

階層型成分法によるマルチモダル脳画像解析

佐賀大学医学部 地域医療科学教育研究センター

川口 淳 <http://biostat.med.saga-u.ac.jp>

- はじめに
 - » MUltimodal Brain Imaging Genetics
- 提案方法
 - » SMSMA (Supervised Multi-block Sparse Multivariable Analysis)
- 数値実験
- 実データ適用例
- まとめ

2017年度統計関連学会連合大会

2017年9月6日

南山大学

マルチモダル脳画像解析

➤ 画像の種類

- » (形態) sMRI, (機能) fMRI, dMRI, FDG PET, Amyloid/Tau PET

➤ 適用疾患

- » 精神障害 (認知症, 統合失調症, 気分障害, 物質関連障害)
- » 発達障害 (注意欠陥・多動性障害 (ADHD), 自閉症スペクトラム (ASD))
- » 不安障害, 境界性パーソナリティ障害

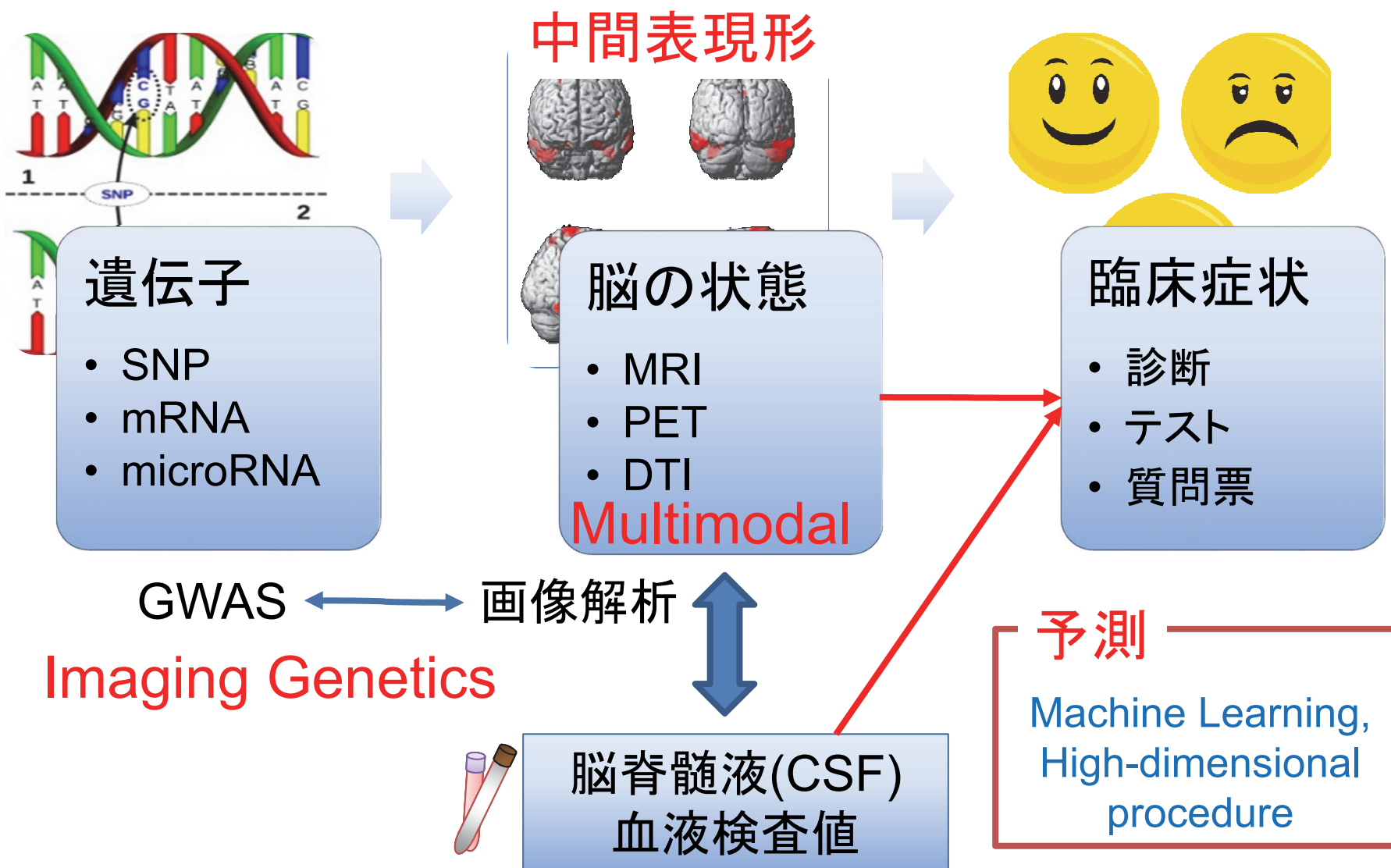
➤ 解析方法

- » SVM, Multi-kernel SVM, Random Forest, deep learning,
- » PCA, CCA, PLS, ICA, NMF...

