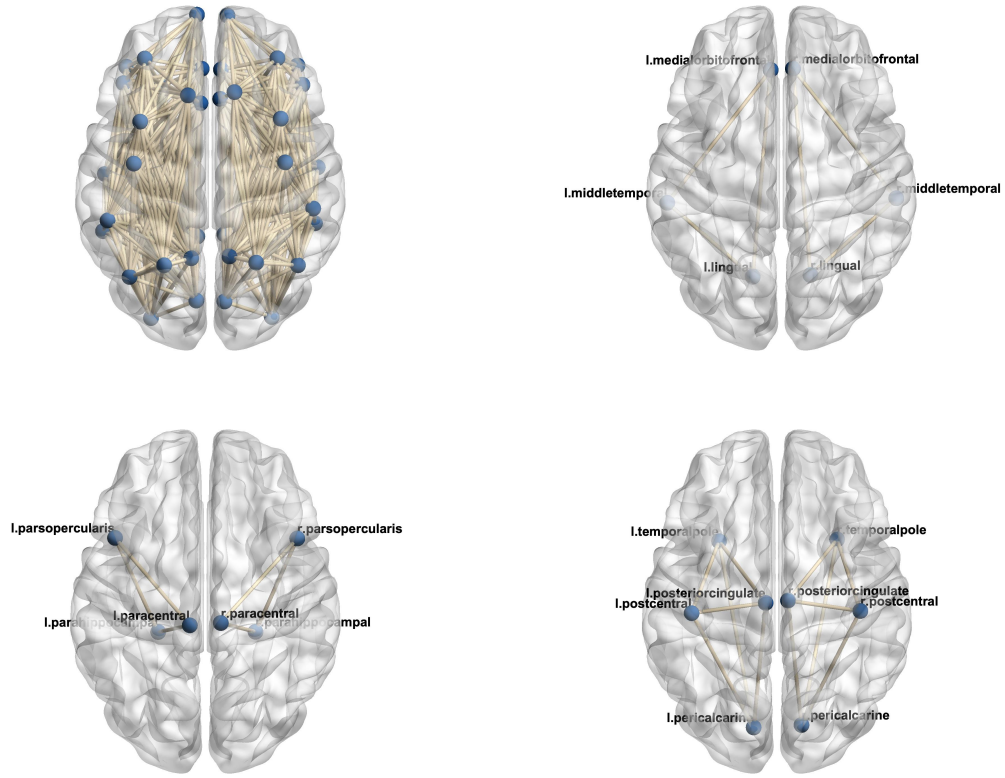
$$\frac{1}{\prod_k d_k} \|(\hat{\Theta}, \hat{\mathbf{b}}) - (\hat{\Theta}, \mathbf{b}^{\text{true}})\|_F = \frac{1}{\prod_k d_k} \|\hat{\mathbf{b}} - \mathbf{b}^{\text{true}}\|_F = o\left(\frac{1}{\prod_k d_k}\right).$$

Group	#1			#2		
Brain nodes	"RMF_L","Fpole_L","Insula_L", "SupF_L","CaudMF_L", "parsTriangularis_L","parsOpercularis_L", "preCentral_L","Tpole_L","SupT_L", "SupT_L","postCentral_L","SupP_L", "IP_L","LO_L","MOF_L","SupF_L", "IsthmusC_L","Precuneus_L", "Cuneus_L","paraHippo_L", "Lingual_L","SupT_L","LO_L"			"RMF_R","Fpole_R","Insula_R", "SupF_R","CaudMF_R", "parsTriangularis_R","parsOpercularis_R", "preCentral_R","Tpole_R","SupT_R", "SupT_R","postCentral_R","SupP_R", "IP_R","LO_R","MOF_R","SupF_R", "IsthmusC_R","Precuneus_R", "Cuneus_R","paraHippo_R", "Lingual_R","SupT_R","LO_R"		
Group	#3	#4	#5	#6	#7	#8
Brain nodes	"SupraM_L" "SupraM_L" "SupraM_L" "SupraM_L"	"SupraM_R" "SupraM_R" "SupraM_R" "SupraM_R"	"IT_L" "IT_L" "IT_L"	"IT_R" "IT_R" "IT_R"	"MT_L" "MT_L" "MT_L"	"MT_R" "MT_R" "MT_R"

## 4 Brain data clustering



Supplementary Figure S1: Brain cluster graph



Supplementary Figure S2: Brain cluster groups

□

## References

Brascamp, H. J. and Lieb, E. H. On extensions of the brunn-minkowski and prékopa-leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. In *Inequalities*, pp. 441–464. Springer, 2002.