

# Additive Noise Models: Identifiability, Learning Algorithms, Hidden Variables and Cycles

بهراد منیری  
دانشکده‌ی مهندسی برق دانشگاه صنعتی شریف  
bemoniri@ee.sharif.edu

## فهرست مطالب

۱	قضایای قابل شناسایی بودن	۱
۱	۱.۱ حالت دو متغیره	۱
۴	۲.۱ حالت چند متغیره	۴
۶	۲ الگوریتم‌های یادگیری	۶
۶	۱.۲ الگوریتم RESIT	۶
۷	۲.۲ الگوریتم‌های مبتنی بر Independence Score	۷

## ۱ قضایای قابل شناسایی بودن

در این بخش به بررسی مدل‌های نويز جمعی در حالاتی که هیچ Common Cause مشاهده نشده‌ای در سیستم وجود ندارد می‌پردازیم. در مدل‌های نويز جمعی، هدف محدود کردن مجموعه‌ی توابعی است که در آن به جست‌وجوی توابع تولید کننده SCM می‌گردیم. مراجع ما در این بخش، مقالات [۲]، [۴] و [۵] هستند. این مقالات به بررسی قضایای Identifiability در مدل‌های نويز جمعی پرداخته‌اند. ما در این بخش به بررسی قضایای مطرح شده در این مقالات پرداخته و در آخر محدودیت‌های آن‌ها را با هم مقایسه می‌کنیم.

**تعریف ۱.۱.** یک مدل نويز جمعی پیوسته را به صورت زیر تعریف می‌شود:

$$S_j : S_j = f_j(\mathbf{PA}_j) + N_j, \quad j = 1, 2, \dots, p \quad (1)$$

که در آن  $\mathbf{PA}_j$  مجموعه‌ی والدین  $X_j$  هستند و متغیرهای نويز،  $N_j$  دارای چگالی احتمالی اکیداً مثبت می‌باشند و مستقل‌اند. همچنین در این بخش فرض می‌شود گراف مولد داده‌ها یک DAG است.

**قضیه ۱.۱.** در مدل‌های نويز جمعی، شرط *Causal Minimality* معادل این است که توابع  $f_j$  نسبت به هیچ یک از متغیرهایشان ثابت نباشند.

### ۱.۱ حالت دو متغیره

Hoyer et al. در مقاله‌ی [۲] قضیه‌ی زیر را در مورد قابل شناسایی بودن ANM در حالت دو متغیره اثبات می‌کند. در ادامه به بررسی دقیق این قضیه پرداخته و بحث خواهیم کرد که در چه شرایطی، در یک مدل جمعی، از روی چگالی احتمال مشترک قادر به شناسایی کامل جهت علی نیستیم.

**قضیه ۲.۱.** فرض کنید داشته باشیم

$$\begin{cases} X = N_x \\ Y = f(X) + N_y \end{cases}$$

اگر برای هر  $x, y$  با شرط  $\nu''(y - f(x))f'(x) \neq 0$  حل معادله‌ی دیفرانسیل زیر نباشد:

$$\xi''' = \xi'' \left( -\frac{\nu''' f'}{\nu''} + \frac{f''}{f'} \right) - 2\nu'' f'' f' + \nu' f''' + \frac{\nu' \nu''' f'' f'}{\nu''} - \frac{\nu' (f'')^2}{f'}, \quad (2)$$

که در آن  $\xi := \log(p_X)$  و  $\nu := \log(p_{N_Y})$  است، مدلی با نویز جمعی پیوسته در جهت دیگر وجود نداشته و جهت علی، با داشتن توزیع مشترک متغیرها قابل شناسایی است. در شرط فوق، آرگومان توابع و مشتق توابع  $\nu, \xi$  و  $f$  به ترتیب  $x, y - f(x)$  و  $x$  هستند.

