دانشگاه تهران

پردیس دانشکده‌های فنی

دانشکده مهندسی برق و کامپیوتر

تمرین شماره: ۳

هوش مصنوعی قابل اعتماد

نام و نام‌خانوادگی: مرتضی ملکی‌نژاد شوشتری

شماره دانشجویی: ۸۱۰۱۰۴۲۵۶

نیم‌سال دوم سال تحصیلی ۴۰۵-۴۰۴

فهرست مطالب

۱ شبکه‌های بیزی ۵
۱.۱ بخش اول ۶
۱.۱.۱ زیربخش اول ۶
۲.۱.۱ زیربخش دوم ۷
۳.۱.۱ زیربخش سوم ۷
۴.۱.۱ زیربخش چهارم ۷
۲.۱ بخش دوم ۸
۳.۱ بخش سوم ۹
۱.۳.۱ زیربخش اول ۹
۲.۳.۱ زیربخش دوم ۹
۳.۳.۱ زیربخش سوم ۹
۴.۱ بخش چهارم ۱۰
۱.۴.۱ زیربخش اول ۱۰
۲.۴.۱ زیربخش دوم ۱۰

۲ علیت ۱۱
۱.۲ سوالات تئوری ۱۱
۱.۱.۲ نامساوی ۱۱
۲.۱.۲ اثر علّی ۱۲
۲.۲ سوالات پیاده‌سازی ۱۳
۱.۲.۲ پیاده‌سازی SCM ۱۳

۲.۲.۲ تاثیر دوره آموزشی ۱۳
۳.۲.۲ سیگار نکشیم ۱۵
۴.۲.۲ قالب جدید وبسایت ۱۷

فهرست تصاویر

۱.۱ شبکه بیزی آلودگی هوا ۶
۲.۱ گراف معادل گراف سمت راست سوال ۱۰

۱.۲ مدل علّی ساختاری توصیف شده در سوال ۱۲
۲.۲ گراف علّی مدل سوال دوره آموزشی ۱۴
۳.۲ گراف علّی مدل سوال دوره آموزشی ۱۵
۴.۲ توزیع سن بین دو گروه ۱۸
۵.۲ نمودار ویالنی زمان صرف شده به تفکیک حالت و سن ۱۸

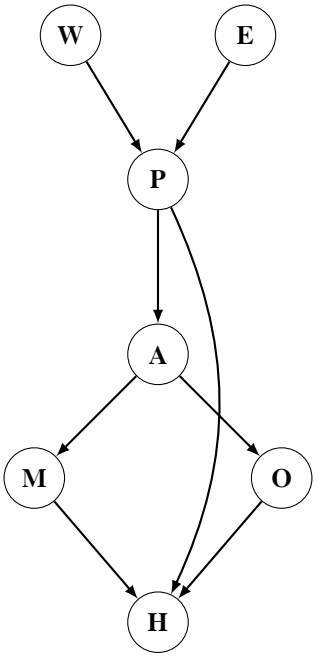
فهرست جداول

سوال ۱

شبکه‌های بیزی

۱.۱ بخش اول

۱.۱.۱ زیربخش اول



شکل ۱.۱: شبکه بیزی آلودگی هوا

۲.۱.۱ زیربخش دوم

با توجه به شبکه رسم شده در شکل ۱.۱ داریم:

$$P(X_1, X_2, X_3, ...) = \prod_{1}^{n} P(X_i|parent_i)$$
$$\Rightarrow P(W, E, P, A, M, O, H) = P(W)P(E)P(P|W, E)P(A|P)P(M|A)P(O|A)P(H|M, O, P)$$

۳.۱.۱ زیربخش سوم

سوال Markov Blanket مربوط به متغیر M را می‌خواهد. پس داریم:

$$MB(M) = Parents(M) + Children(M) + Spouse(M)$$
$$Parents(M) = A \quad Children(M) = H \quad Spouse(M) = P, O$$
$$\Rightarrow MB(M) = A, H, P, D$$

۴.۱.۱ زیربخش چهارم

باید درستی و یا نادرستی عبارات را تعیین کرد.

