

# DeepPINK: reproducible feature selection in deep neural networks[1]

Lu, Yang, et al.

NeurIPS

volume 31 (2018)

---

Presenter: Dahun Seo<sup>1</sup>

March 7, 2024

<sup>1</sup>Department of Statistics, Seoul National University

---

# DeepPINK: reproducible feature selection in deep neural networks

---

**Yang Young Lu \***

Department of Genome Sciences  
University of Washington  
Seattle, WA 98195  
ylu465@uw.edu

**Yingying Fan \***

Data Sciences and Operations Department  
Marshall School of Business  
University of Southern California  
Los Angeles, CA 90089  
fanyingy@marshall.usc.edu

**Jinchi Lv**

Data Sciences and Operations Department  
Marshall School of Business  
University of Southern California  
Los Angeles, CA 90089  
jinchilv@marshall.usc.edu

**William Stafford Noble**

Department of Genome Sciences and Department of Computer Science and Engineering  
University of Washington  
Seattle, WA 98195  
william-noble@uw.edu

# Contents

1. Introduction
2. Methods
3. Simulation studies
4. Real data analysis
5. Discussion
6. References

# 1. Introduction

---

## Introduction

### Deep neural network (DNN)

- ▶ Pros: Outstanding empirical performance in both supervised and unsupervised machine learning tasks
- ▶ Cons: Hard to interpret the Model's output (Black Box)

### Existing methods for feature selection in DNN

- ▶ Most existing methods focus on the feature importance in DNN using a simple idea "locally perturb the input to see how the prediction changes".
- ▶ However, these are susceptible to noise and lack of robustness. (Small random perturbation can dramatically change the feature importance.)  
And hard to control FDR.

## Introduction (cont.)

p-value?

- ▶ If there is a method that produces meaningful p-value for each features, we can use it along with FDR procedure.
- ▶ However, this is still completely unknown.

knockoff framework

- ▶ knockoff framework is a feature selection framework with FDR control that do not need any p-values.

Goal

- ▶ Develop a feature selection method of DNNs by incorporating the idea of knockoff framework
- ▶ They named the new method "DeepPINK".

## Background (FDR)

### FDR

- ▶ Let  $S_0 \subset \{1, \dots, p\}$  be the true feature set. Define FDP (false discovery proportion) and FDR (false discovery rate):

$$\text{FDP} = \frac{|\widehat{S} \cap S_0^c|}{|\widehat{S}| \vee 1}$$

$$\text{FDR} = \mathbb{E}[\text{FDP}]$$

where  $\widehat{S}$  be a set of features selected by some feature selection procedure.

- ▶ To control the false discovery in data analysis, we have to develop a feature selection method that can control FDR.

## Background (knockoff framework)

### Definition 1 (model-X knockoff features)

Model-X knockoff features for the family of random variables

$X = (X_1, \dots, X_p)^\top$  are a new family of random variables

$\tilde{X} = (\tilde{X}_1, \dots, \tilde{X}_p)^\top$  that satisfies two properties:

- 1  $(X, \tilde{X})_{\text{swap}(S)} \stackrel{d}{=} (X, \tilde{X})$  for any subset  $S \subset \{1, \dots, p\}$ , where  $\text{swap}(S)$  means swapping  $X_j$  and  $\tilde{X}_j$  for each  $j \in S$ .
- 2  $\tilde{X} \perp\!\!\!\perp Y | X$   
(the knockoff variables  $\tilde{X}$  is independent of the response variable given  $X$ .)

- In simple terms (not strictly precise), knockoff variables are variables that are unrelated to the response variable  $Y$  but share similar characteristics with the original variables.



## Background (knockoff framework) (cont.)

How to generate knockoff features?