اثبات. فرض کنید که گراف قابل شناسایی نباشد و دو گراف با جهت‌های متضاد بر این داده‌ها قابل برازش باشد. در این صورت خواهیم داشت:

$$p(x, y) = p_n(y - f(x))p_x(x) = p_{\bar{n}}(x - g(y))p_y(y). \quad (3)$$

$\pi$  را برابر تابع log-likelihood در نظر بگیرید:

$$\pi(x, y) := \log p(x, y) = \nu(y - f(x)) + \xi(x), \quad (4)$$

از طرف دیگر، بنا بر معادله‌ی (۳) داریم:

$$\pi(x, y) = \tilde{\nu}(x - g(y)) + \eta(y) \quad (5)$$

با مشتق‌گیری از این معادله خواهیم داشت:

$$\frac{\partial^2 \pi}{\partial x \partial y} = -\tilde{\nu}''(x - g(y))g'(y) \quad \text{و} \quad \frac{\partial^2 \pi}{\partial x^2} = \tilde{\nu}''(x - g(y)).$$

در نتیجه داریم:

$$\frac{\partial}{\partial x} \left( \frac{\partial^2 \pi / \partial x^2}{\partial^2 \pi / (\partial x \partial y)} \right) = -\frac{\partial}{\partial x} g'(y) = 0. \quad (6)$$

از معادله‌ی (۴) به دست می‌آید:

$$\frac{\partial^2 \pi}{\partial x \partial y} = -\nu''(y - f(x))f'(x), \quad (7)$$

و همچنین

$$\frac{\partial^2 \pi}{\partial x^2} = \frac{\partial}{\partial x} (-\nu'(y - f(x))f'(x) + \xi'(x)) = \nu''(f')^2 - \nu' f'' + \xi'', \quad (8)$$

که در آن آرگومان‌ها برای خوانایی بیشتر حذف شده‌اند. از معادله‌ی (۷) و (۸)

$$\frac{\partial}{\partial x} \left( \frac{\frac{\partial^2 \pi}{\partial x^2}}{\frac{\partial^2 \pi}{\partial x \partial y}} \right) = -2f'' + \frac{\nu' f'''}{\nu'' f'} - \xi''' \frac{1}{\nu'' f'} + \frac{\nu' \nu''' f''}{(\nu'')^2} - \frac{\nu' (f'')^2}{\nu'' (f')^2} - \xi'' \frac{\nu'''}{(\nu'')^2} + \xi'' \frac{f''}{\nu'' (f')^2}.$$

بر اساس معادله‌ی (۶)، عبارت فوق برابر صفر است یعنی:

$$\xi''' = \xi'' \left( -\frac{\nu''' f'}{\nu''} + \frac{f''}{f'} \right) - 2\nu'' f'' f' + \nu' f''' + \frac{\nu' \nu''' f'' f'}{\nu''} - \frac{\nu' (f'')^2}{f'}$$

با بازنویسی این معادله بر حسب توابع اصلی مساله داریم:

$$p_x''' = p_x'' p_n'' f' \left( -\frac{p_n'''}{(p_n'')^2} + \frac{f''}{p_n''(f')^2} \right) + p_n'' f' \left( -2f'' + \frac{p_n' f'''}{p_n'' f'} + \frac{p_n' p_n''' f''}{(p_n'')^2} - \frac{p_n'(f'')^2}{p_n''(f')^2} \right).$$

□

سوال مهمی که در این مطرح می شود این است که در چه شرایطی، قادر به تشخیص جهت از روی چگال مشترک نیستیم. قضیه زیر به ما می گوید که در صورتی که به صورت کاملاً تصادفی از مدل های نوین جمعی یک مدل را انتخاب کنیم، احتمال قابل شناسایی نبودن جهت علی صفر است.

**قضیه ۳.۱.** اگر برای مجموعه ی توابع  $f$  و توزیع نویزهای خارجی داده شده، برای یک  $y$  خاص،  $v''(y-f(x))f'(x) = 0$ ، شمارا جواب داشته باشد، یا جوابی نداشته باشد، توزیع  $X$  در یک فضای سه بعدی زندگی می کند.

از آنجا که مجموعه ی توزیع های پیوسته بی نهایت بعدی است، برای "اکثر" مدل های نوین جمعی، جهت علی قابل شناسایی است.