۱. $A \perp W|P$

بله درست است. در شبکه بیزی، A و W تنها از طریق P باهم رابطه دارند. پس اگر مقدار P مشخص باشد، این دو از هم مستقل خواهند بود.

۲. $M \perp O|A$

درست است. چون A تنها والد مشترک آن‌ها است، با مشخص بودن A دیگر متغیری وجود ندارد که بتواند هردوی متغیرهای M و O را به‌هم وابسته کند. و خود متغیرها نیز با همدیگر رابطه‌ای ندارند.

۳. $A \perp H|\{M, O\}$

غلط است. زیرا A و H والد مشترک P را نیز همچنان دارند و این مورد روی هردو آن‌ها تاثیر می‌گذارد که یعنی حتی با معلوم بودن M و O نیز، این دو متغیر از هم مستقل نیستند.

۴. $M \perp E|A$ درست است. مسیر دیگری بین دو متغیر E و M نیست که از A گذر نکند.

۵. $O \perp W|\{A, H\}$ غلط است. زیرا با مشخص بودن H عملا رابطه هم‌بستگی بین والدهای آن بوجود می‌آید و O والد مستقیم و W والد غیر مستقیم آن است که مسیر رابطه بین این‌ها از A عبور نمی‌کند.

۲.۱ بخش دوم

باید مقایسه کرد که در هر مورد، کدام بزرگ‌تر است.

۱. $P(w^1) \qquad p(w^1|f^1)$

مقدار $p(w^1)$ بیشتر است. زیرا f^1 باعث کم شدن احتمال e^1 می‌شود و e^1 باعث زیاد شدن w^1 می‌شود. پس با رخ دادن f^1 احتمال w^1 کاهش می‌یابد.

۲. $P(f^1|e^1) \qquad p(f^1|e^1, h^1)$

برابرند. تنها مسیر ارتباط f با باقی از طریق e است و با مشخص بودن e دیگر مقدار سایر متغیرها تاثیری در احتمال f نخواهد داشت.

۳. $P(h^1|f^0, w^1) \qquad P(h^1|w^1)$

مقدار $P(h^1|w^1)$ بیشتر است. زیرا می‌توان دید که احتمال w با بالا رفتن احتمال h زیاد می‌شود. حال اگر مقدار w را داشته باشیم، اگر f منفی باشد، ممکن است f^0 باعث مثبت شدن w شود و برای همین به دلیل پدیده توضیح‌دهی، احتمال h^1 وقتی f^0 باشد کاهش می‌یابد.

۴. $p(e^1|w^1) \qquad p(e^1|w^1, c^1)$

مقدار $p(e^1|w^1, c^1)$ بیشتر است. زیرا c^1 باعث کم شدن احتمال d^1 می‌شود و این به دلیل پدیده توضیح‌دهی باعث زیاد شدن احتمال e^1 می‌شود.

۵. $p(w^1|d^1, e^1) \qquad p(w^1|d^1, e^1, h^1)$

برابرند. زیرا همه مسیرهای رسیدن به w در هر دو حالت کاملا مشخص شده‌اند.

۶. $p(t^1|w^0, e^1) \qquad p(t^1|w^0, e^1, f^1)$

برابرند. زیرا f نمی‌تواند بدون تغییر e اثری روی t بگذارد. پس با مشخص بودن e عملا f بی‌تاثیر خواهد بود.

۳.۱ بخش سوم

ما n متغیر داریم که هرکدام L مقدار مختلف می‌توانند داشته باشند.

۱.۳.۱ زیربخش اول

ابتدا به ازای n های کوچک حساب می‌کنیم:

$$n = 1 \Rightarrow L - 1$$
$$n = 2 \Rightarrow L^2 - 1$$
$$n = 3 \Rightarrow L^3 - 1$$

چون در هرمورد L^n متغیر داریم. ولی چون جمع آن‌ها یک است، پس یکی از آن‌ها وابسته است و $L^n - 1$ متغیر مستقل داریم.