- ▶ There are various methods to generate  $\tilde{X}$ . [2]–[4]

knockoff statistic

- ▶ Construct the knockoff statistics  $W_j = g_j(Z_j, \tilde{Z}_j)$  for  $1 \leq j \leq p$  where  $Z_j$  and  $\tilde{Z}_j$  represent feature importance measures for the  $j$ th feature  $X_j$  and its knockoff  $\tilde{X}_j$  respectively and  $g_j(\cdot, \cdot)$  is any antisymmetric function.
- ▶ A desirable property for  $W_j$  is that important features are expected to have large positive values, whereas unimportant ones should have small values around 0.

## Background (knockoff framework) (cont.)

threshold for FDR control

- ▶ To determine which variables are important, we have to decide threshold  $\delta$ .

$$\widehat{S} = \{j : |W_j| \geq \delta\}$$

- ▶ To control the FDR under  $q \in (0, 1)$ , set the threshold  $\delta$ :

$$\delta = \min \left\{ t \in \mathcal{W} : \frac{1 + |\{j : W_j \leq -t\}|}{1 \vee |\{j : W_j \geq t\}|} \leq q \right\}$$

where  $\mathcal{W} = \{|W_j| : 1 \leq j \leq p\} \setminus \{0\}$ .

## 2. Methods

---

## Proposed method: DeepPINK

### DeepPINK

- ▶ Application of knockoff framework on DNN

### DNN structure of DeepPINK

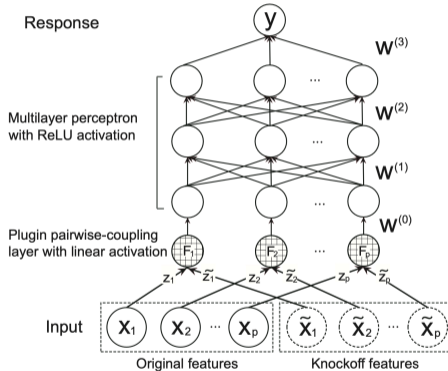
- ▶ They developed "plugin pairwise-coupling layer"  $F_1, \dots, F_p$  for each variables  $(X_1, \tilde{X}_1), \dots, (X_p, \tilde{X}_p)$ .

$$F_j(X_j, \tilde{X}_j) = z_j X_j + \tilde{z}_j \tilde{X}_j$$

Intuitively,  $z_j$  being much larger than  $\tilde{z}_j$  provides some evidence that the  $j$ th feature is important, whereas similar values of  $z_j$  and  $\tilde{z}_j$  indicate that the  $j$ th feature is not important.

## Proposed method: DeepPINK (cont.)

- ▶ They proposed 2 hidden layer having  $p$  hidden node DNN structure.



- ▶ Let  $w$  be  $p \times 1$  vector such that

$$w = W^{(0)} \odot (W^{(1)}W^{(2)}W^{(3)})$$

## Proposed method: DeepPINK (cont.)

Proposed knockoff statistic

- ▶ Feature importance measures:

$$Z_j = z_j \times w_j \text{ and } \tilde{Z}_j = \tilde{z}_j \times w_j$$

- ▶ Assymmetric function  $g_j$ :

$$g_j(x_1, x_2) = x_1^2 - x_2^2$$

for all  $j = 1, \dots, p$ .

- ▶ Thus, we have knockoff statistics  $W_j$ :

$$W_j = Z_j^2 - \tilde{Z}_j^2$$

for all  $j = 1, \dots, p$ .

### **3. Simulation studies**

---

## Simulation setting

### Explanatory variables and error

- ▶ explanatory variables  $x_i$  are from  $\mathcal{N}(0, \Sigma)$  with  $\Sigma^{-1} = (0.5^{|j-k|})_{1 \leq j, k \leq p}$
- ▶ error  $\varepsilon_i$  follows standard normal distribution.

### True coefficient $\beta_0$

- ▶  $\beta_0 \in \mathbb{R}^p$  nonzero elements randomly located with  $s = \|\beta_0\|_0$  fixed.
- ▶ nonzero elements are randomly chosen from  $\{-1.5, 1.5\}$ .



