A Feature Selection Study for Medical Image deep learning

Sanghyuck Lee

2021. 05 11.

Outline

- Knockoff study

- Thyroid Research Project Report

Knockoff study

$$[X\tilde{X}]^T[X\tilde{X}] = \begin{vmatrix} \Sigma & \Sigma - S \\ \Sigma - S & \Sigma \end{vmatrix} \leftrightarrow \quad \tilde{X} = X(\mathrm{I} - \Sigma^{-1}\mathrm{S}) + \tilde{U}C$$

where $\tilde{U}$ is orthonormal to the column space of X, $C \in R^{(n\text{-}p)\times p}$ satisfies $\mathrm{C}^{\mathrm{T}}C = 2S - S\,\Sigma^{-1}S \geq 0$, S is $\mathrm{diag}\{s\}$ and s=$(s_1, \ldots\ s_p)$ has nonnegative entries.

Knockoff study

1)

$$[X\tilde{X}]^T[X\tilde{X}] = \begin{vmatrix} \Sigma & \Sigma - S \\ \Sigma - S & \Sigma \end{vmatrix} \rightarrow \tilde{X} = X(\mathrm{I} - \Sigma^{-1}\mathrm{S}) + \tilde{U}C$$

where $\tilde{U}$ is orthonormal to the column space of X, $C \in R^{(n\text{-}p)\times p}$ satisfies $\mathrm{C}^{\mathrm{T}}C = 2S - S\,\Sigma^{-1}S \geq 0$, S is $\mathrm{diag}\{s\}$ and s=$(s_1, \ldots\, s_p)$ has nonnegative entries.

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

$$\tilde{X} = \mathrm{A} + \mathrm{X}$$

$$A^T X = (\tilde{X} - \mathrm{X})^T X = \Sigma - S - \Sigma = -\mathrm{S} \qquad \Rightarrow A^T X = -\mathrm{S}$$

$$\Sigma = \tilde{X}^{\mathrm{T}}\tilde{X} = (A + X)^T(A + X) = A^T A + \Sigma - 2S \qquad \Rightarrow A^T A = 2S$$

Choose an orthonormal matrix $U \in R^{n\times(n-p)}$, then A can be expressed as

$$A = XB + UC \text{ with } B \in R^{p\times p},\ C \in R^{(n-p)\times p}$$

$$(n \times p) = (n \times p) \times (p \times p) + (n \times (n - p)) \times ((n - p) \times p)$$

$$U^T X = X^T U = 0 \text{ and } U^T U = I$$

Knockoff study

1)

$$[X\tilde{X}]^T[X\tilde{X}] = \begin{vmatrix} \Sigma & \Sigma - S \\ \Sigma - S & \Sigma \end{vmatrix} \quad \rightarrow \quad \tilde{X} = X(\mathrm{I} - \Sigma^{-1}\mathrm{S}) + \tilde{U}C$$

where $\tilde{U}$ is orthonormal to the column space of X, $C \in R^{(n-p)\times p}$ satisfies $\mathrm{C}^{\mathrm{T}}C = 2S - S\,\Sigma^{-1}S \geq 0$, S is diag$\{s\}$ and s=$(s_1, \ldots\, s_p)$ has nonnegative entries.

- $\tilde{X} = \mathrm{A} + \mathrm{X}$
- $A^T X = -\mathrm{S}$
- $A^T A = 2S$
- $A = XB + UC$
- $U^T X = X^T U = 0$ and $U^T U = I$

$$-\mathrm{S} = A^T X = X^T A = X^T(XB + UC) = \Sigma B + 0 \quad \Rightarrow \quad B = -\Sigma^{-1}S$$

$$2S = A^T A = (B^T X^T + C^T U^T)(XB + UC) = B^T \Sigma B + C^T C$$

$$= S\Sigma^{-1}S + C^T C \qquad \Rightarrow C^T C = 2S - S\Sigma^{-1}S$$

$$\tilde{X} = \mathrm{A} + \mathrm{X} = (XB + UC) + X = X - X\Sigma^{-1}S + UC = X(\mathrm{I} - \Sigma^{-1}\mathrm{S}) + \tilde{U}C$$

Knockoff study

2)

$$[X\tilde{X}]^T[X\tilde{X}] = \begin{vmatrix} \Sigma & \Sigma - S \\ \Sigma - S & \Sigma \end{vmatrix} \quad \leftarrow \quad \tilde{X} = X(\mathrm{I} - \Sigma^{-1}\mathrm{S}) + \tilde{U}C$$

where $\tilde{U}$ is orthonormal to the column space of X, $C \in R^{(n-p)\times p}$ satisfies $\mathrm{C}^{\mathrm{T}}C = 2S - S\,\Sigma^{-1}S \geq 0$, S is diag$\{s\}$ and s=$(s_1, \ldots\ s_p)$ has nonnegative entries.