اثبات.  $y$  فیکس ای در نظر بگیرید به طوری که  $v''(y-f(x))f'(x) \neq 0$  در همه ی  $x$  ها به جز تعدادی شمارا برقرار باشد. برای هر  $\nu, f$  داده شده، بنا بر قضیه ۲.۱ یک معادله ی دیفرانسیل برای  $\xi$  به دست می آوریم:

$$\xi'''(x) = \xi''(x)G(x, y) + H(x, y), \quad (9)$$

که در آن  $H$  و  $G$  برابرند با

$$G := -\frac{\nu''' f'}{\nu''} + \frac{f''}{f'}$$

و

$$H := -2\nu'' f'' f' + \nu' f''' + \frac{\nu' \nu''' f'' f'}{\nu''} - \frac{\nu'(f'')^2}{f'},$$

با حل این معادله ی دیفرانسیل برای  $\xi''$  داریم

$$\xi''(x) = \xi''(x_0) e^{\int_{x_0}^x G(\tilde{x}, y) d\tilde{x}} + \int_{x_0}^x e^{\int_{\tilde{x}}^x G(\tilde{x}, y) d\tilde{x}} H(\tilde{x}, y) d\tilde{x}. \quad (10)$$

مجموعه ی توابع  $\xi$  ای که در معادله ی دیفرانسیل مذکور صدق می کنند، در یک زیرفضای سه بعدی آفین زندگی می کنند که با سه عدد  $\xi(x_0), \xi'(x_0), \xi''(x_0)$  تابع به طور یکتا تعیین می شود. در نتیجه اثبات کردیم که برای یک تابع و مجموعه ی توزیع نویزهای خارجی داده شده، مجموعه ی توزیع احتمال  $X$  هایی که به اجازه ی وجود یک مدل برعکس می دهند، در یک زیرفضای سه بعدی از فضای بی نهایت بعدی توزیع های پیوسته هستند. □

با وجود اینکه احتمال اینکه برای یک مدل نوین جمعی، مدل نوین جمعی دیگری در جهت مخالف وجود داشته باشد بسیار نادر است، این سوال مطرح است که در چه شرایطی برای یک مدل نوین جمعی چنین اتفاقی رخ می دهد. Zhang et al. در [۵] پنج دسته مدل نوین جمعی معرفی می کند و اثبات می کند هر مدل نوین جمعی غیر قابل شناسایی از تابع چگالی احتمال، به ناچار در یکی از این دسته ها قرار می گیرد.

**قضیه ۴.۱.** فرض کنید  $X_2 = f_2(X_1) + N_2$  باشد و  $N_2$  نیز نویزی  $full support$  و مستقل از  $X_1$  باشد، تابع  $f_2$  سه بار مشتق پذیر بوده و همچنین معادله ی  $\frac{d}{dx_1} f_2(x_1) \frac{d'}{dx_1} \log(p_{N_2}(x_2)) = 0$  تنها در تعدادی متناهی نقطه ی  $(x_1, x_2)$  برقرار باشد. در صورتی که یک مدل در جهت برعکس وجود داشته باشد، به این معنا که  $X_1 = g_1(X_2) + \tilde{N}_1$  که  $X_2$  و  $\tilde{N}_1$  در آن مستقل باشد، یکی از پنج حالت زیر برقرار است.

تعاریف دقیق عبارات به کار رفته در جدول فوق را در تعریف زیر آورده ایم:

**تعریف ۲.۱.** فرض کنید  $p$  چگالی احتمال یک توزیع پیوسته  $P$  باشد.

Table 1: All situations in which the PNL causal model is not identifiable.

