Group Knockoffs SDP Derivation

Yibin Xiong

July 17, 2022

Construct Knockoffs by Solving an SDP

▷ Recall from [1], we want to make the correlation $\langle X_j, \tilde{X}_j \rangle = 1 - s_j$ for any j as small as possible to increase the power.

▷ This motivates the following SDP problem

$$\begin{aligned} \min_{s_j} \sum_{j=1}^{n} |1 - s_j| \\ \text{s.t. } s_j \geq 0 \ \ \forall j \\ \operatorname{diag}\{s\} \preceq 2\Sigma \end{aligned}$$

where the constraints are equivalent to the condition that the matrix $G := \begin{bmatrix} X & \tilde{X} \end{bmatrix}^T \begin{bmatrix} X & \tilde{X} \end{bmatrix} \succeq 0$. ($\frac{1}{n}G$ is the correlation matrix if X has been normalized).

Solve the SDP

▷ Note that at the optimum $s_j \leq 1$

This is because if $s_j > 1$ is a feasible solution, then 1 is also feasible and attains a lower value of the objective.

▷ Consider the equivalent SDP

$$\begin{aligned} \min_{s_j} \sum_{j=1}^{n} 1 - s_j \\ \text{s.t. } 0 \leq s_j \leq 1 \;\; \forall j \\ \text{diag}\{s\} \preceq 2\Sigma \end{aligned} \iff \begin{aligned} \max_{s_j} \sum_{j=1}^{n} s_j \\ \text{s.t. } 0 \leq s_j \leq 1 \;\; \forall j \\ \text{diag}\{s\} \preceq 2\Sigma \end{aligned}$$

▷ Solve it with convex optimization packages like `cvxpy`.

Example Code

```python
import cvxpy as cvx

# corrMatrix is the correlation matrix
p,_ = corrMatrix.shape
s = cvx.Variable(p)
# set up the objective and the constraints
objective = cvx.Maximize(sum(s))
constraints = [ 2.0*corrMatrix >> cvx.diag(s) + cvx.diag([
    tol]*p), 0<=s, s<=1]
prob = cvx.Problem(objective, constraints)
# get the solution
prob.solve(solver='CVXOPT')
```

Equi-correlated Case: A Close-form Solution

▷ If we further assume that $s_1 = \cdots = s_n := \tilde{s}$, then the SDP reduces to

$$\begin{aligned} &\min_{\tilde{s}} 1 - \tilde{s} \\ &\text{s.t.}\, 0 \leq \tilde{s} \leq 1,\ \tilde{s}\mathbf{I} \preceq 2\Sigma \end{aligned}$$

▷ By **Rayleigh inequality**, the largest $\tilde{s}$ that satisfies $\tilde{s}I \preceq 2\Sigma$ is $2\lambda_{\min}(\Sigma)$.

▷ Thus, the optimal solution is $\min\{1, 2\lambda_{\min}(\Sigma)\}$.

Group Structure: A Generalization of diag{s}

▷ Now we assume

$$\begin{bmatrix} X & \tilde{X} \end{bmatrix}^T \begin{bmatrix} X & \tilde{X} \end{bmatrix} := G = \begin{bmatrix} \Sigma & \Sigma - S \\ \Sigma - S & \Sigma \end{bmatrix}$$

where $S := \text{diag}\{S_1, \ldots, S_m\}$ is a **block-diagonal** matrix. $S_1, \ldots, S_m$ are square matrices that corresponds to the groups. [2]

∘ This means that we can have $\tilde{X}_{j-1}^T X_j \neq X_{j-1}^T X_j = \Sigma_{j-1,j}$ if the variables $j-1, j$ are in a group.

∘ Intuitively, we want that not only $\tilde{X}_j$ and X_j are not correlated, but also $\tilde{X}_j$ and X_{j-1} are not correlated given that X_j and X_{j-1} are highly correlated.

Group Structure: A Generalization of diag{s}

▷ The original SDP can be adapted to

$$
\begin{aligned}
&\min_{s_j} \sum_{j=1}^{n} |1 - s_j| \\
&\text{s.t. } s_j \geq 0 \ \forall j, \\
&\quad \operatorname{diag}\{s\} \preceq 2\Sigma
\end{aligned}
\iff
\begin{aligned}
&\min_{\operatorname{diag}\{s\}} \sum_{j=1}^{n} |\Sigma_{j,j} - s_j| \\
&\text{s.t. } \operatorname{diag}\{s\} \succeq 0, \\
&\quad \operatorname{diag}\{s\} \preceq 2\Sigma
\end{aligned}
\implies
\begin{aligned}
&\min_{S} \sum_{j=1}^{m} \|\Sigma_{G_j,G_j} - S_j\|_F \\
&\text{s.t. } S \succeq 0, \\
&\quad S \preceq 2\Sigma
\end{aligned}
$$

where G_j denotes the set of indices in the j-th group.

