یادگیری ماشین

نیم‌سال دوم ۱۳۹۹-۱۴۰۰

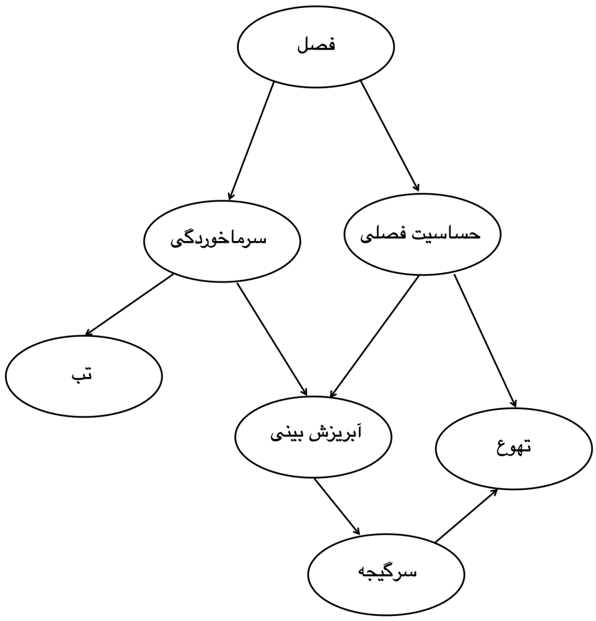
مدرس: دکتر عباس حسینی

دانشکده مهندسی کامپیوتر

تمرین سری پنجم

مدل‌های احتمالاتی گرافی | موعد تحویل: ۲۴ اردیبهشت

پرسش ۱



مدل گرافی داده شده را در نظر بگیرید.

الف. درستی یا نادرستی هرکدام از موارد زیر را با ذکر دلیل مشخص کنید. (۶ نمره)

- فصل و تهوع به شرط مشاهده‌ی حساسیت فصلی و آبریزش از هم مستقل اند.
- سرماخوردگی و حساسیت فصلی به شرط مشاهده‌ی فصل و تهوع از هم مستقل اند.
- فصل و تب به شرط مشاهده‌ی سرماخوردگی از هم مستقل اند.
- تب و تهوع به شرط مشاهده‌ی آبریزش مستقل اند.

ب. حال فرض کنید جدول احتمالات این گراف به صورت زیر است. با استفاده از این جداول به پرسش‌های گفته‌شده پاسخ دهید. (۵ نمره)

$P(S = \text{winter})$	$P(S = \text{summer})$
0.5	0.5

	$P(F = \text{true} \mid S)$	$P(F = \text{false} \mid S)$
$S = \text{winter}$	0.4	0.6
$S = \text{summer}$	0.1	0.9

	$P(D = \text{true} \mid S)$	$P(D = \text{false} \mid S)$
$S = \text{winter}$	0.1	0.9
$S = \text{summer}$	0.3	0.7

	$P(C = \text{true} \mid F)$	$P(C = \text{false} \mid F)$
$F = \text{true}$	0.8	0.2
$F = \text{false}$	0.1	0.9

	$P(H = \text{true} \mid F, D)$	$P(H = \text{false} \mid F, D)$
$F = \text{true}, D = \text{true}$	0.9	0.1
$F = \text{true}, D = \text{false}$	0.8	0.2
$F = \text{false}, D = \text{true}$	0.8	0.2
$F = \text{false}, D = \text{false}$	0.3	0.7

	$P(Z = \text{true} \mid H)$	$P(Z = \text{false} \mid H)$
$H = \text{true}$	0.8	0.2
$H = \text{false}$	0.2	0.8