	$p_{e_2}$	$p_{t_1} (t_1 = g_2^{-1}(x_1))$	$h = f_1 \circ g_2$	Remark
I	Gaussian	Gaussian	linear	$h_1$ also linear
II	log-mix-lin-exp	log-mix-lin-exp	linear	$h_1$ strictly monotonic, and $h'_1 \rightarrow 0$ , as $z_2 \rightarrow +\infty$ or as $z_2 \rightarrow -\infty$
III	log-mix-lin-exp	one-sided asymptotically exponential (but not log-mix-lin-exp)	$h$ strictly monotonic, and $h' \rightarrow 0$ , as $t_1 \rightarrow +\infty$ or as $t_1 \rightarrow -\infty$	—
IV	log-mix-lin-exp	generalized mixture of two exponentials	Same as above	—
V	generalized mixture of two exponentials	two-sided asymptotically exponential	Same as above	—

•  $P$  یک  $\log\text{-mix-lin-exp}$  است اگر وجود داشته باشند  $c_1, c_2, c_3, c_4$  به نحوی که  $c_1 < 0$  و  $c_2 c_3 > 0$  به صورتی که:

$$\log p(x) = c_1 \exp(c_2 x) + c_3 x + c_4.$$

•  $P$  one-sided asymptotically exponential است اگر وجود داشته باشد  $c \neq 0$  به نحوی که

$$\frac{d}{dx} \log p(x) \rightarrow c$$

وقتی  $x \rightarrow \infty$  یا  $x \rightarrow -\infty$ .

•  $P$  یک  $two\text{-sided asymptotically exponential}$  است اگر وجود داشته باشند  $c_1 \neq 0$  و  $c_2 \neq 0$  به نحوی که

$$\frac{d}{dx} \log p(x) \rightarrow c_1$$

وقتی  $x \rightarrow -\infty$  و

$$\frac{d}{dx} \log p(x) \rightarrow c_2$$

وقتی  $x \rightarrow \infty$ .

•  $P$  یک  $generalized mixture of two exponentials$  است اگر  $d_1, d_2, d_3, d_4, d_5, d_6$  وجود داشته باشند به نحوی که  $d_2 < -\frac{d_1}{d_5}$  و  $d_1 d_5 > 0, d_3 > 0, d_4 > 0$  داشته باشیم:

$$\log p(x) = d_1 x + d_2 \log(d_3 + d_4 \exp(d_5 x)) + d_6.$$

## ۲.۱ حالت چند متغیره

تا اینجا برای حالت دو بعدی، نشان دادیم در حالت generic یک توزیع احتمال، اجازه‌ی وجود مدل نوینز جمعی در هر دو طرف را نمی‌دهد. در این بخش به تعمیم این قضیه از دوبعدی به حالت چندبعدی می‌پردازیم. مرجع اصلی ما در این بخش مقاله‌ی [۴] است. این مقاله قضیه‌ای بسیار جالب را مطرح می‌کند که عنوان می‌کند هنگامی یک قضیه‌ی Identifiability دو بعدی داریم در چه صورتی می‌توان آن را به حالت چندبعدی تعمیم داد. برای ورود به این بحث مقاله‌ی [۴] مثالی جالب را مطرح کرده که در این گزارش نیز به همان شیوه‌ی مقاله عمل می‌کنیم.

**مثال ۱.۱. SCM** زیر را در نظر بگیرید:

$$\begin{cases} X_1 = N_1 \\ X_2 = f_2(X_1) + N_2 \\ X_3 = f_3(X_1) + aX_2 + N_3 \end{cases} \quad (11)$$

که در آن  $N_1 \sim t\text{-student}(\nu = 3)$ ،  $N_2 \sim \text{Normal}(0, \sigma_2^2)$  و  $N_3 \sim \text{Normal}(0, \sigma_3^2)$ . در این جا  $X_2$  و  $X_3$  غیر گاوسی هستند اما

$$X_3 | X_2 = x_2 = c + aX_2 | X_1 = x_1 + N_3$$

برای هر  $x_1$  یک معادله‌ی خطی- گاوسی است در حالی که هیچ‌یک از معادلات اصلی SCM گاوسی- خطی نیستند. می‌توانیم SCM دیگری بسازیم که در توزیع مشاهداتی تفاوتی با SCM اصلی نداشته باشد:

$$\begin{cases} X_1 = M_1 \\ X_2 = g_2(X_1) + bX_3 + M_2 \\ X_3 = g_3(X_1) + aX_2 + M_3 \end{cases} \quad (12)$$

به نظر می‌رسد باید شرطی بر روی توزیع‌های شرطی قرار دهیم!