Group Structure: A Generalization of diag$\{s\}$

$\triangleright$ The original SDP can be adapted to

$$\begin{aligned} &\min_{s_j} \sum_{j=1}^{n} |1 - s_j| \\ &\text{s.t. } s_j \geq 0 \ \forall j, \\ &\quad \text{diag}\{s\} \preceq 2\Sigma \end{aligned} \iff \begin{aligned} &\min_{\text{diag}\{s\}} \sum_{j=1}^{n} |\Sigma_{j,j} - s_j| \\ &\text{s.t. } \text{diag}\{s\} \succeq 0, \\ &\quad \text{diag}\{s\} \preceq 2\Sigma \end{aligned} \implies \begin{aligned} &\min_{S} \sum_{j=1}^{m} \|\Sigma_{G_j,G_j} - S_j\|_F \\ &\text{s.t. } S \succeq 0, \\ &\quad S \preceq 2\Sigma \end{aligned}$$

where G_j denotes the set of indices in the j-th group.

$\triangleright$ Assume that $S_j := \gamma_j \cdot \Sigma_{G_j,G_j}$. This just means we want some correlations between variables and knockoffs to be a fraction of the correlations between variables. In particular, s_j is a fraction of $\Sigma_{j,j} = 1$.

$$\min_{\gamma_j} \sum_{j=1}^{m} |1 - \gamma_j| \, \|\Sigma_{G_j,G_j}\|_F$$

$$\text{s.t. } \gamma_j \geq 0 \ \ \forall j, \ S \preceq 2\Sigma$$

Solve the SDP: General Case

- ▷ Argue that at optimum $\gamma_j \leq 1 \ \forall j$ for the same reason as before.
- ▷ Use `cvxpy` (code to be added later)

Solve the SDP: Equi-correlated Case Close-form

▷ In the last constraint,

$$S = \begin{bmatrix} \gamma_1 \cdot I_{G_1,G_1} & 0 & \dots & 0 \\ 0 & \gamma_2 \cdot I_{G_2,G_2} & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & \gamma_m I_{G_m,G_m} \end{bmatrix} \cdot \begin{bmatrix} \Sigma_{G_1,G_1} & 0 & \dots & 0 \\ 0 & \Sigma_{G_2,G_2} & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & \Sigma_{G_m,G_m} \end{bmatrix}$$

$$:= D^{-1} R D^{-1}$$

where $D^{-1} := \text{diag}\{\Sigma^{1/2}_{G_1,G_1}, \dots, \Sigma^{1/2}_{G_m,G_m}\}$ and $R := \text{diag}\{\gamma_1 \cdot I_{G_1,G_1}, \dots, \gamma_m \cdot I_{G_m,G_m}\}$

▷ Then the last constraint is equivalent to $R \preceq 2D\Sigma D$, where $D := \text{diag}\{\Sigma^{-1/2}_{G_1,G_1}, \dots, \Sigma^{-1/2}_{G_m,G_m}\}$.

Solve the SDP: Equi-correlated Case Close-form

▷ Now the SDP simplifies to

$$\begin{aligned} \min_{\gamma_j} \sum_{j=1}^{m} (1-\gamma_j) \, \|\Sigma_{G_j,G_j}\|_F \\ \text{s.t.}\, 0 \le \gamma_j \le 1 \;\; \forall j, \\ R \preceq 2D\Sigma D \end{aligned}$$

▷ In the equi-correlated case, we have $\gamma_1 = \cdots = \gamma_m := \tilde{\gamma}$. The SDP reduces to

$$\begin{aligned} \min_{\tilde{\gamma}} \left(\sum_{j=1}^{m} \|\Sigma_{G_j,G_j}\|_F \right) (1-\tilde{\gamma}) \\ \text{s.t.}\, 0 \le \tilde{\gamma} \le 1, \;\; \tilde{\gamma} I \preceq 2D\Sigma D \end{aligned}$$

▷ By Rayleigh inequality, the largest $\tilde{\gamma}$ that satisfies the last constraint is $2\lambda_{\min}(D\Sigma D)$. Thus, the optimal solution is $\min\{1, 2\lambda_{\min}(D\Sigma D)\}$.

References

[1] Rina Foygel Barber and Emmanuel J Candès. "Controlling the false discovery rate via knockoffs". In: *The Annals of Statistics* 43.5 (2015), pp. 2055–2085.

[2] Ran Dai and Rina Barber. "The knockoff filter for FDR control in group-sparse and multitask regression". In: *International conference on machine learning*. PMLR. 2016, pp. 1851–1859.