



به نام خدا



دانشگاه تهران
دانشکده مهندسی برق و کامپیوتر
هوش مصنوعی قابل اعتماد

تمرین سوم

نام و نام خانوادگی	مهدی موسوی
شماره دانشجویی	810102264
تاریخ ارسال گزارش	خرداد 1403

فهرست

3	سؤال اول
3	الف
4	ب
5	سؤال دوم
7	سؤال سوم
10	سؤال چهارم
11	سؤال پنجم
12	سؤال ششم

سؤال اول

الف

برای حل قسمت اول سوال داریم:

$$P_x(Y = 1 | A = O) = \sum_s P_x(Y = 1 | A = O, S = s) \cdot P_x(S = s | A = O)$$

در عبارت بالا مقادیر $P_x(Y = 1 | A = O, S = s)$ را داریم و لازم است که

$P_x(S = s | A = O)$ را حساب کنیم. برای محاسبه خواهیم داشت:

$$P_x(A = O) = P_x(A = O | S = L) \cdot P_x(S = L) + P_x(A = O | S = R) \cdot P_x(S = R)$$

$$P_x(A = O) = (0.23 \cdot 0.49) + (0.76 \cdot 0.51) = 0.5003$$

$$P_x(S = L | A = O) = \frac{P_x(A = O | S = L) \cdot P_x(S = L)}{P_x(A = O)} = 0.2252$$

$$P_x(S = R | A = O) = \frac{P_x(A = O | S = R) \cdot P_x(S = R)}{P_x(A = O)} = 0.7748$$

در نهایت با جایگذاری داریم :

$$P_x(Y = 1 | A = O) = (0.69 \cdot 0.2252) + (0.87 \cdot 0.7748) = 0.829$$

همچنین برای احتمال بعدی با تکرار روند داریم :

$$P_x(A = N) = 1 - P_x(A = O) = 0.4997$$

$$P_x(S = L | A = N) = \frac{P_x(A = N | S = L) \cdot P_x(S = L)}{P_x(A = N)} = 0.7553$$

$$P_x(S = R | A = N) = 1 - P_x(S = L | A = N) = 0.2447$$

به طور مشابه با جایگذاری داریم:

$$P_x(Y = 1 | A = N) = (0.73 \cdot 0.7553) + (0.93 \cdot 0.2447) = 0.779$$

ب

برای محاسبه مقادیر دیگر اشاره شده خواهیم داشت :

$$P_x(Y = 1|do(A = 0)) = P_x(Y = 1|A = 0, S = R)P(S = R) \\ + P_x(Y = 1|A = 0, S = L)P(S = L) = 78.18$$

به همین ترتیب

$$P_x(Y = 1|do(A = N)) = P_x(Y = 1|A = N, S = R)P(S = R) \\ + P_x(Y = 1|A = N, S = L)P(S = L) = 83.2$$

سوال دوم

ابتدا برای فرد A مرحله abduction را انجام می‌دهیم:

$$2500 = 0.3 * 75000 + u_2 \rightarrow u_2 = 2500$$

در مرحله action ابتدا خروجی طبقه بند را برای مقادیر در دست محاسبه می‌کنیم:

$$\text{sgn}(x_1 + 5x_2 - 225000) = \text{sgn}(-25000) = -1$$

حال خواهیم داشت:

$$x_1 + \delta_1 + 5 * (0.3 * (x_1 + \delta_1) + 2500) = 225000$$

$$75000 + \delta_1 + 5 * (0.3 * (75000 + \delta_1) + 2500) = 225000$$

$$2.5\delta_1 = 25000 \rightarrow \delta_1 = 10000$$

به همین صورت برای δ_2 خواهیم داشت :

$$x_1 + 5 * (x_2 + \delta_2) = 225000$$

$$75000 + 5 * (25000 + \delta_2) = 225000$$

$$5\delta_2 = 25000 \rightarrow \delta_2 = 5000$$

برای محاسبه هزینه خواهیم داشت :

$$\text{cost}_1 = \frac{10000}{100000} = 0.1, \text{cost}_2 = \frac{5000}{32500} = 0.15$$

پس تغییر مناسب تغییر x_1 است و به همین ترتیب کل هزینه تغییر برابر با

$$\text{cost}_1 + \text{cost}_2 = 0.1 + \frac{30}{325} = 0.19$$

است.