برای بیان شرط قابل شناسایی بودن از توزیع احتمال مشاهداتی، به یک تعریف نیاز داریم:

**تعریف ۳.۱.** یک مدل نوینز جمعی با  $n$  متغیر را در نظر بگیرید. این مدل را یک مدل نوینز جمعی محدود شده می‌نامیم اگر برای هر  $j \in V$  و  $i \in PA_j$  و تمام مجموعه‌های  $S \subseteq V$  به طوری که  $PA_j \setminus \{i\} \subseteq S \subseteq ND_j \setminus \{i, j\}$  وجود داشته باشد  $x_S$  که  $ps(x_S) > 0$  به نحوی که

$$\left( f_j(x_{PA_j \setminus \{i\}}, \bigcap_{X_i} \cdot), P(X_i | X_S = x_S), P(N_j) \right)$$

در شرایط قضیه‌ی (۲.۱) صدق کند.

**قضیه ۵.۱.** فرض کنید که  $X = \{X_1, X_2, \dots, X_n\}$  توسط یک مدل نوینز جمعی محدود شده با گراف  $G$  تولید شده باشند و فرض کنید  $P(X)$  نسبت به  $G$  شرط  $Causal Minimality$  را ارضا کند. در این صورت  $G$  از روی توزیع احتمال  $P(X)$  قابل شناسایی است.

برای اثبات این قضیه نیاز به یک لم گرافی داریم که Chickering در سال ۱۹۹۵ اثبات کرده است [۱].

**لم ۱.۱.** فرض کنید  $G$  و  $G'$  دو DAG روی مجموعه‌ی متغیرهای  $X$  باشند. فرض کنید  $P(X)$  چگالی احتمالی همواره مثبت دارد که نسبت به  $G$  و  $G'$  مارکوف هستند و شرط  $Causal Minimality$  را ارضا می‌کنند. در این صورت متغیرهای  $L$  و  $Y$  وجود دارند که برای مجموعه‌های  $\{Y\} := PA_L^G \setminus \{Y\}$ ،  $\{L\} := PA_Y^{G'} \setminus \{L\}$ ،  $S := Q \cup R$  و  $R := PA_Y^{G'} \setminus \{L\}$  داشته باشیم:

•  $Y \rightarrow L$  در  $G$  و  $L \rightarrow Y$  در  $G'$ .

•  $S \subseteq ND_Y^G \setminus \{Y\}$  و  $S \subseteq ND_Y^{G'} \setminus \{L\}$

اثبات. (قضیه‌ی (۵.۱))

برهان خلف: فرض کنید دو مدل نوینز جمعی محدود شده با این توزیع احتمال وجود داشته باشند. یکی با گراف  $G$  و دیگری با گراف متفاوت  $G'$ . دو متغیر  $L$  و  $Y$  را بر اساس لم فوق انتخاب می‌کنیم و مشابه لم فوق، می‌گیریم  $S := Q \cup R$  و  $R := PA_Y^{G'} \setminus \{L\}$ ،  $Q := PA_L^G \setminus \{Y\}$ ،  $s = (q, r)$  را در نظر بگیرید و بنویسید  $Y^* := Y | S = s$  و  $L^* := L | S = s$  از لم فوق داریم:  $S \subseteq ND_L^G \setminus \{Y\}$  و این بدین معناست که  $\{Y\} \cup S \subseteq ND_L^G$  در نتیجه  $N_L \perp\!\!\!\perp \{Y\} \cup S$  در نتیجه می‌توان به معادله‌ی زیر رسید:

$$L^* = f_L(q, Y^*) + N_L \quad N_L \perp\!\!\!\perp Y^*$$

برای  $G'$  هم با استدلال مشابه می‌توان به معادله‌ی زیر رسید:

$$Y^* = g_Y(r, L^*) + N_Y \quad N_Y \perp\!\!\!\perp L^*$$

و این در تناقض با فرض مدل نوینز جمعی محدود شده است پس فرض خلف باطل بوده و  $G$  و  $G'$  برابرند.  $\square$