۲.۳.۱ زیربخش دوم

باید تعداد حالات مستقل خود گره را در تعداد حالات والدهای آن ضرب کرد. در بخش قبل حساب کردیم که n متغیر، $L^n - 1$ حالت مستقل دارند. البته در اینجا چون جمع این احتمالات ثابت نیست، منهای یک نداریم. خود گره نیز $L - 1$ حالت دارد) پس درنهایت یک گره حداکثر $(L^k)(L - 1)$ پارامتر دارد. حال باید این عدد را در تعداد گره‌ها ضرب کرد پس درنهایت داریم:

$$\text{number of parameters} = \text{number of nodes} * \text{each node parent parametrs} * \text{each node number of states} = n * (L^k) * (L - 1)$$

۳.۳.۱ زیربخش سوم

ایده شبیه بخش قبل است و فقط شیوه محاسبه فرق می‌کند. پس داریم:

$$\text{number of parameters} = \sum_{i=0}^{n} param(X_i) + param(C) = \sum_{i=0}^{n} states(parent(X_i)) states(X_i - 1) + (m-1) = m - 1 + m*(L-1) + m*L*(L-1) + m*l*(L-1) + ... = (m-1) + (m*(L-1)) + (n-1)(m*L*(L-1)) = m - 1 + ml - m + ((n-1)(mL^2 - mL)) =$$

$$ml - 1 + nmL^2 - nmL - mL^2 + mL = 2 - nml - 1 + nmL^2 - nmL - mL^2$$

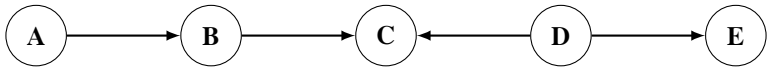
۴.۱ بخش چهارم

۱.۴.۱ زیربخش اول

در هرمورد باید دید آیا یالی وجود دارد که با برعکس کردن آن هیچ v structure ای حذف نشود و هیچ v structure جدیدی بوجود نیاید.

در گراف سمت چپ، با تغییر هر یک از یال‌های A–B و C–B باعث می‌شود V structure مربوط به A–B–C خراب شود. از طرف دیگر تغییر جهت یال‌های B–D و D–E منجر به ایجاد V structure های جدید می‌شود. پس هیچ گراف دیگری با گراف داده شده I equivalent نیست.

در گراف سمت راست، با تغییر جهت یال E–D هیچ v strucutre ای حذف یا تولید نمی‌شود. پس این گراف با گراف سوال I equivalent است. شکل این گراف در تصویر ۲.۱ رسم شده است.



شکل ۲.۱: گراف معادل گراف سمت راست سوال

۲.۴.۱ زیربخش دوم

دو شرط آمده در سوال بیان می‌کنند که A و B نسبت به هم مستقلند و همین‌طور هر دو والد مشترک C نیستند. این دو شرط در کنارهم باعث می‌شود تا گراف همبند نشود. زیرا C نمی‌تواند هم به A و هم به B وصل باشد (فارغ از جهت) و A و B نیز رابطه‌ای باهم ندارند. پس A و B درون دو مولفه جدا از هم قرار می‌گیرند. ولی C در هرکدام از این مولفه‌ها باشد، به این معنی است که از متغیر دیگر مستقل است و این استقلال در صورت سوال نیامده است. پس نمی‌توان شبکه‌ای یافت که perfect map این توزیع باشد.

