A Feature Selection Study for Medical Image deep learning

Sanghyuck Lee

2021. 04. 07.

Outline

- Overall Framework of Feature Selection using Knockoffs
- Knockoff Types
- Model-X Knockoffs
- Normal Distribution Knockoffs
- A Simple Example

Overall Framework of FS using Knockoffs

- Given a set p features $X = (X_1, \dots, X_p)$ and an outcome of interest y, knockoffs aim to select the small fraction of features on which y actually depends while controlling the False Discovery Rate (FDR).

1. Construct synthetic variables $\tilde{X} = (\tilde{X}_1, \dots, \tilde{X}_p)$ called knockoffs.
2. Use an arbitrary machine learning algorithm – usually called a features statistic – to assign variable importances to each of the p features and each of the p knockoffs. (e.g. train cross-validated Lasso on $[X, \tilde{X}]$ and y)
3. A non-null feature should be assigned a higher variable importance than its knockoff, whereas knockoffs are constructed such that null features are indistinguishable from their knockoff.

2 main types of knockoffs

1) **Fixed-X knockoffs** (Barber et al., 2015)

- treat the design matrix X as fixed.
- assuming $y|X$ follows a homoscedastic gaussian linear response.
- it is possible to construct valid knockoffs $\tilde{X}$ with NO assumptions on X.

2) **Model-X knockoffs** (Candes et al.,2018)

- treat the design matrix X as random.
- for any conditional distribution $y|X$, but they assume that the distribution of X is known.
- one must know (or estimate) the distribution of X.

Model-X knockoffs

Random variables with Two properties

Original Features: $\mathrm{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_p)$

Knockoff Features: $\tilde{\mathrm{X}} = (\tilde{\mathrm{X}}_1, \ldots, \tilde{\mathrm{X}}_\mathrm{p})$

(1) $(\mathrm{X}, \tilde{\mathrm{X}})_{\mathrm{swap(S)}} \overset{d}{=} (\mathrm{X}, \tilde{\mathrm{X}})$ for any subset $S \subset \{1, \ldots, p\}$

e.g. $(\mathrm{X}_1,\ X_2,\ \mathrm{X}_3,\ \tilde{\mathrm{X}}_1,\ \tilde{\mathrm{X}}_2,\ \tilde{X}_3)_{\mathrm{swap(S)}} \overset{d}{=} (X_1,\ \tilde{X}_2,\ \tilde{X}_3,\ \tilde{X}_1, X_2,\ X_3)$

(2) $\tilde{\mathrm{X}} \perp\!\!\!\perp \mathrm{Y}|\mathrm{X}$, i.e. $\tilde{\mathrm{X}}$ is independent of response Y given feature X

Normal Distribution Knockoffs

- $X \sim \mathcal{N}(\mu, \Sigma)$
- $(X, \tilde{X}) \sim \mathcal{N}(\mu, \mathbf{G})$, where $\mathbf{G} = \begin{bmatrix} \Sigma & \Sigma - \mathrm{diag}\{s\} \\ \Sigma - \mathrm{diag}\{s\} & \Sigma \end{bmatrix}$
- here, $\mathrm{diag}\{s\}$ is any diagonal matrix selected in such a way that the joint covariance matrix G is positive semidefinite.
- $\tilde{X} | X \overset{\mathrm{d}}{=} \mathcal{N}(\tilde{\mu}, \tilde{\Sigma})$

  where $\tilde{\mu} = X - (X - \mu)\Sigma^{-1}\mathrm{diag}\{s\}$,

  $$\tilde{\Sigma} = 2\mathrm{diag}\{s\} - \mathrm{diag}\{s\}\Sigma^{-1} - \mathrm{diag}\{s\}.$$
- $\tilde{\mu}$ and $\tilde{\Sigma}$ are given by classical regression formulas.

Example

- Original features X

n	X_1	X_2	X_3	X_4
1	0.87	0.49	0.18	0.28
2	0.37	0.21	0.07	0.56
3	0.55	0.52	0.37	0.29
4	0.19	0.26	0.93	0.21
5	0.34	0.15	0.11	0.54

- Mean matrix μ

	X_1	X_2	X_3	X_4
μ	0.464	0.326	0.332	0.376

- Covariance matrix Σ

	X_1	X_2	X_3	X_4
X_1	0.0679	0.0335	-0.0425	-0.0096
X_2	0.0335	0.0283	0.0031	-0.0179
X_3	-0.0425	0.0031	0.125	-0.0431
X_4	-0.0096	-0.0179	-0.0431	0.0262