نکته‌ی بسیار جالبی که در این قضیه وجود دارد این است که می‌توان آن را برای هر قضیه‌ی Indentifiability دیگر نیز به کار برد، به شرط عدم وجود دور. به طور مثال به کمک آن می‌توان LINGAM و یا Post Non-Linear Additive Noise Model، که در ادامه به آن پرداخته خواهد شد، را از حالت دو بعدی به حالت چند بعدی تعمیم داد.

## ۲ الگوریتم‌های یادگیری

در این بخش به بررسی الگوریتم‌های یادگیری گراف مربوط به یک SCM از داده‌ی محدود به فرض اینکه در SCM مدل نویز جمعی برقرار باشد، پرداخته و الگوریتم‌های مختلف مطرح شده را با هم مقایسه می‌کنیم. مراجع ما در این بخش مقالات [۴] و [۳] هستند.

### ۱.۲ الگوریتم RESIT

این الگوریتم توسط Peters et al.، سال ۲۰۱۴، در [۴] معرفی شده است. ایده‌ی اصلی این الگوریتم این است که برای هر  $X_i$  اگر  $X_i$  یک Sink Node باشد داریم  $N_i \perp\!\!\!\perp \mathbf{X} \setminus \{X_i\}$  به طور کلی برای هر  $N_Y \perp\!\!\!\perp \mathbf{ND}_Y$ ،  $Y \in \mathbf{X}$

---

#### Algorithm 1 Regression with subsequent independence test (RESIT)

---

```

1: Input: i.i.d. samples of a  $p$ -dimensional distribution on  $(X_1, \dots, X_p)$ 
2:  $S := \{1, \dots, p\}, \pi := []$ 
3: PHASE 1: Determine causal order.
4: repeat
5:   for  $k \in S$  do
6:     Regress  $X_k$  on  $\{X_i\}_{i \in S \setminus \{k\}}$ .
7:     Measure dependence between residuals and  $\{X_i\}_{i \in S \setminus \{k\}}$ .
8:   end for
9:   Let  $k^*$  be the  $k$  with the weakest dependence.
10:   $S := S \setminus \{k^*\}$ 
11:   $\text{pa}(k^*) := S$ 
12:   $\pi := [k^*, \pi]$  ( $\pi$  will be the causal order, its last component being a sink)
13: until  $\#S = 1$ 
14: PHASE 2: Remove superfluous edges.
15: for  $k \in \{2, \dots, p\}$  do
16:   for  $\ell \in \text{pa}(\pi(k))$  do
17:     Regress  $X_{\pi(k)}$  on  $\{X_i\}_{i \in \text{pa}(\pi(k)) \setminus \{\ell\}}$ .
18:     if residuals are independent of  $\{X_i\}_{i \in \{\pi(1), \dots, \pi(k-1)\}}$  then
19:        $\text{pa}(\pi(k)) := \text{pa}(\pi(k)) \setminus \{\ell\}$ 
20:     end if
21:   end for
22: end for
23: Output:  $(\text{pa}(1), \dots, \text{pa}(p))$ 

```

---

الگوریتم RESIT در هر مرحله یک Sink Node را تشخیص داده و حذف می‌کند. برای تشخیص یک Sink نیز از ویژگی  $N_i \perp\!\!\!\perp \mathbf{X} \setminus \{X_i\}$  استفاده می‌کند.

