A Feature Selection Study for Medical Image deep learning

Sanghyuck Lee

2021. 04. 20.

Outline

- Introduction
- False Discovery Rate
- Related works for controlling False Discovery Rate
- Knockoffs for controlling False Discovery Rate in Feature Selection
- Example

Introduction

- Last week's Feedback

  1) Consideration for Type II error

  2) The meaning of imitating the correlation structure

  3) Simple knockoff example

- Barber et al. (2015):

  Variable selection for linear model

  → Controlling False Discovery Rate among the selected variables

  → The knockoff filter

Introduction

"Unfortunately, however, it(A FDR control procedure, i.e. Benjamini & Hochberg, 1995 (BH method)) is inapplicable in variable selection, since the required positive dependence condition among the test statistics exhibited through $\Sigma = X'X$ is not generally satisfied, except in the very special case when the explanatory variables are uncorrelated. "

"These procedures not only do not rely on any distributional assumptions and any specific correlation structure among the explanatory variables, but also can consider any type of estimates of the regression coefficients, such as those obtained using Lasso"

(Sarkar et al. 2021)

False Discovery Rate

- False Discovery Rate (FDR) in statistics: a method of conceptualizing the rate of type I errors in null hypothesis testing when conducting multiple comparisons.
- Multiple comparisons using Analysis of Variance (ANOVA):
    1) "The average of all groups is the same" or Not
    2) If not, which groups are different?

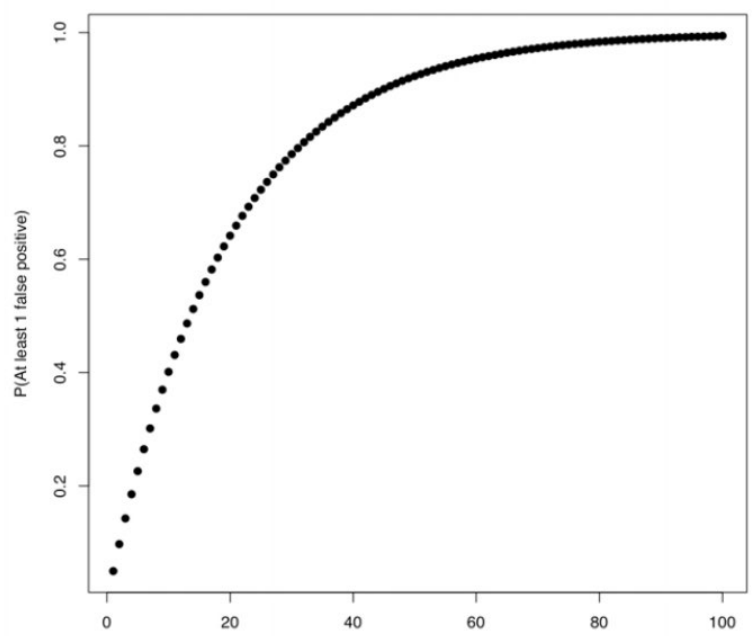
As the number of tests increases, the error increases

Family-wise error rate (FWER): $1-(1-\alpha)^m$

m is the number of tests, α is significance level

$$FDR = \frac{False\ Positive}{True\ Positive+False\ Positive}$$

P(Making at least 1 error in m tests) = 1 - (1 - α)^m



False Discovery Rate in Feature Selection

- rejected hypothesis -> selected feature -> likely belongs to the model

$$y = X\beta + z$$

y is a vector of responses, $X \in \mathbb{R}^{n\times p}$ is a known design matrix, $\beta \in \mathbb{R}^{p}$ is an unknown vector of coefficients and $z \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ is Gaussian noise.

- The FDR is the expected proportion of falsely selected variables, a false discovery being a selected variable not appearing in the true model

- $FDR = \dfrac{False\ Positive}{True\ Positive + False\ Positive} = \mathbb{E}\left[\dfrac{\#\{\mathrm{j}: \beta_{\mathrm{j}} = 0 \text{ and } \mathrm{j} \in \hat{S}\}}{\#\{\mathrm{j}: \mathrm{j} \in \hat{S}\} \vee 1}\right]$

- #: cardinality, ∨: max operator, $\hat{S}$: a set of selected features

Related works for controlling FDR

- Permutation method
- The Benjamini–Hochberg procedure and variants
- Barber & Candes (2015) method: Using knockoffs

  "The goal of the knockoff variables is to imitate the correlation structure of the original features in a very specific way that will allow for FDR control."