برای فرد B به همین ترتیب خواهیم داشت:

$$23800 = 0.3 * 70000 + u_2 \rightarrow u_2 = 2800$$

$$x_1 + \delta_1 + 5 * (0.3 * (x_1 + \delta_1) + 2800) = 225000$$

$$70000 + \delta_1 + 5 * (0.3 * (70000 + \delta_1) + 2800) = 225000$$

$$2.5\delta_1 = 36000 \rightarrow \delta_1 = 14400$$

و به طور مشابه

$$x_1 + 5 * (x_2 + \delta_2) = 225000$$

$$70000 + 5 * (23800 + \delta_2) = 225000$$

$$5\delta_1 = 36000 \rightarrow \delta_1 = 7200$$

در نهایت برای محاسبه هزینه داریم:

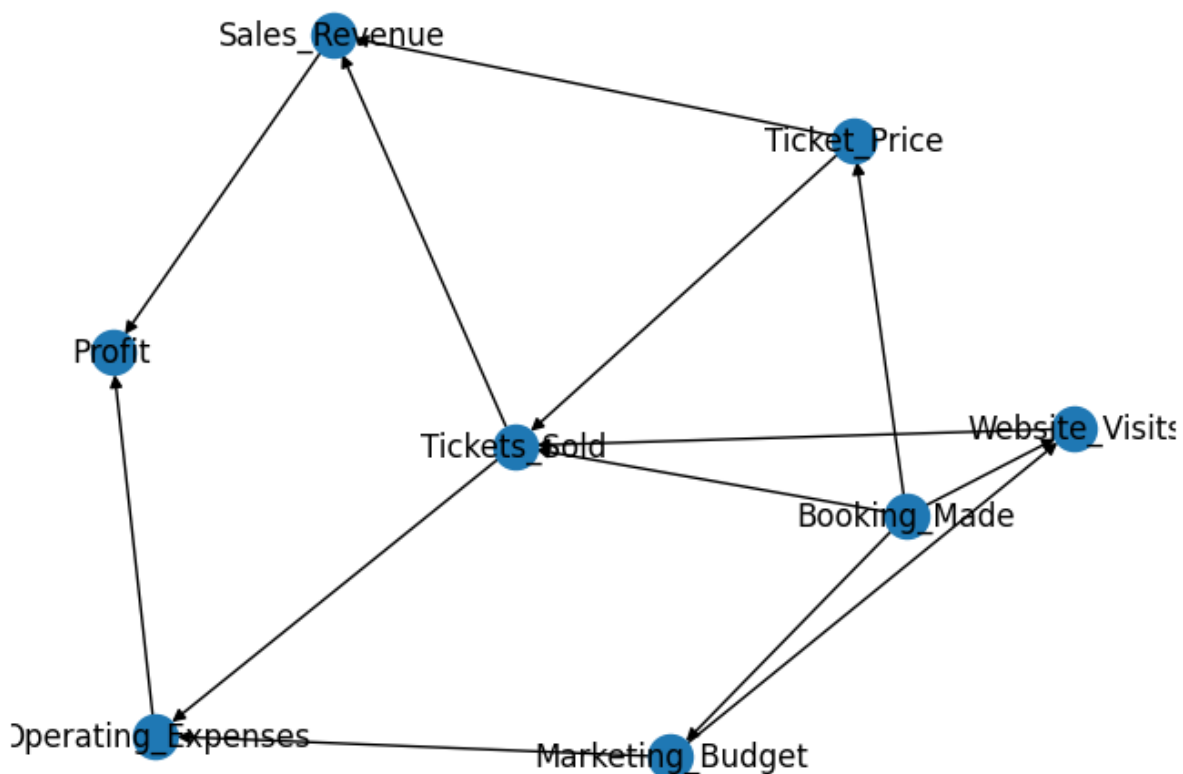
$$cost_1 = \frac{14400}{100000} = 0.144, cost_2 = \frac{7200}{32500} = 0.221$$