سوال ۲

علیت

۱.۲ سوالات تئوری

۱.۱.۲ نامساوی

کافیست یک مثال نقض بزنیم. حالت زیر را درنظر بگیرید:

$$X \sim Bernouli(0.5) Y = XZ = X(xor)Y$$

حال اگر پرسش احتمالی را برای $Y = 1|X$ حساب کنیم داریم:

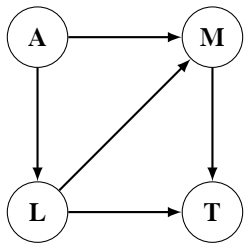
$$P(Z = 1) = 0$$
$$P(Z = 1|X = 1) = 0$$
$$P(Z = 1|X = 0) = 0 \Rightarrow min(P(Y = 1|X = x)) \leq P(Y = 1) \leq max(P(Y = 1|x = x'))$$

که در آن نامساوی درست است. حال اگر با مداخله حساب کنیم داریم:

$$P(Z=1|do(Y=0)) = P(X=1) = 0.5$$
$$P(Z=1|do(Y=1)) = P(X=0) = 0.5$$

که میبینیم هر دو احتمال با مداخله برابر 0.5 می‌شود ولی مقدار $P(Z=1)$ برابر صفر است که از هردو این مقادیر کمتر است.

۲.۱.۲ اثر علّی



شکل ۱.۲: مدل علّی ساختاری توصیف شده در سوال

باید مقدار $E[M \mid L, A]$ را در نظر گرفت، زیرا هدف ما محاسبه‌ی اثر علّی بار CPU بر مصرف حافظه است. اگر تنها از $E[M \mid L]$ استفاده کنیم، آنگاه اثر متغیر (A) نیز در این مقدار وارد می‌شود. دلیل این موضوع آن است که (A) والد مشترک (L) و (M) است؛ بنابراین هم بر مقدار بار CPU و هم بر مصرف حافظه تأثیر می‌گذارد. در نتیجه بخشی از وابستگی مشاهده‌شده بین (L) و (M) ناشی از مسیر $L \leftarrow A \rightarrow M$ است و لزوماً بیانگر اثر علّی (L) بر (M) نیست.

برای حذف این اثر م باید A را معین کنیم. در این صورت مسیر پشتی بین L و M مسدود می‌شود و می‌توان اثر علّی L بر M را محاسبه کرد. به طور دقیق، اثر علّی از طریق $P(M \mid do(L=l))$ تعریف می‌شود و در این مسئله با تعدیل روی A قابل محاسبه است. بنابراین برای محاسبه‌ی اثر علّی یک متغیر بر متغیر دیگر، باید ابتدا تمام علل مشترک آن‌ها را کنترل کرد و سپس اثر مداخله را بررسی نمود.

۲.۲ سوالات پیاده‌سازی

۱.۲.۲ پیاده‌سازی SCM

مطابق صورت سوال کلاس SCM پیاده‌سازی شد. ساختار این کلاس به این صورت است که موقع اضافه شدن یک متغیر، یک تابع نیز داده می‌شود که با ورودی گرفتن والدهای متغیر و تعداد داده مورد نیاز، برای آن متغیر خروجی تولید کند.

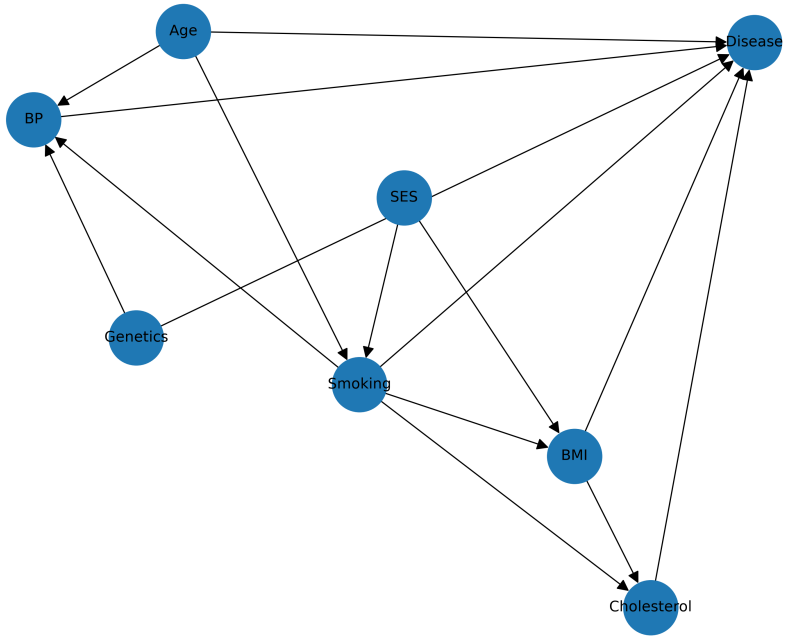
موقع نمونه‌برداری، مدل ابتدا متغیرها را به صورت توپولوژی مرتب می‌کند و سپس از گره اول شروع به ایجاد نمونه می‌کند.