Knockoff for controlling FDR in FS

In Barber & Candes (2015)

$$\tilde{X}^T\tilde{X} = \Sigma \text{ and } \tilde{X}^T X = \Sigma - \mathrm{D}$$

where $D = \mathrm{diag}\{s\}$, with $s \in \mathbb{R}_{+}^{d}$ being such that $2\Sigma - D$ is positive definite.

$$\tilde{X} = X\Sigma^{-1}(\Sigma - D) + \widetilde{U}(2D - D\Sigma^{-1}D)^{\frac{1}{2}}$$

where $\widetilde{U}$ is orthonormal to the column space of X

Example

$$\tilde{X}^T\tilde{X} = \Sigma \text{ and } \tilde{X}^T X = \Sigma - \mathrm{D}$$

where $D = \mathrm{diag}\{s\}$, with $s \in \mathbb{R}_+^d$ being such that $2\Sigma - D$ is positive definite.

$$\tilde{X} = X\Sigma^{-1}(\Sigma - D) + \tilde{U}(2D - D\Sigma^{-1}D)^{\frac{1}{2}}$$

- Equi-correlated knockoffs:

$$s_j = 2\lambda_{min}(\Sigma) \wedge 1, \text{ where } \tilde{X}_j^T X_j = 1 - s_j$$

Example

- Original features $X \sim \mathcal{N}(0, \Sigma)$

i	X_1	X_2	X_3
1	1	-1	-1
2	-1	1	-1
3	1	1	-1
4	-1	1	0
5	0	-1	1
6	-1	0	1
7	1	-1	1
8	-1	-1	-1
9	1	1	-1
10	-1	1	1
11	1	-1	1

- Mean matrix μ

	X_1	X_2	X_3
μ	0	0	0

- Covariance matrix Σ

	X_1	X_2	X_3
X_1	1	-0.3	-0.1
X_2	-0.3	1	-0.3
X_3	-0.1	-0.3	1

- Covariance matrix and Correlation matrix are the same in this example.

Example

- Calculate $\Sigma^{-1} = \begin{vmatrix} 91/99 & -3/11 & 1/99 \\ 1/30 & 1 & 1/30 \\ 19/99 & 3/110 & 109/99 \end{vmatrix}$

using Gauss-Jordan elimination method, Cramer method or Cholesky decomposition method

Example

- Calculate $D = \text{diag}\{s\}$

where $s_j = 2\lambda_{min}(\Sigma) \wedge 1$, $\wedge$ is $\min$ operator and $\tilde{X}_j^T X_j = 1 - s_j$ (Equi-correlated knockoffs)

calculate minimum eigen value of $\Sigma = \begin{vmatrix} 1 & -0.3 & -0.1 \\ -0.3 & 1 & -0.3 \\ -0.1 & -0.3 & 1 \end{vmatrix}$

$$eigen\ values = \ 0.52, 1.1, 1.38$$

$$minimum\ eigen\ value\ \ = \ 0.52$$

$\therefore\ s_j = 1$ and D= $\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}$

Example

- Calculate $\widetilde{U}$ that is orthonormal to the column space of X
- $U^T X = 0$

$$A = \begin{vmatrix} 1 & -1 & -1 & 0 & 0 & 0 \\ -1 & 1 & -1 & 0 & 0 & 0 \\ 1 & 1 & -1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & 0 \\ -1 & -1 & -1 & 0 & 0 & 0 \\ 1 & 1 & -1 & 0 & 0 & 0 \\ -1 & 1 & 1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & 0 \end{vmatrix} = QR, \text{ where Q is } \mathrm{n} \times n \text{ matrix and } R \text{ is } n \times 2p \text{ matrix.}$$

Use $p\sim 2p$ row of Q as U

Example

- $\tilde{X} = X\Sigma^{-1}(\Sigma - D) + \tilde{U}(2D - D\Sigma^{-1}D)^{\frac{1}{2}}$

$(n \times p) =$

$$(n \times p)(n \times n)\{(n \times n) - (n \times n)\} + (n \times p)\{2(n \times n) - (n \times n)(n \times n)(n \times n)\}^{\frac{1}{2}}$$

- By substituting the previously calculated Σ^{-1}, D, and $\tilde{U}$ into the formula,

  knockoff matrix $\tilde{X}$ can be calculated

References

Barber, Rina Foygel, and Emmanuel J. Candès. "Controlling the false discovery rate via knockoffs." *Annals of Statistics* 43.5 (2015): 2055-2085.

Benjamini, Yoav, and Yosef Hochberg. "Controlling the false discovery rate: a practical and powerful approach to multiple testing." *Journal of the Royal statistical society: series B (Methodological)* 57.1 (1995): 289-300.

Sarkar, Sanat K., and Cheng Yong Tang. "Adjusting the Benjamini-Hochberg method for controlling the false discovery rate in knockoff assisted variable selection." *arXiv preprint arXiv:2102.09080* (2021).