

## توضیحات

تمرین‌های عملی را با زبان R پیاده‌سازی نمایید. توصیه می‌شود که از نرم‌افزار R-Studio استفاده کنید. به ازای هر بخش از تمرین موارد خواسته شده را در قالب یک گزارش ارائه کنید. این گزارش می‌تواند به صورت یک فایل متنی ضمیمه‌ی برنامه‌ها باشد یا به صورت یک jupyter-notebook ارائه شود.

## ۱ تولید و تصویر داده‌ی علی

در این بخش از شما می‌خواهیم که مطابق با مدل‌های ساختاری علی که در ادامه معرفی می‌شوند داده تولید کنید و این داده را از منظرهای مختلف تصویر کنید. به مدل‌های دو متغیره‌ی زیر توجه کنید:

$$\text{A. } X = N_X, Y = X^3 + N_Y \\ N_X \perp\!\!\!\perp N_Y$$

$$\text{B. } X = N_X, Y = 2X + N_Y \\ N_X \perp\!\!\!\perp N_Y$$

حال توزیع خطاهای  $N_X, N_Y$  را مستقل از یک دیگر و در شرایط زیر در نظر بگیرید:

(۱) مدل A با توزیع نرمال استاندارد برای هر دو متغیر خطا.

(۲) مدل B با توزیع نرمال استاندارد برای خطای علت و توزیع  $t$  با ۱ درجه آزادی برای خطای معلول.

(۳) مدل B با توزیع نرمال استاندارد برای خطای علت و توزیع  $t$  با ۵ درجه آزادی برای خطای معلول.

(۴) مدل B با توزیع نرمال استاندارد برای خطای علت و توزیع  $t$  با ۲۰ درجه آزادی برای خطای معلول.

برای هرکدام از این چهار مدل علی نمودارهای مربوط به موارد زیر را رسم کنید:

- $p_{X,Y}(x, y)$
- $p_{Y|X}(y|x)$
- $p_{X|Y}(x|y)$
- $p_Y^{do(X)}(y)$
- $p_X^{do(Y)}(x)$

نهایتاً تشخیص‌پذیری در هر حالت را مورد بررسی قرار دهید و دلایل خود را شرح دهید.

## ۲ اولین استنتاج علی

در این بخش از شما خواسته شده که روشی برای پیدا کردن جهت علی پیاده‌سازی کنید و آن را مورد آزمایش قرار دهید. ابتدا الگوریتم زیر را پیاده‌سازی کنید:

- (۱) با استفاده از یک روش رگرسیون غیر خطی مدلی به داده برازش کنید.
  - (۲) باقی مانده<sup>۱</sup> را به دست آورید.
  - (۳) استقلال باقی مانده از متغیر توضیح دهنده را با استفاده از تست آماری HSIC مورد بررسی قرار دهید و p-value را بدست آورید.
  - (۴) همین مراحل را در خلاف جهت هم انجام دهید.
  - (۵) جهت با p-value کوچکتر را به عنوان جهت پادعلی<sup>۲</sup> گزارش نمایید.
- ابتدا توضیح دهید که روش مذکور در چه حالتی درست عمل خواهد کرد و ارتباط این مراحل به مفروضات را شرح دهید. سپس همین الگوریتم را ۱۰ بار روی ۲۰۰ داده‌ی بدست آمده از هرکدام از مدل‌های تمرین ۱ اجرا کنید. نتایج را تفسیر کنید.

### ۱.۲ داده‌ی واقعی

در این بخش از شما خواسته می‌شود روش استنتاج خود را روی مجموعه داده‌ی توبینگن<sup>۳</sup> اجرا کنید. حال سعی کنید علت خطای الگوریتم خود در یکی از موارد خطا را توجیه کنید. راهنمایی: رجوع به نمودار داده.

---

<sup>۱</sup>Residual

<sup>۲</sup>Anti-Causal

<sup>۳</sup>Tuebingen Cause-Effect Pairs