هنگام مداخله نیز، مدل همه متغیرها را عینا منتقل می‌کند ولی متغیری که برای آن مداخله ثبت شده را تابع تولید آن را تغییر می‌دهد تا مقدار ثابت خروجی بدهد و والدی نیز برای آن در نظر نمی‌گیرد.

۲.۲.۲ تاثیر دوره آموزشی

موارد این سوال در نوتبوک subsec2 پیاده‌سازی شده‌اند.

۱. گراف علّی در تصویر ۲.۲ آمده است.



شکل ۲.۲: گراف علّی مدل سوال دوره آموزشی

۲. در تکه کد پایین ،کد محاسبه و خروجی آن نمایش داده شده‌اند.

```
samples = s.sample(10_000)
e1 = samples[samples["x"] == 1]["y"].mean()
e0 = samples[samples["x"] == 0]["y"].mean()
print(e1 - e0)
```

نمونه کد ۱: مقدار $E[Y|X]$

```
6.22758964557649
```

۳. در تکه کد پایین ،کد محاسبه و خروجی آن نمایش داده شده‌اند.

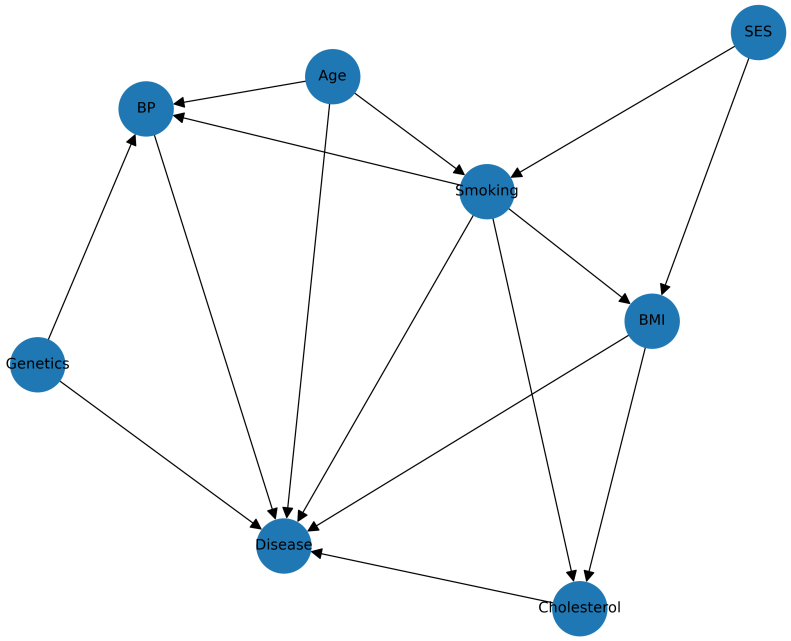
```
interv_scm = s.do_operator({"x": 1})
interv_samples = interv_scm.sample(10_000)
e1 = interv_samples["y"].mean()
interv_scm = s.do_operator({"x": 0})
interv_samples = interv_scm.sample(10_000)
e0 = interv_samples["y"].mean()
print(e1 - e0)
```

نمونه کد ۲: مقدار $E[Y|do(X)]$

```
2.9780376353445415
```

۴. این سوال به نحوی همان سوال آخر قسمت تئوری است. از آنجایی که دو متغیری که می‌خواهیم رابطه علّی بین آن‌ها را بسنجیم والد مشترک دارند، اگر فقط مقدار $E[Y|X]$ را بسنجیم، آنگاه اثر متغیر u نیز دخیل می‌شود. این تفاوت نیز به این دلیل است.

