DeepDRK: Deep Dependency Regularized Knockoff for Feature Selection

Hongyu Shen[1]; Yici Yan[2]; Zhizhen Zhao[1]

[1]Department of Electrical and Computer Engineering, UIUC, IL, USA
[2]Department of Statistics, UIUC, IL, USA

Introduction

- **Goal:** Select the features associated with the linear response Y, given the covariate design matrix X, with controlled false discovery rate (FDR) under the Model-X knockoff framework[1]

[1] Candés et al., "Model-X knockoffs for high dimensional controlled variable selection," *J. R. Stat. Soc. Ser. B*, 2018.
[2] Romano et al., "Deep knockoffs," *J. Amer. Stat. Assoc.*, 2020.
[3] Jordon et al., "KnockoffGAN: Generating knockoffs for feature selection using GANs," *ICLR*, 2018.
[4] Masud et al., "Multivariate soft rank via entropy-regularized optimal transport: sample efficiency and generative modeling," *JMLR*, 2023.
[5] Sudarshan et al., "Deep direct likelihood knockoffs," *NeurIPS*, 2020.

Introduction

- **Goal:** Select the features associated with the linear response Y, given the covariate design matrix X, with controlled false discovery rate (FDR) under the Model-X knockoff framework[1]

- **Challenges:** Unknown data distribution and small sample size

[1] Candés et al., "Model-X knockoffs for high dimensional controlled variable selection," *J. R. Stat. Soc. Ser. B*, 2018.
[2] Romano et al., "Deep knockoffs," *J. Amer. Stat. Assoc.*, 2020.
[3] Jordon et al., "KnockoffGAN: Generating knockoffs for feature selection using GANs," *ICLR*, 2018.
[4] Masud et al., "Multivariate soft rank via entropy-regularized optimal transport: sample efficiency and generative modeling," *JMLR*, 2023.
[5] Sudarshan et al., "Deep direct likelihood knockoffs," *NeurIPS*, 2020.

Introduction

- **Goal:** Select the features associated with the linear response Y, given the covariate design matrix X, with controlled false discovery rate (FDR) under the Model-X knockoff framework[1]
- **Challenges:** Unknown data distribution and small sample size
- **Approach:** Deep generative models have been used for knockoff generations for non-Gaussian data
  - Deep Knockoff[2], KnockoffGAN[3], sRMMD[4], and DDLK[5]
  - Performance declines as the sample size decreases and the data distributions become more complex.

---

[1] Candés et al., "Model-X knockoffs for high dimensional controlled variable selection," *J. R. Stat. Soc. Ser. B*, 2018.
[2] Romano et al., "Deep knockoffs," *J. Amer. Stat. Assoc.*, 2020.
[3] Jordon et al., "KnockoffGAN: Generating knockoffs for feature selection using GANs," *ICLR*, 2018.
[4] Masud et al., "Multivariate soft rank via entropy-regularized optimal transport: sample efficiency and generative modeling," *JMLR*, 2023.
[5] Sudarshan et al., "Deep direct likelihood knockoffs," *NeurIPS*, 2020.

Introduction

- **Goal:** Select the features associated with the linear response Y, given the covariate design matrix X, with controlled false discovery rate (FDR) under the Model-X knockoff framework[1]
- **Challenges:** Unknown data distribution and small sample size
- **Approach:** Deep generative models have been used for knockoff generations for non-Gaussian data
  - Deep Knockoff[2], KnockoffGAN[3], sRMMD[4], and DDLK[5]
  - Performance declines as the sample size decreases and the data distributions become more complex.
- **Our approach:** DeepDRK generates knockoffs with a novel transformer-based generator and a random perturbation technique

---

[1] Candés et al., "Model-X knockoffs for high dimensional controlled variable selection," *J. R. Stat. Soc. Ser. B*, 2018.
[2] Romano et al., "Deep knockoffs," *J. Amer. Stat. Assoc.*, 2020.
[3] Jordon et al., "KnockoffGAN: Generating knockoffs for feature selection using GANs," *ICLR*, 2018.
[4] Masud et al., "Multivariate soft rank via entropy-regularized optimal transport: sample efficiency and generative modeling," *JMLR*, 2023.
[5] Sudarshan et al., "Deep direct likelihood knockoffs," *NeurIPS*, 2020.