## Simulation setting (cont.)

Two scenarios:

1 Linear:

$$Y_i = x_i^\top \beta_0 + \varepsilon_i, \quad i = 1, \dots, n$$

where  $s = 30$

2 Single-Index (non Linear):

$$Y_i = g(x_i^\top \beta_0) + \varepsilon_i, \quad i = 1, \dots, n$$

where  $s = 10$  and  $g(x) = x^3/2$ .

## Simulation setting (cont.)

Other competitive models with feature importance measure:

- ▶ DNN without "plugin pairwise-coupling layer"
- ▶ DeepLIFT[5] with the multiplier score
- ▶ random forest (RF) with feature importance score
- ▶ linear kernel SVM with coefficient

# Simulation results

- ▶ repetition number: 20
- ▶ In all scenarios, FDR is controlled at  $q = 0.2$ .

$p$	DeepPINK				MLP				DeepLIFT				RF				SVR			
	Linear		Single-Index		Linear		Single-Index		Linear		Single-Index		Linear		Single-Index		Linear		Single-Index	
	FDR	Power	FDR	Power	FDR	Power	FDR	Power	FDR	Power	FDR	Power	FDR	Power	FDR	Power	FDR	Power	FDR	Power
50	0.046	1	0.13	0.98	0.15	1	0.17	0.89	0.16	1	0.24	0.9	0.005	0.45	0	0	0.18	1	0.180	0.81
100	0.047	1	0.08	1	0.048	1	0.056	0.26	0.16	1	0.13	0.47	0.016	0.61	0.025	0.045	0.22	1	0.094	0.26
200	0.042	0.99	0.042	1	0.11	1	0	0	0.24	0.96	0.034	0.067	0.013	0.54	0.020	0.045	0.21	1	0.061	0.05
400	0.022	0.97	0.022	1	0.29	0.95	0	0	0.034	0.5	0.039	0.069	0.017	0.53	0.033	0.050	0.22	1	0.083	0.01
600	0.031	0.95	0.046	1	0.17	0.8	0.014	0.013	0.003	0.26	0.068	0.16	0.023	0.56	0.11	0.095	0.19	1	0	0
800	0.048	0.95	0.082	1	0.037	0.62	0.016	0.068	0	0.17	0.16	0.24	0.022	0.61	0.061	0.12	0.22	0.98	0	0
1000	0.023	0.97	0.065	1	0.007	0.4	0.037	0.16	0	0.12	0.013	0.33	0.029	0.59	0.081	0.17	0.15	0.67	0	0
1500	0.007	1	0.065	1	0.002	0.41	0.068	0.25	0.001	0.32	0.13	0.44	0.045	0.58	0.098	0.17	0.064	0.043	0	0
2000	0.026	0.99	0.098	1	0.023	0.4	0.063	0.35	0.015	0.37	0.1	0.56	0.033	0.65	0.046	0.14	0.04	0.002	0	0
2500	0.029	0.97	0.067	1	0.21	0.5	0.042	0.35	0.088	0.58	0.32	0.47	0.034	0.62	0.11	0.18	0.02	0.005	0	0
3000	0.046	0.97	0.051	1	0.11	0.43	0.046	0.31	0.069	0.46	0.14	0.44	0.05	0.65	0.087	0.17	0.05	0	0	0

Table 1: Simulation results for linear model and the Single-Index model.