۳.۲.۲ سیگار نکشیم



شکل ۳.۲: گراف علّی مدل سوال دوره آموزشی

۱. ATE

```
interv_scm = s.do_operator({"Smoking": 1})
interv_samples = interv_scm.sample(10_000)
e1 = interv_samples["Disease"].mean()
interv_scm = s.do_operator({"Smoking": 0})
interv_samples = interv_scm.sample(10_000)
e0 = interv_samples["Disease"].mean()
print("ATE",e1 - e0)
```

نمونه کد ۳: مقدار ATE

```
ATE 0.09618307044069563
```

۲. ATT

```
samples = s.sample(10_000)
samples = samples[samples["Smoking"] == 1]

e1 = samples["Disease"].mean()
samples["Smoking"] = 0
samples["Disease"] = eq_disease(samples, len(samples))
e0 = samples["Disease"].mean()
print("ATT", e1 - e0)
```

نمونه کد ۴: مقدار ATT

```
ATT 0.06294287833251536
```

این مقدار از بخش قبل کمتر است. یعنی اگر یک فرد سیگاری از سیگار کشیدن دست بکشد، همچنان احتمال بروز بیماری در او به میزان قبل از سیگار کشیدن بر نمی‌گردد.

۳. CATE

```
samples = s.sample(10_000)
samples1 = samples[samples["Age"] > 55]
samples2 = samples[samples["Age"] <= 55]
for samples in (samples1, samples2):
samples["Smoking"] = 1
samples["Disease"] = eq_disease(samples, len(samples))
e1 = samples["Disease"].mean()
samples["Smoking"] = 0
samples["Disease"] = eq_disease(samples, len(samples))
e0 = samples["Disease"].mean()
print("CATE", e1 - e0)
```

نمونه کد ۵: مقدار CATE

```
CATE 0.06257998919727491
CATE 0.042922818228705154
```

همانطور که می‌توان دید، سیگار کشیدن در افراد مسن حدودا ۵۰ درصد تاثیر بیشتری دارد.

۴.۲.۲ قالب جدید وبسایت

حالت ساده‌لوحانه

```
for layout in df["layout"].unique():
print(layout, df[df['layout'] == layout]['time_spent'].mean())
```

نمونه کد ۶: محاسبه میانگین زمان صرف‌شده در هر حالت سایت

```
Old_Layout 12.501732101616629
New_Layout 9.730546737213404
```

با محاسبه مدل OLS می‌توان دید که حالت جدید سایت ضریب منفی دارد. اما این موضوع لزوما علت را نشان نمی‌دهد. چون داده ما بایاس دارد و افراد مسن حالت جدید را می‌بینند، این موضوع می‌تواند ناشی از این باشد که بصورت کلی افراد مسن زمان کم‌تری را در سایت می‌گذرانند. و بهتر است حالت قبل و بعد یک فرد را مقایسه کرد. [۱]

یک فرد ناشی ممکن است از این ضریب به این نتیجه برسد که حالت جدید سایت اثر منفی دارد و تصمیم به برگرداندن حالت سایت به قبل بکند.

```
X_layout = (df["layout"] == "New_Layout").astype(float).values
X = np.column_stack([np.ones(len(df)), X_layout])
y = df["time_spent"].values

beta, *_ = np.linalg.lstsq(X, y, rcond=None)
intercept, coef_layout = beta

print("coef", coef_layout)
```