- $U^TX = X^TU = 0$ and $U^TU = I$

$$X^T\tilde{X} = X^T\left(X(\mathrm{I} - \Sigma^{-1}\mathrm{S}) + \tilde{U}C\right) = \Sigma(I - \Sigma^{-1}\mathrm{S}) + 0 = \Sigma - S$$

$$\tilde{X}^T\tilde{X} = \left((\mathrm{I} - \mathrm{S}\Sigma^{-1})X^T + C^T\tilde{U}^T\right)\left(X(\mathrm{I} - \Sigma^{-1}\mathrm{S}) + \tilde{U}C\right)$$

(1) (2) (3) (4)

(1)×(3)= $(\mathrm{I} - \mathrm{S}\Sigma^{-1})X^TX(\mathrm{I} - \Sigma^{-1}\mathrm{S}) = (\Sigma - S)(\mathrm{I} - \Sigma^{-1}\mathrm{S}) = \Sigma - 2S + S\Sigma^{-1}\mathrm{S}$

(1)×(4)=(2)×(3)= $0 \quad \because U^TX = X^TU = 0$

(2)×(4)= $C^T\tilde{U}^T \times \tilde{U}C = \mathrm{C}^{\mathrm{T}}C = 2S - S\,\Sigma^{-1}S \qquad \because U^TU = I$

Knockoff study

2)

$$[X\tilde{X}]^T[X\tilde{X}] = \begin{vmatrix} \Sigma & \Sigma - S \\ \Sigma - S & \Sigma \end{vmatrix} \quad \leftarrow \quad \tilde{X} = X(\mathrm{I} - \Sigma^{-1}\mathrm{S}) + \tilde{U}C$$

where $\tilde{U}$ is orthonormal to the column space of X, $C \in R^{(n-p)\times p}$ satisfies $\mathrm{C}^{\mathrm{T}}C = 2S - S\,\Sigma^{-1}S \geq 0$, S is $\mathrm{diag}\{s\}$ and s=$(s_1, \ldots\ s_p)$ has nonnegative entries.

- $U^TX = X^TU = 0$ and $U^TU = I$

$$X^T\tilde{X} = \Sigma - 2S + S\Sigma^{-1}\mathrm{S} + 2S - S\,\Sigma^{-1}S$$
$$= \Sigma$$

$$\Rightarrow\ X^T\tilde{X} = \Sigma - S \text{ and } X^T\tilde{X} = \Sigma$$

$$\therefore \quad [X\tilde{X}]^T[X\tilde{X}] = \begin{vmatrix} \Sigma & \Sigma - S \\ \Sigma - S & \Sigma \end{vmatrix} \quad \leftrightarrow \quad \tilde{X} = X(\mathrm{I} - \Sigma^{-1}\mathrm{S}) + \tilde{U}C$$

갑상선 연구 과제 보고

- 이번주 완료 계획으로 말씀드렸던 내용 및 해당 결과:

  CO2와 CO3에 대한 세그멘테이션 150장 완료

  => CO2에 대해 완료, CO3에 대해 일부 완료
- 시행착오 및 향후 계획 판단 근거

  =>CO3에 대해, 기존 방식으로 세그멘테이션이 어려움을 확인, 새로운 방식으로 진행할 예정이나, 1273 환자에 대해 모두 수동 세그멘테이션 진행 계획

  => 따라서, 시간이 비교적 오래 소요될 것으로 예상되며, 또한, 예측 모델상에서 CO2가 비교적 더 중요한 역할을 수행할 수 있겠다고 판단
- 향후 계획

  => CO3 보다는 AX를 먼저 진행, 자동화 모델(CO2와 AX 및 지방) 구축이 가능해 보이는 케이스를 먼저 진행하고, 자동 세그멘테이션과 남은 수동 세그멘테이션(CO3)을 병렬적으로 진행

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.

Chen, Jiajie, Anthony Hou, and Thomas Y. Hou. "Some analysis of the knockoff filter and its variants." *arXiv preprint arXiv:1706.03400* (2017).

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).