Example

- $\tilde{X}|X \stackrel{d}{=} \mathcal{N}(\tilde{\mu}, \tilde{\Sigma})$

  where $\tilde{\mu} = X - (X-\mu)\Sigma^{-1}\mathrm{diag}\{s\}$, $\tilde{\Sigma} = 2\mathrm{diag}\{s\} - \mathrm{diag}\{s\}\Sigma^{-1} - \mathrm{diag}\{s\}$.
- Calculate Σ^{-1} using Cholesky decomposition

  $\Sigma = LL^T$ where L is a lower triangular matrix

  $\Sigma^{-1} = (L^{-1})^T L^{-1}$

$$\Sigma = \begin{vmatrix} 0.0679 & 0.0335 & -0.0425 & -0.0096 \\ 0.0335 & 0.0283 & 0.0031 & -0.0179 \\ -0.0425 & 0.0031 & 0.125 & -0.0431 \\ -0.0096 & -0.0179 & -0.0431 & 0.0262 \end{vmatrix}$$

$$= \begin{vmatrix} 0.2605 & 0 & 0 & 0 \\ 0.1287 & 0.1085 & 0 & 0 \\ -0.1633 & 0.2222 & 0.2213 & 0 \\ -0.037 & -0.1213 & -0.1004 & 0.0084 \end{vmatrix} \begin{vmatrix} 0.2605 & 0.1287 & -0.1633 & -0.037 \\ 0 & 0.1085 & 0.2222 & -0.1213 \\ 0 & 0 & 0.2213 & -0.1004 \\ 0 & 0 & 0 & 0.0084 \end{vmatrix}$$

Example

$\Sigma^{-1} =$

$$\begin{vmatrix} 3.8382 & -4.5503 & 7.399 & 39.4163 \\ 0 & 9.2146 & -9.2504 & 22.415 \\ 0 & 0 & 4.5179 & 53.7404 \\ 0 & 0 & 0 & 118.4925 \end{vmatrix} \begin{vmatrix} 3.8382 & 0 & 0 & 0 \\ -4.5503 & 9.2146 & 0 & 0 \\ 7.399 & -9.2504 & 4.5179 & 0 \\ 39.4163 & 22.415 & 53.7404 & 118.4925 \end{vmatrix}$$

$$= \begin{vmatrix} 1643.8301 & 773.144 & 2151.6791 & 4670.5399 \\ 773.144 & 672.9094 & 1162.7979 & 2656.0051 \\ 2151.6791 & 1162.7979 & 2908.445 & 6367.8355 \\ 4670.5399 & 2656.0051 & 6367.8355 & 14040.4627 \end{vmatrix}$$

Example

- Calculate $\mathrm{diag}\{s\}$ using a specific method, such as:
  - Semidefinite Program (SDP)
  - Minimum Variance-Based Reconstructability (MVR)
  - Minimizes the mutual information (MMI)
  - Maximize Entropy(MAXENT)
- $\mathrm{diag}\{s\} = \begin{vmatrix} 0.0002 & 0 & 0 & 0 \\ 0 & 0.0002 & 0 & 0 \\ 0 & 0 & 0.0002 & 0 \\ 0 & 0 & 0 & 0 \end{vmatrix}$ by SDP using kncokpy

Example

- Σ^{-1} and $\mathrm{diag}\{s\}$ were assigned, and the distribution of knockoff was determined.
- $\tilde{X}|X \overset{\mathrm{d}}{=} \mathcal{N}(\tilde{\mu},\ \tilde{\Sigma})$,

$$\text{where } \tilde{\mu} = X - (X-\mu)\Sigma^{-1}\,\mathrm{diag}\{s\} = \begin{vmatrix} 0.867 & 0.4914 & 0.1782 & 0.2795 \\ 0.3617 & 0.2038 & 0.0583 & 0.5544 \\ 0.5546 & 0.5176 & 0.3743 & 0.2916 \\ 0.1884 & 0.2603 & 0.9274 & 0.2093 \\ 0.3483 & 0.1569 & 0.1217 & 0.5453 \end{vmatrix}$$

$$\tilde{\Sigma} = 2\mathrm{diag}\{s\} - \mathrm{diag}\{s\}\Sigma^{-1} - \mathrm{diag}\{s\} = \begin{vmatrix} 0.0003 & 0 & -0.0001 & 0 \\ 0 & 0.0004 & 0 & 0 \\ -0.0001 & 0 & 0.0002 & 0 \\ 0 & 0 & 0 & 0.0001 \end{vmatrix}$$

Further Study

- Study some parts that are not calculated by hand
- Knock-off verification
- Understanding of mathematical expression

References

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

Candes, Emmanuel, et al. "Panning for gold: Model-X knockoffs for high-dimensional controlled variable selection." *arXiv preprint arXiv:1610.02351* (2016).

Lu, Yang Young, et al. "DeepPINK: reproducible feature selection in deep neural networks." *arXiv preprint arXiv:1809.01185* (2018).