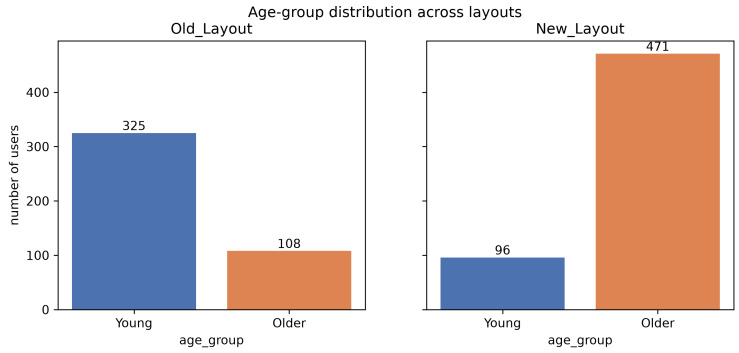
نمونه کد ۷: محاسبه میانگین زمان صرف‌شده در هر حالت سایت

```
-2.7711853644032156
```

[۱] برای محاسبه OLS با استفاده از numpy برای سینتکس از هوش مصنوعی استفاده شد.

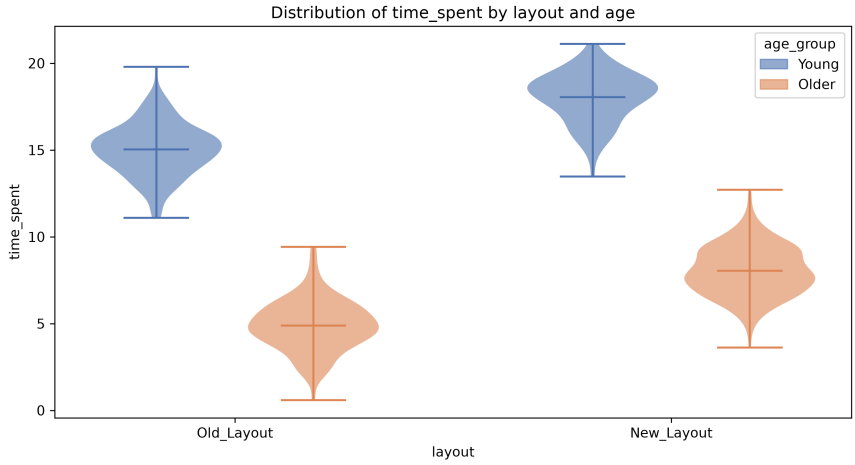
تشخیص بصری

همانطور که در شکل ۴.۲ می‌توان دید، بین دو گروه نابرابری وجود دارد. این موضوع به ما می‌گوید که اختلاف زمان بین دو گروه می‌تواند ناشی از بایاس موجود در سن دو گروه باشد و نه به دلیل تاثیر علّی چیدمان.



شکل ۴.۲: توزیع سن بین دو گروه

تصویر ۵.۲ تاثیر را به خوبی نشان می‌دهد. در هر رده سنی،حالت جدید سایت نسبت به حالت قدیم میانگین زمان بیشتری داشته است. ولی چون به طور کلی افراد مسن‌تر زمان کم‌تری را در سایت صرف می‌کنند و افراد مسن نیز بیشتر حالت جدید سایت را می‌بینند، در بخش‌های قبل این‌طور به نظر می‌رسید که حالت جدید اثر منفی داشته است.



شکل ۵.۲: نمودار ویالنی زمان صرف شده به تفکیک حالت و سن

میانگین کل جامعه، از آنجایی که سعی در خلاصه‌سازی کل توزیع در یک عدد دارد، ناگزیر بسیار نادقیق است و روابط علّی و معلولی را نیز ندارد. برای مثال در همین سوال، یک ویژگی اثر مثبت داشته است ولی چون این ویژگی در گروهی بیشتر بوده که بصورت کلی زمان کم‌تری را در سایت می‌گذراندند، اینطور به نظر می‌آید که این ویژگی باعث بدتر شدن زمان می‌شود.