و برای هزینه کل:

$$\frac{14400}{100000} + \frac{4320}{32500} = 0.177$$

سوال سوم

در اولین قسمت از این سوال به رسم نمودار وابستگی ویژگی های میپردازیم که در نمودار ۱-۳ دیده میشود. مشاهده میشود که متغیر نوع رزرو به هیچ متغیر دیگری وابسته نیست و متغیر سود چون متغیر نهایی است بر هیچ متغیر دیگری تاثیر نمیگذارد.



در قسمت بعدی کاری که انجام میدهیم به ازای هر معلول و عامل هایش مدل رگرسیون خطی آموزش می دهیم که معادله های به دست آمده مطابق زیر است:

$$\text{Marketing Budget} = 1368.01 * \text{Booking Made} + 1298.81$$

$$\text{Website Visits} = 7699.75 * \text{Booking Made} + 0.5 * \text{Marketing Budget} + 11090.27$$

$$\begin{aligned} \text{Tickets Sold} = & 764.69 * \text{Booking Made} + 0.27 * \text{Website Visits} \\ & - 5.02 * \text{Ticket Price} + 4206.54 \end{aligned}$$

$$\text{Ticket Price} = -92.02 * \text{Booking Made} + 999.82$$

$$\text{Sales Revenue} = 6118.32 * \text{Ticket Price} + 1001.81 * \text{Tickets Sold} - 6121790.91$$

$$\text{Operating Expenses} = \text{Marketing Budget} + 500 * \text{Tickets Sold} + 500011.10$$

$$\text{Profit} = \text{Sales Revenue} - \text{Operating Expenses}$$

به همین ترتیب سعی میکنیم که با توزیع نرمال باقیمانده ها را تخمین بزنیم که توزیع برای انحراف معیارها مطابق جدول 1-3 است.

جدول 1-3: مقادیر انحراف معیار برای متغیرهای تخمین زده شده در توزیع نرمال

متغیر	انحراف از معیار
Budget Marketing	120.96
Visits Website	320.59
Tickets_Sold	100.64
Ticket_Price	10.77
Sales_Revenue	11858.66
Operating_Expenses	6.08
Profit	0

در ادامه کار سعی میکنیم که تاثیر دو عامل را بر سود کلی ببینیم. واریانس کلی سود برابر 62277529945 است که حدود 75 درصد از این مقدار مربوط به فروش و 25 درصد مربوط به مخارج عملیاتی است که منطقی است چون اکثر عدم قطعیت در قسمت فروش است و معمولاً در قسمت تولید عدم قطعیت کمتری داریم.

در قسمت بعدی سعی داریم که ببینیم که به ازای هر کدام از ویژگیها چه میزان تاثیر روی سود نهایی داریم. برای این کار برای هر متغیری که آن را با بیشتر از یک متغیر دیگر تخمین زده ایم میزان مشارکت حساب میکنیم و در نهایت از پایین به سمت سود شروع به جمع زدن میکنیم و در مرحله ضرب در میزان تاثیر در متغیر هدف آن مرحله میکنیم. در نهایت به جدول 2-3 خواهیم رسید.

جدول 2-3: مقادیر تاثیر هر متغیر در سود

متغیر	تاثیر در واریانس خروجی
-------	------------------------

Budget Marketing	1
Visits Website	82.56
Tickets_Sold	96
Ticket_Price	6.83
Sales_Revenue	74.05
Operating_Expenses	25.95
Booking made	97.74

برای محاسبه مقادیر بالا به ازای هر ویژگی تاثیر تمام ویژگی های والد در واریانس آن مشاهده شد سپس برای هر ویژگی همه مسیر ها تا متغیر سود را محاسبه و با هم جمع کردیم. همانطور که مشاهده میشود متغیر های مربوط به خدمات دهنده مثل بودجه تبلیغاتی و قیمت بلیط تاثیر محدودی بر تغییرات داشته اند در حالی که متغیر های مربوط به مشتری مثل متغیر باینری انجام شدن رزرو سهم خیلی زیادی از واریانس را دارند.