Model-X Knockoff

- Core ingredients: Learned knockoff variables $\tilde{X}$ and knockoff statistics $w_j((X, \tilde{X}), Y)$ for $j \in [p]$

Model-X Knockoff

- Core ingredients: Learned knockoff variables $\tilde{X}$ and knockoff statistics $w_j((X, \tilde{X}), Y)$ for $j \in [p]$
- Two required conditions for the knockoff variables and the knockoff statistics
  - Swap property: $(X, \tilde{X})_{\text{swap}(B)} \overset{d}{=} (X, \tilde{X}), \quad \forall B \subset [p]$
  - Flip-sign property:

$$w_j\left((X, \tilde{X})_{\text{swap}(B)}, Y\right) = \begin{cases} w_j((X, \tilde{X}), Y), & \text{if } j \notin B \\ -w_j((X, \tilde{X}), Y), & \text{if } j \in B \end{cases}$$

Model-X Knockoff

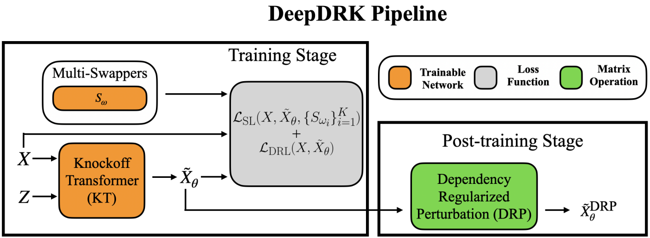
- Core ingredients: Learned knockoff variables $\tilde{X}$ and knockoff statistics $w_j((X, \tilde{X}), Y)$ for $j \in [p]$

- Two required conditions for the knockoff variables and the knockoff statistics
  - Swap property: $(X, \tilde{X})_{\mathrm{swap}(B)} \overset{d}{=} (X, \tilde{X}), \quad \forall B \subset [p]$
  - Flip-sign property:

$$w_j\left((X, \tilde{X})_{\mathrm{swap}(B)}, Y\right) = \begin{cases} w_j((X, \tilde{X}), Y), & \text{if } j \notin B \\ -w_j((X, \tilde{X}), Y), & \text{if } j \in B \end{cases}$$

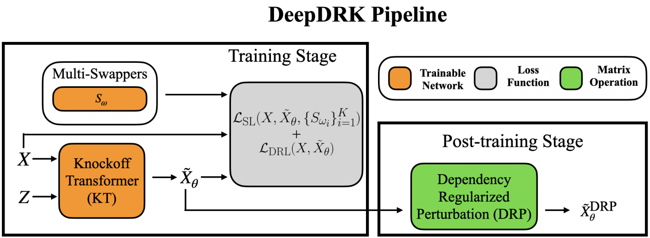
- Feature selection with controlled FDR at nominal level q:
  - Selection rule: $\mathcal{S} = \{w_j \geq \tau_q\}$
  - Threshold: $\tau_q = \min_{t>0}\left\{t : \frac{1+|\{j: w_j \leq -t\}|}{\max(1, |\{j: w_j \geq t\}|)} \leq q\right\}$

Knockoff Transformer and Multi-Swapper



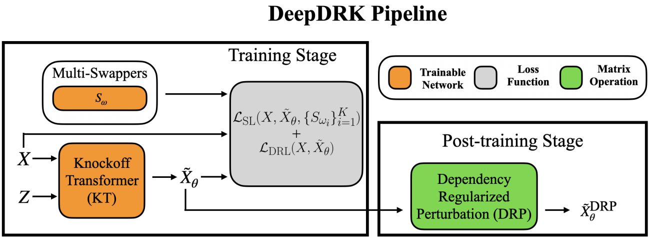
- The Knockoff Transformer takes X and i.i.d. standard Gaussian random variables Z as the inputs to generate the knockoffs $\tilde{X}_\theta$