## **4. Real data analysis**

---

## Application to HIV-1 data

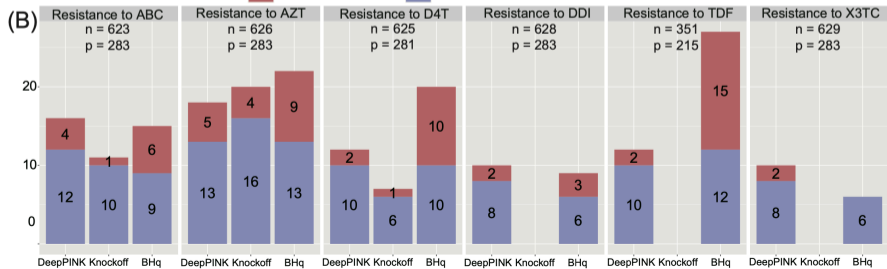
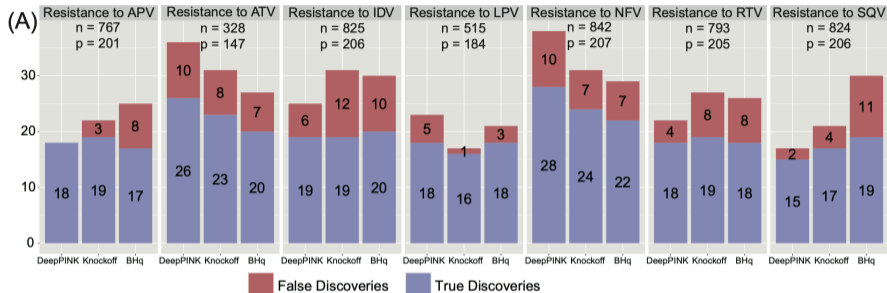
### Goal

- ▶ Identify mutation associated with drug resistance level

### Data

- ▶ 13 datasets (13 kinds of drug)
- ▶ response variable: Drug resistance level
- ▶ explanatory variables: mutation (present or absence)
- ▶ Actually, we knew the associated mutation for each drug.  
(So we can calculate the false discovery rate)

# Application to HIV-1 data (cont.)



## Application to gut microbiome data

### Goal

- ▶ Identify the important nutrient intake and the important gut microbiome that are associated with BMI

### Data

- ▶  $n = 98, p = p_1 + p_2 = 301$
- ▶ response variable: BMI
- ▶ explanatory variables: nutrient intake ( $p_1 = 214$ ) and CLR transformed gut microbiome ( $p_2 = 87$ )

## Application to gut microbiome data (cont.)

Nutrient intake		Bacteria genera		
Micronutrient	Reference	Phylum	Genus	Reference
Linoleic	[7]	Firmicutes	Clostridium	[26]
Dairy Protein	[29]	Firmicutes	Acidaminococcus	[26]
Choline, Phosphatidylcholine	[31]	Firmicutes	Allisonella	[26]
Choline, Phosphatidylcholine w/o suppl.	[31]	Firmicutes	Megamonas	[25]
Omega 6	[39]	Firmicutes	Megasphaera	[43]
Phenylalanine, Aspartame	[41]	Firmicutes	Mitsuokella	[43]
Aspartic Acid, Aspartame	[41]	Firmicutes	Holdemania	[30]
Theaflavin 3-gallate, flavan-3-ol(2)	[42]	Proteobacteria	Sutterella	[13]



## **5. Discussion**

---

# Summary

## Goal

- ▶ Develop a feature selection method of DNNs by incorporating the idea of knockoff framework

## Method

- ▶ They proposed "plugin pairwise-coupling layer" in DNN structure.
- ▶ Because of "plugin pairwise-coupling layer", DeepPINK achieves FDR control with much higher power than the naive combination of the knockoffs idea with a vanilla MLP.

Thank you

## 6. References

---

## References

- [1] Y. Lu, Y. Fan, J. Lv, and W. Stafford Noble, “**Deeppink: Reproducible feature selection in deep neural networks,**” *Advances in neural information processing systems*, vol. 31, 2018.
- [2] R. F. Barber and E. J. Candès, “**Controlling the false discovery rate via knockoffs,**”, 2015.
- [3] E. Candès, Y. Fan, L. Janson, and J. Lv, “**Panning for gold: ‘model-x’ knockoffs for high dimensional controlled variable selection,**” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 80, no. 3, pp. 551–577, 2018.

## References (cont.)

- [4] Y. Romano, M. Sesia, and E. Candès, “**Deep knockoffs,**” *Journal of the American Statistical Association*, vol. 115, no. 532, pp. 1861–1872, 2020.
- [5] A. Shrikumar, P. Greenside, and A. Kundaje, “**Learning important features through propagating activation differences,**” in *International conference on machine learning*, PMLR, 2017, pp. 3145–3153.