الگوریتم RESIT دو فاز دارد. در فاز اول (خط ۳ تا ۱۳)، یک Causal Order پیدا می‌شود. با رگرسیون کردن هر متغیر روی بقیه‌ی متغیرهای گراف هر مرحله، متغیری که باقی مانده‌ی رگرسیون مربوط به از دیگر متغیرها مستقل‌تر (مثلاً با معیار  $p$ -value آزمون HSIC) باشد را به عنوان یک sink در نظر می‌گیریم. با حذف این راس، مجدداً یک DAG دیگر به وجود می‌آید که در آن همین روند را روی آن تکرار می‌کنیم. با این کار می‌توان به یک Causal Order برای متغیرها رسید. در فاز دوم، برای شروع فرض می‌شود که اگر  $\pi(i) < \pi(j)$ ، از  $i$  به  $j$  یک یال وجود دارد. از این گراف شروع کرده. هر بار یک متغیر،  $X_{\pi(k)}$  را در نظر گرفته و آن را بر روی parent هایش به جز یک  $X_l$ ، رگرسیون می‌کنیم به نحوی که هر parent یک بار از رگرسیون کنار گذاشته شود. در هر رگرسیون، اگر باقی مانده رگرسیون از متغیرهایی که در Causal Order بالاتر از  $X_{\pi(k)}$  هستند مستقل شد، ارتباط  $X_l$  و  $X_{\pi(k)}$  را حذف می‌کنیم.

الگوریتم RESIT در مرحله‌ی اول خود  $O(n^2)$  تست آماری انجام می‌دهد و در مرحله‌ی دوم نیز تعداد تست‌های آماری  $O(n)$  است. چند جمله‌ای بودن این الگوریتم بسیار عجیب است زیرا مسائل معمول در Bayesian Network

Learning اکثراً NP-Hard هستند. با این وجود الگوریتم RESIT برای  $n$  های بزرگ قابل استفاده نیست زیرا در صورتی که در انجام تست آماری دچار خطا شویم، خطا به شدت در مراحل بعد منتشر شده و باعث می شود به طور قابل ملاحظه ای از گراف اصلی دور شویم.

## ۲.۲ الگوریتم های مبتنی بر Independence Score

یک دسته ی دیگر از الگوریتم های یادگیری در مدل های نويز جمعی، الگوریتم های مبتنی بر score هستند. یک الگوریتم دیگر برای یادگیری ساختار مدل های نويز جمعی محدود شده، این است که تمام DAG ها را enumerate کنیم و تست های استقلال مطرح شده را با آنها چک کنیم ولی مساله این است که این روش لزوماً یک گراف Causal Minimal به ما نمی دهد. برای حل این مشکل یک penalized independence score تعریف کرده و آن را برای گراف ها محاسبه می کنیم و این معیار را مبنای انتخاب گراف قرار می دهیم.

$$\hat{G} = \operatorname{argmin}_G \sum_{i=1}^n \operatorname{DM}(res_i^{G, \text{RM}}, res_{-i}^{G, \text{RM}}) + \lambda \# \text{edges}$$

در آن RM روش رگرسیون ما و DM یک معیار استقلال است.  $res_i$  مقدار باقی مانده ی رگرسیون  $X_i$  است وقتی آن را بر روی تمام parent هایش رگرس می کنیم و  $res_{-i}$  باقی مانده ی رگرسیون است

## مراجع

- [1] CHICKERING, D. M. A transformational characterization of equivalent bayesian network structures. in *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence* (1995), UAI'95, pp. 87–98.
- [2] HOYER, P. O., JANZING, D., MOOIJ, J. M., PETERS, J., AND SCHÖLKOPF, B. Nonlinear causal discovery with additive noise models. in *Advances in Neural Information Processing Systems* 21. 2009, pp. 689–696.
- [3] NOWZOHOUR, C., AND BÜHLMANN, P. Score-based causal learning in additive noise models. *Statistics* 50, 3 (2016), 471–485.
- [4] PETERS, J., MOOIJ, J. M., JANZING, D., AND SCHÖLKOPF, B. Causal discovery with continuous additive noise models. *J. Mach. Learn. Res.* 15, 1 (Jan. 2014), 2009–2053.
- [5] ZHANG, K., AND HYVÄRINEN, A. On the identifiability of the post-nonlinear causal model. in *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence* (Arlington, Virginia, United States, 2009), UAI '09, pp. 647–655.