Knockoff Transformer and Multi-Swapper



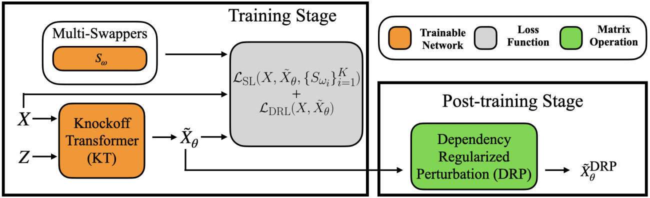
- The Knockoff Transformer takes X and i.i.d. standard Gaussian random variables Z as the inputs to generate the knockoffs $\tilde{X}_\theta$
- Use K swappers $\{S_{\omega_i}\}_{i=1}^K$ to create adversarial environments for testing the swap property
- The swap loss $\mathcal{L}_{\mathrm{SL}}(X, \tilde{X}_\theta, \{S_{\omega_i}\}_{i=1}^K)$ aims to enforce the swap property

Knockoff Transformer and Multi-Swapper



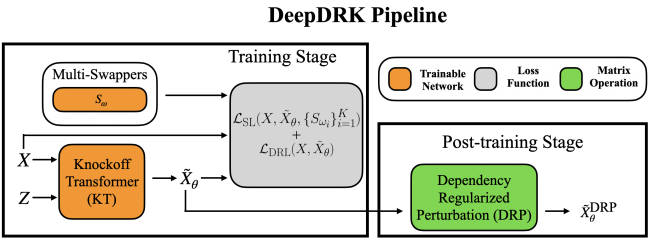
- The Knockoff Transformer takes X and i.i.d. standard Gaussian random variables Z as the inputs to generate the knockoffs $\tilde{X}_\theta$
- Use K swappers $\{S_{\omega_i}\}_{i=1}^K$ to create adversarial environments for testing the swap property
- The swap loss $\mathcal{L}_{\mathrm{SL}}(X, \tilde{X}_\theta, \{S_{\omega_i}\}_{i=1}^K)$ aims to enforce the swap property
- The dependency regularization loss $\mathcal{L}_{\mathrm{DRL}}(X, \tilde{X}_\theta)$ aims to decorrelate the data X and the knockoff $\tilde{X}_\theta$

DeepDRK Pipeline



- The Knockoff Transformer takes X and i.i.d. standard Gaussian random variables Z as the inputs to generate the knockoffs $\tilde{X}_\theta$
- Use K swappers $\{S_{\omega_i}\}_{i=1}^K$ to create adversarial environments for testing the swap property
- The swap loss $\mathcal{L}_{\mathrm{SL}}(X, \tilde{X}_\theta, \{S_{\omega_i}\}_{i=1}^K)$ aims to enforce the swap property
- The dependency regularization loss $\mathcal{L}_{\mathrm{DRL}}(X, \tilde{X}_\theta)$ aims to decorrelate the data X and the knockoff $\tilde{X}_\theta$
- Training: $\min_\theta \max_{\omega_1,\ldots,\omega_K} \{\mathcal{L}_{\mathrm{SL}}(X, \tilde{X}_\theta, \{S_{\omega_i}\}_{i=1}^K) + \mathcal{L}_{\mathrm{DRL}}(X, \tilde{X}_\theta)\}$

Post-training Perturbation



- Perturb the learned knockoff $\tilde{X}_\theta$:

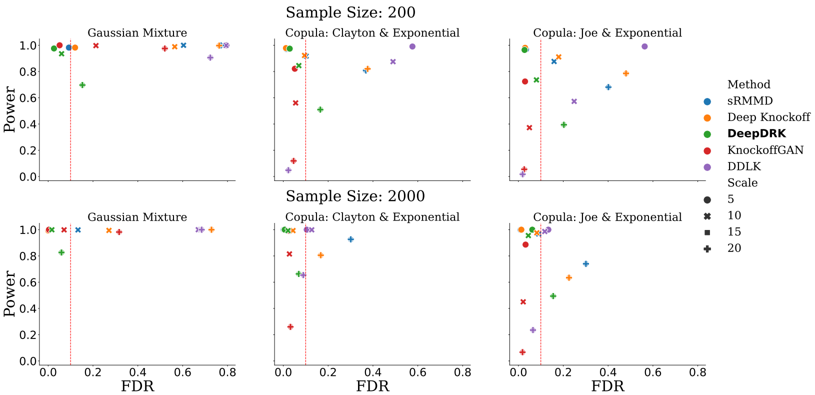
$$\tilde{X}_{\theta,n}^{\mathrm{DRP}} = (1 - \alpha_n) \cdot \tilde{X}_\theta + \alpha_n \cdot X_{\mathrm{rp}},$$

  where X_{rp} is the random row permutation of the design matrix X

- The perturbation aims to reduce collinearity[1]
- As $n \to \infty$, $\alpha_n \to 0$

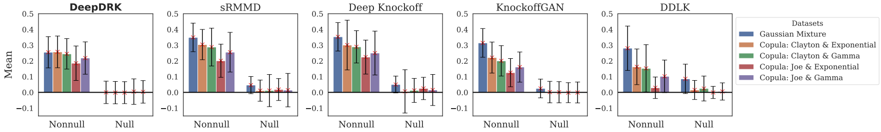
[1] Spector et al., "Powerful knockoffs via minimizing reconstructability," *Ann. Stat.*, 2022.

Results on Synthetic Data



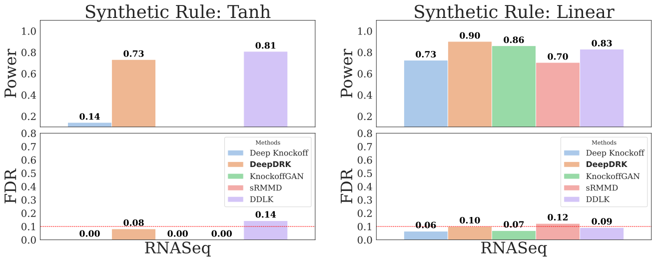
- Sample size: $n = 200$ or 2000; data dimension: $p = 100$
- Model: $Y \sim \mathcal{N}(X^T\beta, 1)$; feature sparsity: 20
- Nonnull $\beta_j \sim \frac{p}{\text{scale} \cdot \sqrt{n}} \cdot \text{Rademacher}(0.5)$
- FDR nominal threshold $q = 0.1$

The Behavior of Knockoff Statistics



- Compare the means and the standard deviations of the knockoff statistics w_j's
- Positive shifts in the null knockoff statistics from baseline models cause:
  - smaller thresholds τ_q, as there are fewer null statistics remaining on the negative side (lower $|\{j : w_j \leq -t\}|$), where $\tau_q = \min_{t>0} \left\{ t : \frac{1+|\{j:w_j \leq -t\}|}{\max(1,|\{j:w_j \geq t\}|)} \leq q \right\}$
  - increase in the number of false positives given the selection rule $\mathcal{S} = \{w_j \geq \tau_q\}$.

Results on Semi-synthetic Data



- X drawn from single-cell RNA sequencing (scRNA-seq)[1] and used to simulate response Y
- $n = 10000$ and $p = 100$

[1] Hansen et al., "Normalizing flows for knockoff-free controlled feature selection," *NeurIPS*, 2022.

Summary

- We developed DeepDRK for feature selection with controlled FDR for non-Gaussian data and limited sample size
- Paper link: https://arxiv.org/pdf/2402.17176v2
- GitHub: https://github.com/nowonder2000/DeepDRK

Thank you! Please feel free to reach out to us at poster session or via email.