Arbabshirani et al. (2017 NeuroImage), Meng et al. (2017 NeuroImage)

Calhoun and Sui (2016 Bio. Psy.) Vandenberghe Marsden (2015 *Ph. Med. Biol.*)

MUBIG (MULTImodal Brain Imaging Genetics)



目標

- MUBIGデータを統合的に解析できる統計手法を開発する
 - » 全脳を対象とするボクセルベース解析
 - » Genome-wide Study
- 二段階次元縮小をベースに
 - » データ駆動型(Data Driven), 探索的解析
 - » スコアに基づく予測(診断)モデル
- Multiblock PLS (CCA) をsupervise version に拡張

ADNI (Baseline)データと前処理

	Normal (n = 48)	AD (n = 52)
Age [year]	74.3	75.5
Male [n (%)]	32 (66.7)	31 (59.6)
Education duration [year]	15.7	14.2
MMSE	28.8	23.4

➤ 【 Z 】 $Z = 3.17CDRSB + 0.11ADAS13 - 0.57MMSE$

※ベンチマーク
(最終目標は, MCI converter)

➤ 【 Y1 】 MRI (GM) の前処理

- » VBM toolbox (segmentation, DARTEL normalization)
- » 3D B-spline 基底展開による次元縮小 ($q \doteq 5000$)

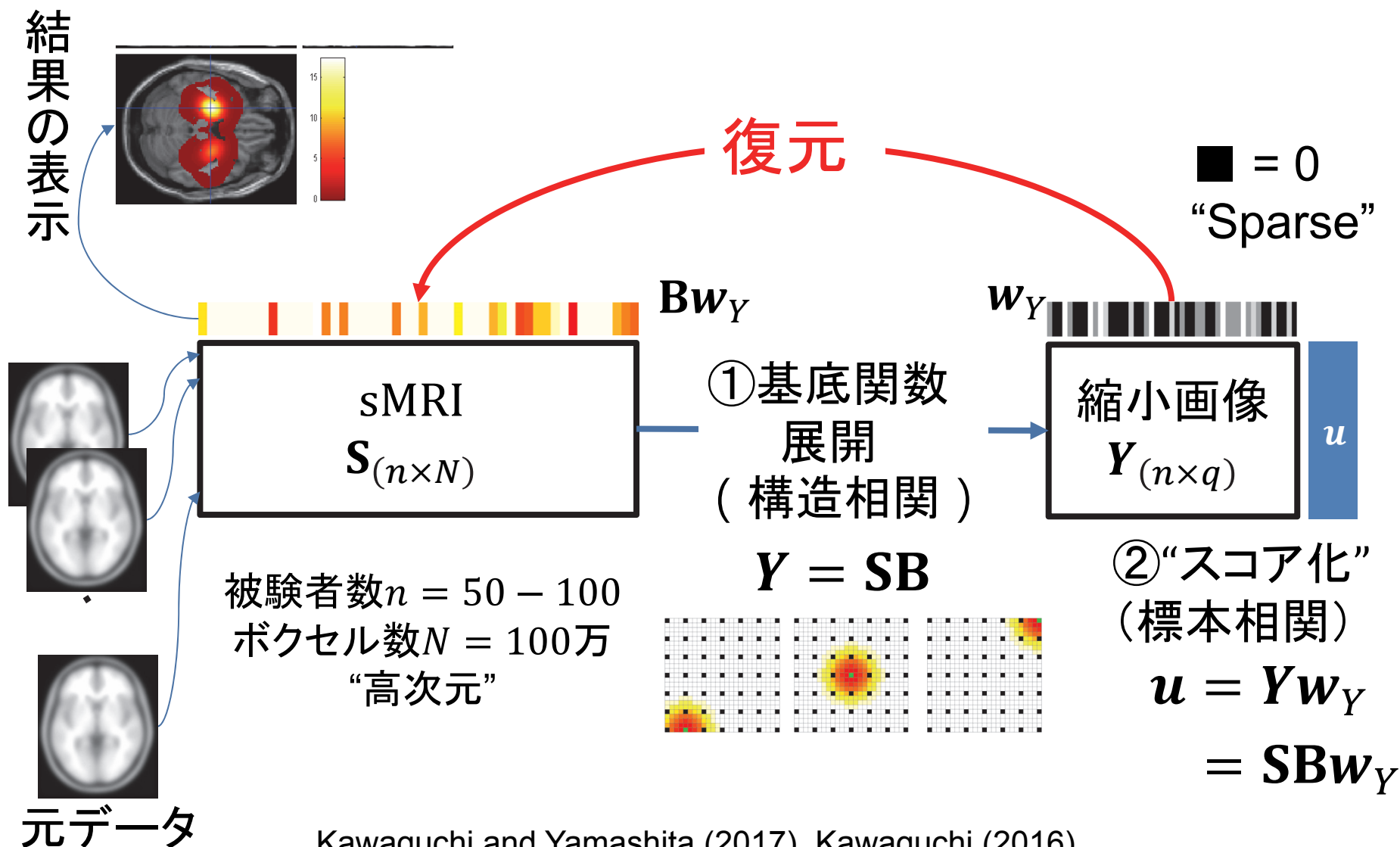
➤ 【 Y2 】 PET の前処理

- » co-registered to the bias-corrected T1 image and an intensity normalization to the GM