در نهایت در آخرین قسمت برای اینکه متوجه شویم که آیا در روز مطرح شده به نسبت پارسال افت فروش داریم یا خیر سعی میکنیم که داده جدید را در مدل سال قبل قرار دهیم و ببینیم که آیا با مدل سال قبل انتظار سود بیشتر داریم یا کمتر.

بعد از برازش دقیقاً به عدد داده شده در صورت سوال رسیدیم که به این معناست که مقدار فروش مطابق انتظار بوده است.

برای مقایسه با دقیقاً روز اول سال قبل اما میتوان دید که سود به شکل چشمگیری افزایش یافته است که مهمترین دلیل آن این است که برخلاف پارسال در روز اول امسال امکان رزرو پرواز وجود دارد و به همین دلیل مقدار سود به شکل چشمگیری افزایش یافته است. این مورد موارد پیدا شده در قسمت قبل برای اینکه بیشترین تاثیر را امکان رزرو میگذارد را تایید میکند زیرا این متغیر بر ۴ متغیر دیگر تاثیر میگذارد و از لحاظ منطقی هم بدون امکان رزرو سود یک شرکت هواپیمایی بسیار پایین می آید.

همینطور میتوان دید که تعداد بلیط های به فروش رفته بسیار بیشتر است که همانطور که اشاره شد از نتایج این است که امکان رزرو فراهم بوده است و این تعداد زیاد قیمت کم شده آن را جبران کرده است.

سوال چهارم

در این قسمت قصد داریم مقادیر امید ریاضی خواسته شده را به دست بیاوریم که میدانیم این مقادیر در حقیقت مقدار حاصل از رگرسیون خطی به کمک مقادیر مد نظر است. به عبارت دیگر اگر با مقادیر متناظر با سوال رگرسیون خطی انجام دهیم مقدار امید ریاضی در حقیقت مقدار خواسته شده خواهد بود.

در اولین قسمت عبارت مد نظر

$$E_{w,z}E_Y[y|t, w, z]$$

است که به شکل زیر میتوان بازنویسی کرد:

$$E_{w,z}[bias + b_1t + b_2w + b_3z]$$

که عبارت درون کروشه در حقیقت مدل رگرسیون ما است. در مرحله بعد با توجه به اینکه w و z متغیرهای ما هستند از آنها میانگین میگیریم تا در نهایت به عبارت زیر برسیم:

$$130.52 + 0.86 * t$$

در مرحله بعدی همین کار را بدون متغیر z انجام میدهیم یعنی اول

$$E_wE_Y[y|t, w]$$

را محاسبه میکنیم و سپس

$$E_w[bias + b_1t + b_2w]$$

که در نهایت خواهیم داشت :

$$129.84 + 1.054 * t$$

در نهایت در آخرین رابطه خواسته شده صرفاً از متغیر t برای پیش بینی استفاده میکنیم که خروجی آن به صورت زیر است:

$$bias + b_1t = 120.53 + 3.64 * t$$

حال میخواهیم که اثر علی انسولین بر روی گلوکز را مشاهده کنیم برای این منظور نیاز داریم که مسیر backdoor را ببندیم که در حقیقت تاثیر مشترکی است که از سن بر روی هر دوی آن ها دیده میشود. برای این کار نیاز داریم که محاسبات خود را با شرط کردن روی w انجام دهیم که در حقیقت همان رابطه

$$E_wE_Y[y|t, w]$$

است. به این صورت با ضرایب به دست آمده برای این قسمت دیده میشود که در حقیقت انسولین فارغ از سن هم روی گلوکز موثر است چون همانطور که اشاره شد در این رابطه مسیر w بسته شده است و به عبارت دیگر با ثابت بودن سن باز هم مقدار انسولین روی گلوکز تاثیر دارد.