➤ 【 X 】 SNP の前処理

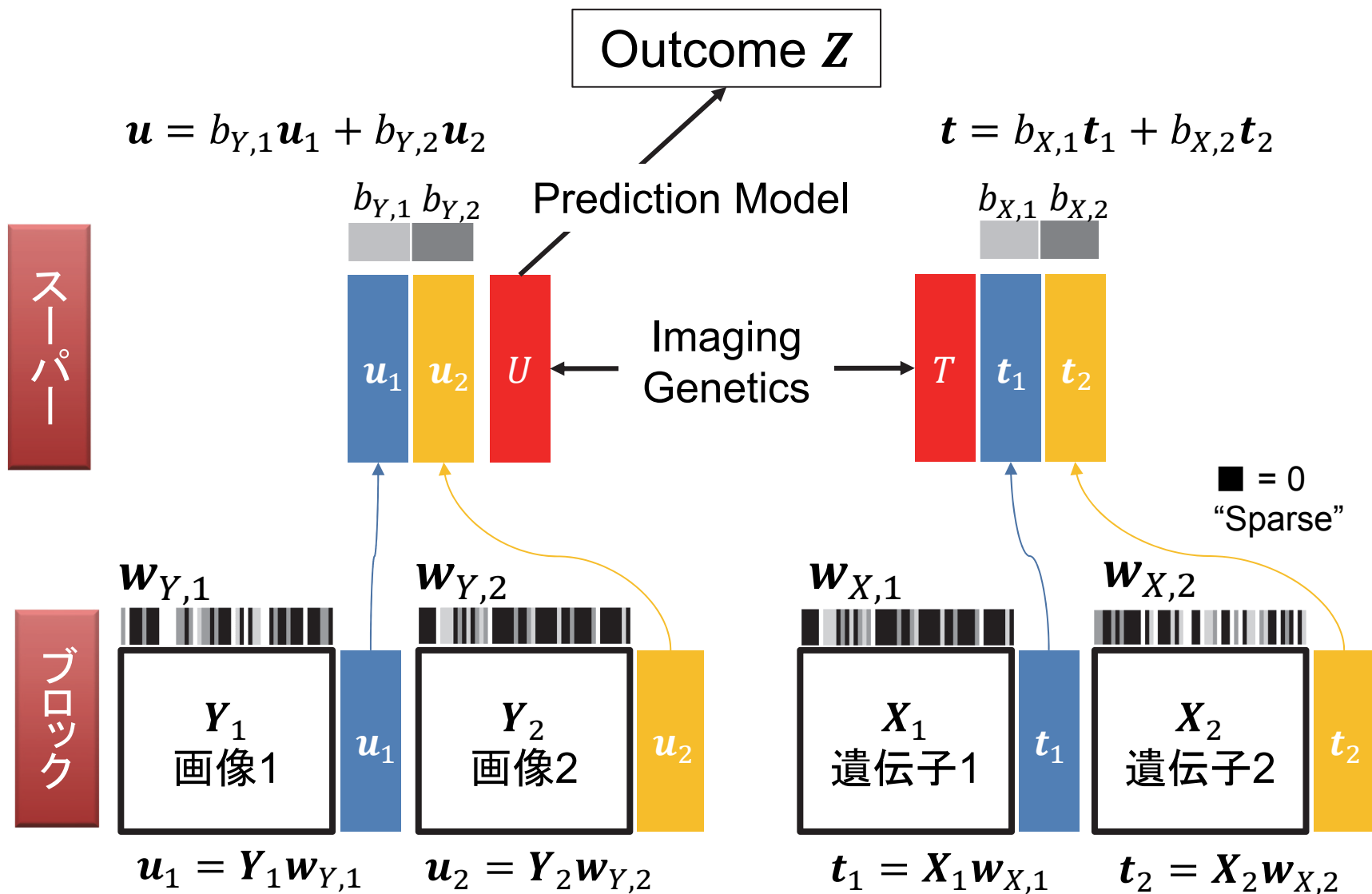
- » PLINK (maf=0.05, HWD p.level=0.001, call rate=0.95)
- » Mean Score within Haplotype Blocks ($p \doteq 5000$), Median imputation for missing

二段階次元縮小(脳画像の解析)



Kawaguchi and Yamashita (2017), Kawaguchi (2016),
Kawaguchi (2015), Araki et al. (2013), Yoshida et al. (2013),

階層成分(スコア)



階層成分(スコア)

- n 人の被験者を考える.
- \mathbf{X}_m $n \times p_m$ ($m = 1, 2, \dots, M_X$) 行列, \mathbf{Y}_m $n \times q_m$ ($m = 1, 2, \dots, M_Y$) 行列.
 - » M_X, M_Y はブロック(モダリティ)数
- \mathbf{Z} n 次元アウトカムベクトル

- X と Y のスコア表現.

$$\mathbf{t} = \sum_{m=1}^{M_X} b_{X,m} \mathbf{X}_m \mathbf{w}_{X,m}, \quad \mathbf{u} = \sum_{m=1}^{M_Y} b_{Y,m} \mathbf{Y}_m \mathbf{w}_{Y,m}$$

- 重みベクトル

- » $\mathbf{w}_X = (\mathbf{w}_{X,1}, \mathbf{w}_{X,2}, \dots, \mathbf{w}_{X,M_X})^\top$, $\mathbf{w}_Y = (\mathbf{w}_{Y,1}, \mathbf{w}_{Y,2}, \dots, \mathbf{w}_{Y,M_Y})^\top$ サブ重みベクトル
- » $\mathbf{b}_X = (b_{X,1}, b_{X,2}, \dots, b_{X,M_X})^\top$, $\mathbf{b}_Y = (b_{Y,1}, b_{Y,2}, \dots, b_{Y,M_Y})^\top$ スーパー重みベクトル

スコア間の共分散最大化

- 次の $L(\mathbf{b}_X, \mathbf{w}_X, \mathbf{b}_Y, \mathbf{w}_Y)$ を $\|\mathbf{w}_X\|^2 = 1, \|\mathbf{w}_Y\|^2 = 1, \|\mathbf{b}_X\|^2 = 1, \|\mathbf{b}_Y\|^2 = 1$ の制限下で最大化することにより求める

$$L(\mathbf{b}_X, \mathbf{w}_X, \mathbf{b}_Y, \mathbf{w}_Y)$$

$$= \mu_{XY} \mathbf{t}^\top \mathbf{u} + \mu_{XZ} \mathbf{t}^\top \mathbf{Z} + \mu_{YZ} \mathbf{u}^\top \mathbf{Z} - \sum_{m=1}^{M_X} P_{\lambda_{X,m}}(\mathbf{w}_{X,m}) - \sum_{m=1}^{M_Y} P_{\lambda_{Y,m}}(\mathbf{w}_{Y,m})$$

$$0 \leq \mu_{XZ}, \mu_{YZ}, \mu_{XY} \leq 1, \mu_{XY} = 1 - \mu_{XZ} - \mu_{YZ}, P_\lambda(x) = 2\lambda|x|.$$

$\mu_{XZ}, \mu_{YZ}, \mu_{XY}$ は先に与える. λ 's はBICで選択

Related works

Gross and Tibshirani (2015, biostat), Luo et al. (2016, biostat)

重みの更新

- X の重みに対する更新 (Y も同様)は,

$$\tilde{b}_{X,m} = \mathbf{t}_m^\top \{ \mu_{XY} \mathbf{u} + \mu_{XZ} \mathbf{Z} \}$$

$$\tilde{\mathbf{w}}_{X,m} = h_{\lambda_{X,m}} (b_{X,m} \mathbf{X}_m^\top \{ \mu_{XY} \mathbf{u} + \mu_{XZ} \mathbf{Z} \})$$

$$h_\lambda(y) = \text{sign}(y)(|y| - \lambda)_+ \quad \text{※要素毎}$$

- $\tilde{\mathbf{a}}$ のノルム1の $\hat{\mathbf{a}}$ に標準化

$$\hat{\mathbf{a}} = \begin{cases} \tilde{\mathbf{a}} / \|\tilde{\mathbf{a}}\| & (\text{if } \|\tilde{\mathbf{a}}\| > 0) \\ 0 & (\text{otherwise}) \end{cases},$$

アルゴリズム

1. t と u に初期値を与える
2. t と u が収束するまで以下を続ける:
 1. u を固定して, \hat{w}_X を更新して, b_X を更新
 2. t を固定して, \hat{w}_Y を更新する, b_Y を更新
 3. $t = X\hat{w}_X, u = Y\hat{w}_Y$.
3. deflation step (Zはしない)

$$X_m \leftarrow X_m - t_m p_m^\top, \quad Y_m \leftarrow Y_m - u_m q_m^\top$$

$$p_m = (X_m^\top t_m) / (t_m^\top t_m), \quad q_m = Y_m^\top u_m / u_m^\top u_m$$

上記1.2.3を繰り返す, 各componentのスコアと重みを得る

【単調性】

$$\begin{aligned} \gamma(s) &= \text{cov}(t^{(s)}, u^{(s)}) + \text{cov}(t^{(s)}, Z) + \text{cov}(u^{(s)}, Z) \\ &\Rightarrow \gamma(s) \leq \gamma(s+1) \end{aligned}$$

数値実験(設定)

$n = 50$ 繰り返し回数100

$$m = 1, 2$$

$$X = tw_X^-$$

$$Y_m = uw_{Y,m}^- + E$$

$$Z = uw_Z^T$$

$$w_X = \begin{bmatrix} \mathbf{1}_{10} & \mathbf{0}_{10} \\ \mathbf{0}_{10} & \mathbf{1}_{10} \\ \mathbf{0}_{80} & \mathbf{0}_{80} \end{bmatrix}^T$$

$$w_{Y,m} = s_m w_Y$$

$$w_Z = (0, \mathbf{1})^T$$

$$w_Y^{-1} = 0.2 \text{ (弱) or } 1 \text{ (強)}$$

$$s_m = (s_m(v_1), s_m(v_2), \dots, s_m(v_q))^T$$



$$E = e_0 E_0 \quad E_0 \sim N(\mathbf{0}, I_{n \times q})$$



$$e_0 = 0.001$$

$$q = 100 \times 100$$

$$t \sim N(\mathbf{0}, I_2)$$

$$u = tw_u^T + \varepsilon \quad \varepsilon \sim N(\mathbf{0}, I_2)$$

$$n \times 2$$

$$w_u = (1, w_{u,0})^T$$

$$w_{u,0} = 0.2 \text{ (弱) or } 0.8 \text{ (強)}$$





数値実験の結果 (検出平均)

Supervised の強さ($\mu_{XZ}=0$)