سوال پنجم

در اولین بخش این سوال خواسته شده است که دو ویژگی از چهار ویژگی را مورد کار قرار دهیم و همچنین برای سه ویژگی نرمال سازی انجام دهیم. به این منظور در لیست مورد کار ایندکس دو ویژگی مورد نظر را قرار می‌دهیم و به همین ترتیب سه ویژگی خواسته شده را هم نرمال می‌کنیم.

در بخش بعدی علاقه داریم که بدانیم که به ازای چه مقدار تغییر می‌توانیم خروجی را عوض کنیم که مشاهده میشود به طور میانگین 1.132 واحد برای تغییر علامت نیاز به هزینه داریم.

بعد از موارد ذکر به فایل *scm* میرویم و دو ستون خواسته شده را به عنوان ستون های قابل تغییر مشخص می‌کنیم و همه ستون ها را هم برای *soft intervention* قرار می‌دهیم.

در مرحله بعدی نیاز داریم که ماتریس جاکوبین را کامل کنیم که برای این منظور به ازای هر مسیر بین نود مبدا و نود مقصد ضرایب را محاسبه و در هم ضرب می‌کنیم تا تاثیر هر متغیر در متغیر تحت تاثیر دیده شود. به عنوان مثال از متغیر سن به متغیر فشار خون سه مسیر وجود دارد که در سطر و ستون مربوطه جای گذاری شده است. به همین ترتیب ضریب هر متغیر برای ایندکس خودش برابر ۱ است.

در نهایت بعد از انجام موارد گفته شده و تاثیر دادن تکه کد مورد نظر در حالت جدید متوجه می‌شویم که میانگین مقدار هزینه به 0.898 کاهش یافته است. این مورد به این معنا است که توانسته ایم با وارد کردن روابط علی تغییرات بهینه تری برای تغییر علامت داشته باشیم که در مجموع به کاهش میانگین هزینه انجامید.

در نهایت در آخرین قسمت برای اولین نمونه از دادگان ارزیابی هزینه و همچنین ویژگی های تغییر یافته را چاپ می‌کنیم. با مشاهده این دو متغیر دیده میشود که برای این داده به خصوص با روش اولیه هزینه تغییرات برابر 0.856 بود و همچنین هر دو متغیر گلوکز و انسولین تغییر کردند ولی بعد از اعمال الگوریتم مد نظر هزینه به مقدار 0.679 کاهش می یابد و همچنین در حالت جدید تنها انسولین تغییر میکند. به عبارت دیگر مدل متوجه تاثیر انسولین روی گلوکز شده است و سعی میکند فقط این متغیر را تغییر دهد و با تغییر به اندازه این متغیر کلاس را عوض کند.

سوال ششم

در اولین قسمت از این سوال به این موضوع خواهیم پرداخت که با چه شرایطی مقاومت تضمین میشود. برای این منظور بیان میشود که نیاز داریم که طبقه بند مقاوم باشد اما این شرط لازم است و نه کافی. در ادامه میتوان گفت که نیاز داریم که ساختار scm را به درستی استخراج کرده باشیم و همچنین روی متغیر های تغییرات انجام دهیم که از قبل تعیین شده اند قابل تغییر هستند. به همین ترتیب نیاز داریم که رفتار مدل طبقه بند قابل پیش بینی باشد تا بتوانیم که نمونه های تخصصی منطقی تولید کنیم.

در 4 preposition معادلات مد نظر برای مساله در دست آمده است که نشان میدهد در نهایت به دنبال این هستیم که به ازای کمترین مقدار تغییرات نمونه تخصصی را پیدا کنیم. در حقیقت هر دو قسمت 4 و 5 به دنبال روابطی برای پیدا کردن نزدیکترین نمونه تخصصی با علامت متفاوت هستند. همانطور که مشخص است در حقیقت قسمت 5 نسخه تعمیم یافته قسمت 4 است که از یک طبقه بند خطی به طبقه بند های مشتق پذیر تعمیم پیدا کرده است. در 5 preposition ما یک بهینه minmax داریم که در حقیقت هم مقدار لاندا و هم عمل تخصصی بهینه را پیدا میکند که همانطور که اشاره شد حالت عمومی تری از آنچه برای طبقه بند های خطی در 4 preposition دیدیم است.