スコアと実現値の関連の強さ

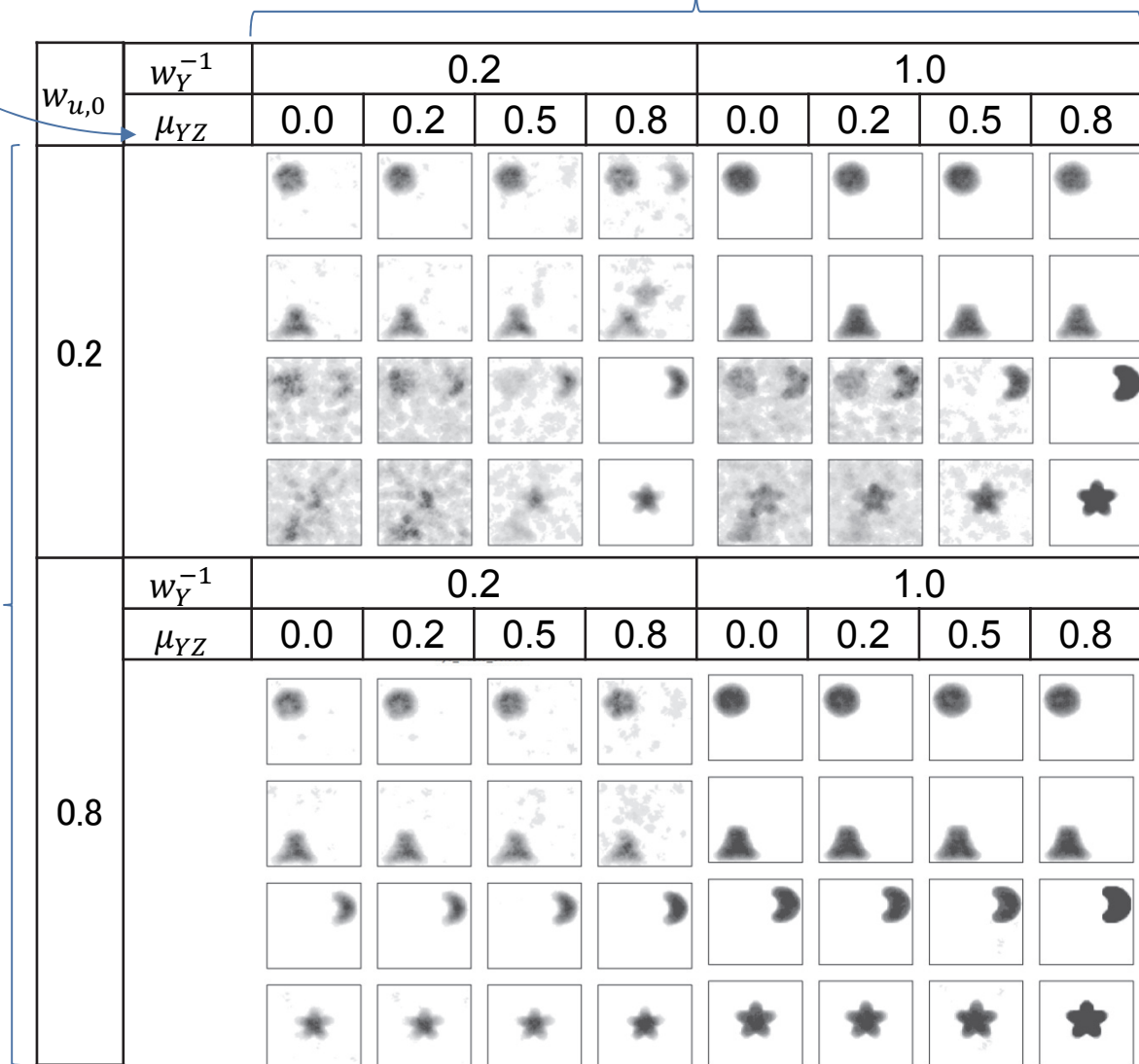
Z : Outcome

Y

comp	block	True
1	1	
	2	
2	1	
	2	

X: 100変数

XとYのスコアの関連の強さ



数値実験の結果(表)

sensitivity – (1- specificity)

Parameters			X			Y		
$w_{u,0}$	w_Y	μ_{YZ}	comp1	comp2	mean	comp1	comp2	mean
0.2	0.2	0	0.979	0.803	0.891	0.722	0.114	0.418
		0.2	0.976	0.849	0.912	0.712	0.135	0.424
		0.5	0.959	0.907	0.933	0.663	0.401	0.532
		0.8	0.909	0.896	0.902	0.555	0.915	0.735
	1.0	0	0.971	0.806	0.888	0.880	0.112	0.496
		0.2	0.973	0.849	0.911	0.860	0.164	0.512
		0.5	0.976	0.880	0.928	0.649	0.380	0.515
		0.8	0.920	0.884	0.902	0.720	0.892	0.806
0.8	0.2	0	0.958	0.956	0.957	0.812	0.850	0.831
		0.2	0.959	0.956	0.957	0.807	0.847	0.827
		0.5	0.978	0.961	0.969	0.759	0.736	0.747
		0.8	0.961	0.946	0.953	0.794	0.932	0.863
	1.0	0	0.956	0.959	0.957	0.905	0.869	0.887
		0.2	0.953	0.956	0.954	0.897	0.874	0.885
		0.5	0.972	0.951	0.962	0.754	0.742	0.748
		0.8	0.970	0.954	0.962	0.904	0.898	0.901

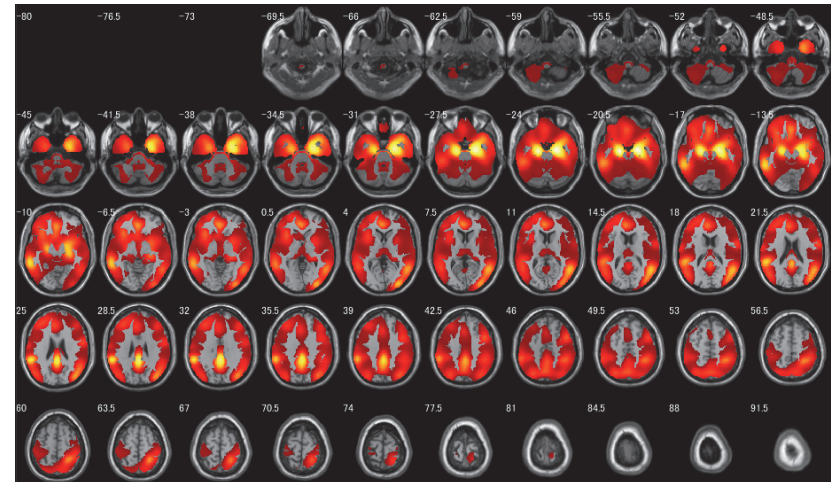
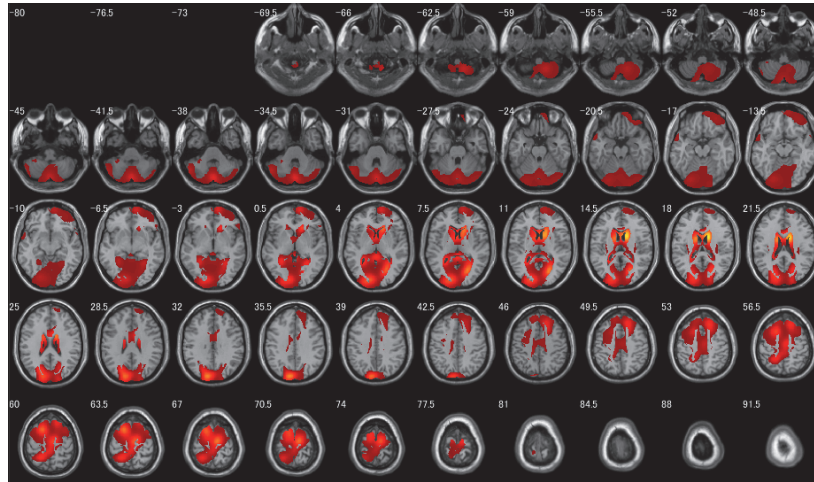
選択された脳領域

$$\mu_{XZ} = 0, \mu_{YZ} = 0.2$$

Positive

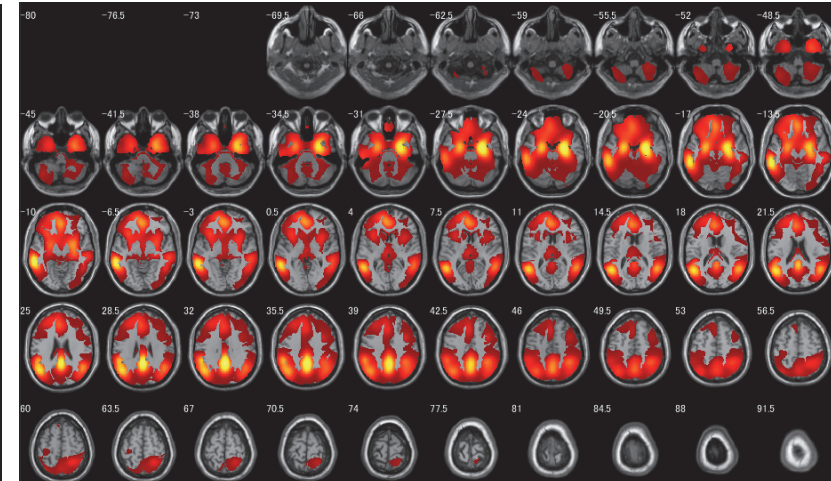
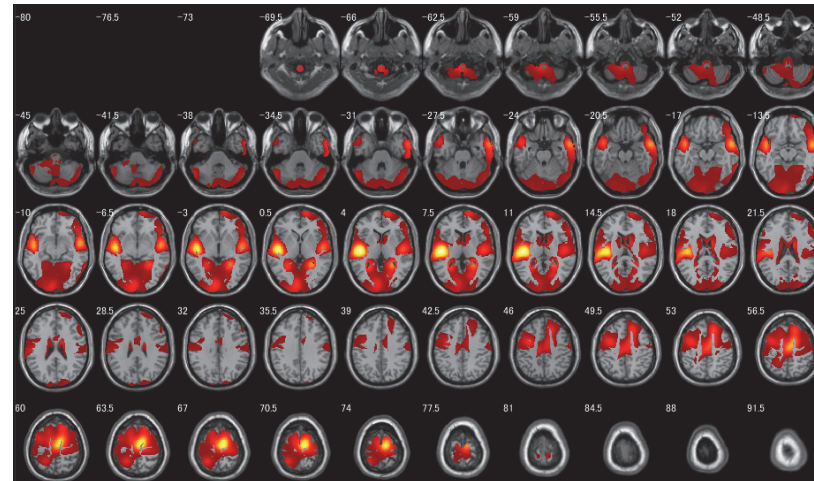
Negative

MRI

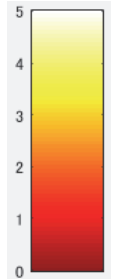


萎縮

PET

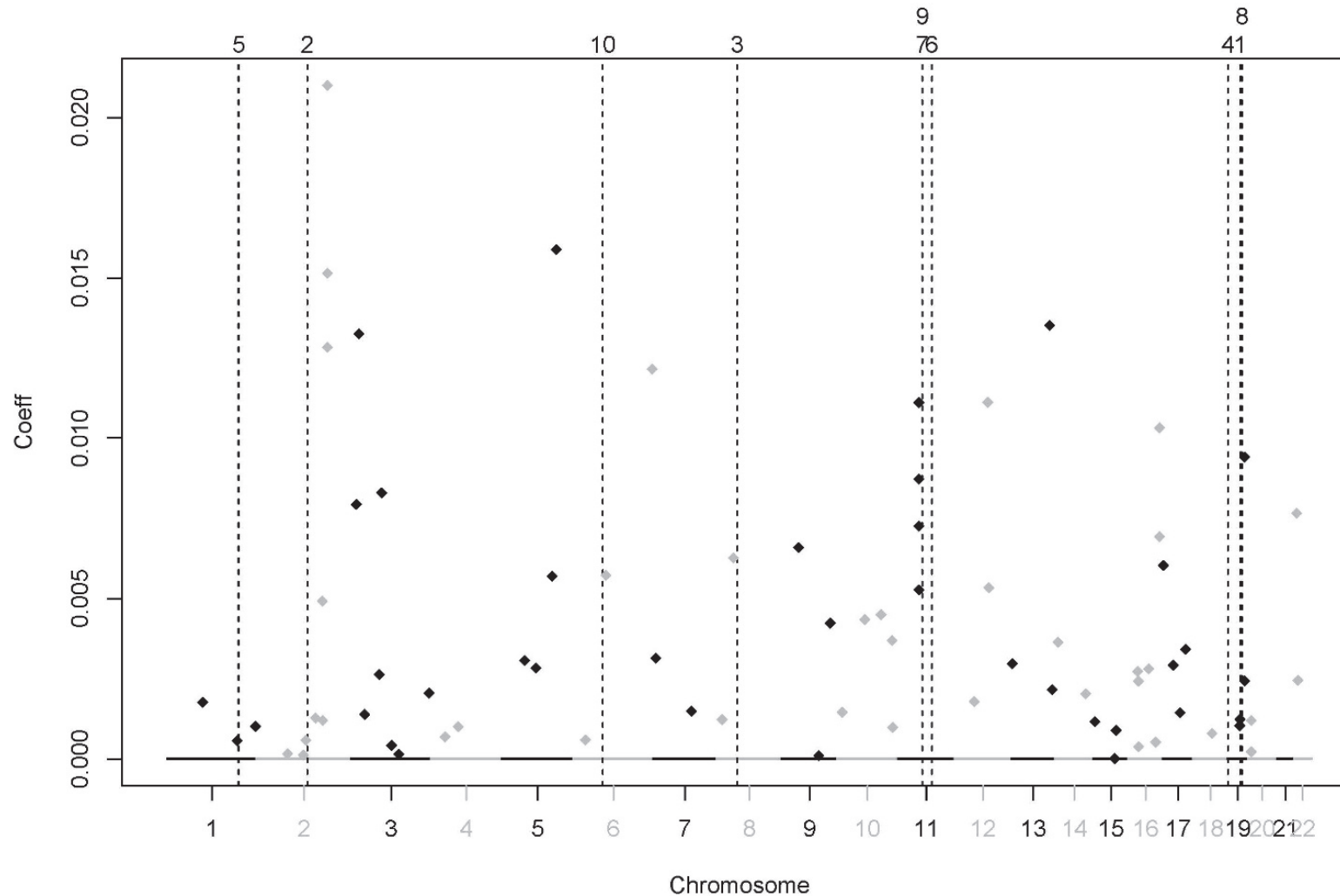


代謝低下

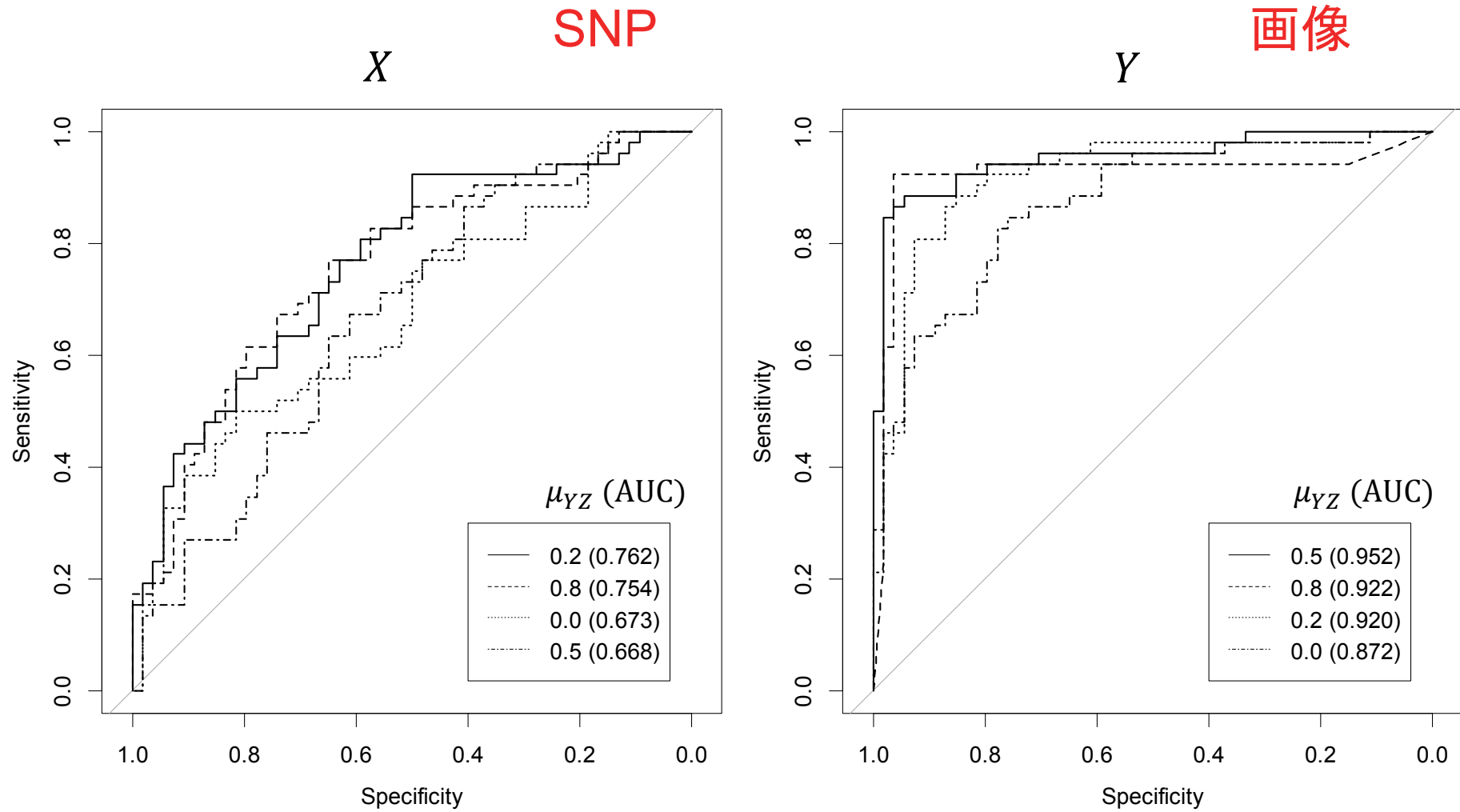


選択されたSNP

1="APOE", 2="BIN1", 3="CLU", 4="ABCA7", 5="CR1", 6="PICALM",
7="MS4A6A", 8="CD33", 9="MS4A4E", 10="CD2AP"



ROC解析



$$\mu_{XZ} = 0, 0 \leq \mu_{YZ} \leq 1$$

まとめ

- μ BIG data analysisに診断情報を組み込むための方法として、SMSMA法を提案した。
 - » Multiblock: 画像のマルチモダリティに対応
 - » Supervised: 結果の解釈を容易にする. より予測精度の高い予測モデル.
 - » Sparse (L1正則化法): 関連を表すパラメータ推定と同時に脳領域とSNP探索.
- SNP・sMRI・PET・診断の実データに適用し、実行可能性と有用性を確認した.
- 数値実験によって、方法の特徴を調べた。
 - » XとYのスコアの関連が弱いときに提案法が有効である事がわかった.
- 発展
 - » 適用モダリティの拡大, モダリティの選択
 - » ゲノム解析への応用, 脳腫瘍画像解析(Kawaguchi 2017), など

R package

msma: Multiblock Sparse Multivariable Analysis

There are several functions to implement the method for analysis in a multiblock multivariable data. If the input is only a matrix, then the principal components analysis (PCA) is implemented. If the input is a list of matrices, then the multiblock PCA is implemented. If the input is two matrices for exploratory and objective variables, then the partial least squares (PLS) analysis is implemented. If the input is two list of matrices for exploratory and objective variables, then the multiblock PLS analysis is implemented. Moreover, if the extra outcome variable is specified, then the supervised version for the methods above is implemented. For each methods, the sparse modeling is also incorporated. Functions to select the number of components and the regularized parameters are also provided.

Version: 0.7
Depends: [mvtnorm](#)
Published: 2016-01-01
Author: Atsushi Kawaguchi
Maintainer: Atsushi Kawaguchi <kawa_a24 at yahoo.co.jp>
License: [GPL-2](#) | [GPL-3](#) [expanded from: GPL (≥ 2)]
NeedsCompilation: no
CRAN checks: [msma results](#)

Reference

- Kawaguchi A, Yamashita F (2017). Supervised Multiblock Sparse Multivariable Analysis with Application to Multimodal Brain Imaging Genetics. *Biostatistics*, in press.
- Kawaguchi A (2017). Supervised Dimension Reduction Methods for Brain Tumor Image Data Analysis. In *Frontiers of Biostatistical Methods and Applications in Clinical Oncology*, Matsui S and Crowley J. (Ed.), Springer, in press.
- Kawaguchi A (2016). Diagnostic Probability Modeling for Longitudinal Structural Brain MRI Data Analysis. In *Statistical Techniques for Neuroscientists*, Truong KY. (Ed.), CRC Press, 361-374.
- Yoshida H., Kawaguchi A., Tsuruya K. (2013). Radial Basis Function-Sparse Partial Least Squares for Application to Brain Imaging Data. *Computational and Mathematical Methods in Medicine*, Volume 2013, Article ID 591032, 7 pages
- Araki, Y. Kawaguchi A., Yamashita, F. (2013). Regularized logistic discrimination with basis expansions for the early detection of Alzheimer's disease based on three-dimensional MRI data. *Advances in Data Analysis and Classification*, Volume 7, Issue 1, pp 109-119.
- Arbabshirani, M. R., Plis, S., Sui, J., & Calhoun, V. D. (2017). Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *NeuroImage*, 145, 137-165.
- Calhoun, V. D., & Sui, J. (2016). Multimodal fusion of brain imaging data: A key to finding the missing link (s) in complex mental illness. *Biological psychiatry: cognitive neuroscience and neuroimaging*, 1(3), 230-244.
- Vandenberghe, S., & Marsden, P. K. (2015). PET-MRI: a review of challenges and solutions in the development of integrated multimodality imaging. *Physics in medicine and biology*, 60(4), R115.
- Meng, X., Jiang, R., Lin, D., Bustillo, J., Jones, T., Chen, J., ... & Sui, J. (2017). Predicting individualized clinical measures by a generalized prediction framework and multimodal fusion of MRI data. *NeuroImage*, 145, 218-229.