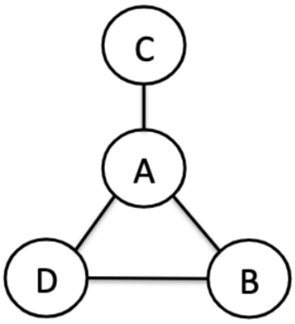
	$P(N = \text{true} \mid D, Z)$	$P(N = \text{false} \mid D, Z)$
$D = \text{true}, Z = \text{true}$	0.9	0.1
$D = \text{true}, Z = \text{false}$	0.8	0.2
$D = \text{false}, Z = \text{true}$	0.6	0.4
$D = \text{false}, Z = \text{false}$	0.2	0.8

که در اینجا منظور از S همان فصل، منظور F سرماخوردگی، منظور از D حساسیت، منظور از C تب، منظور از H آبریزش، منظور از N تهوع و منظور از Z سرگیجه است.

- فرض کنید زمستان است و آبریزش دارید، احتمال اینکه سرماخوردگی داشته باشید چقدر است؟
- حال فرض کنید علاوه بر شرایط قبل حساسیت فصلی نیز داشته باشید، حال احتمال اینکه سرما خوردگی داشته باشید چقدر است؟
- حال بدون دانستن هیچ شرایطی از خودتان احتمال اینکه سرما خورده باشید چقدر است؟

پرسش ۲

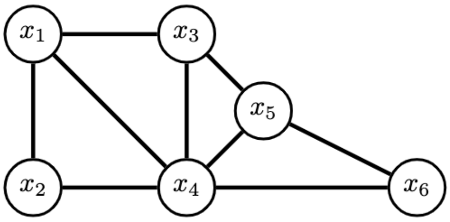
توزیع P حول چهار متغیر تصادفی A, B, C, D را درنظر بگیرید به صورتی که این توزیع به جایگشت $(0, 0, 0, 0)$ و $(1, 1, 0, 0)$ احتمال $\frac{1}{8}$ می‌دهد و به جایگشت‌های $(0, 1, 1, 1)$ و $(1, 0, 1, 0)$ و $(1, 1, 0, 1)$ هرکدام احتمال $\frac{1}{4}$ می‌دهد. اسکلت بندی شبکه‌ی بیزین این توزیع در شکل زیر آمده است اما جهت یال‌ها در آن مشخص نشده است.



الف. مشخص کنید کدام یک از دو شرط روبرو برقرار است $C \perp B$, $C \perp D$. (۱۰ نمره)
ب. مطابق احتمال‌های گفته شده به یال‌ها جهت دهید. آیا این جهت‌دهی که مشخص کردید یکتا است؟ (۵ نمره)

پرسش ۳

مدل گرافی زیر را درنظر بگیرید.



الف. توزیع گیبس حاصل از این گراف چه فرمی دارد؟ (۳نمره)
ب. آیا $x_3 \perp x_2 | x_4$ برقرار است؟ (۳ نمره)
پ. فرض کنید می‌خواهیم $E[x_1 x_2 x_5 | x_3, x_4]$ را تخمین بزنیم. مطابق با این گراف به کدام جدول‌های احتمالاتی نیاز داریم؟ جواب بهینه را گزارش کنید. (۳نمره)

پرسش ۴

فرض کنید d متغیر تصادفی داریم به صورتی که $p(x_1, ..., x_d) \propto \prod_{i \leq j} \phi_{ij}(x_i, x_j)$ برقرار است.
الف. اگر $p(x_1, ..., x_d) \propto \frac{-1}{2} x^T A x - b^T x$ باشد به صورتی که A یک ماتریس متقارن است، ϕ_{ij} را به دست آورید. (۸نمره)
ب. اگر $p(x_1, ..., x_d) \propto \frac{-1}{2} x^T A x - b^T x$ باشد نشان دهید اگر مولفه‌ی ij ماتریس A برابر 0 باشد، ادعای زیر برقرار است.

$$x_i \perp x_j \mid \{x_1, \ldots, x_d\} \setminus \{x_i, x_j\}$$

(۷نمره)

پرسش ۵

ماشین‌های محدود شده‌ی بولتزمن (RBM) یکی از معروف‌ترین و پراستفاده‌ترین مدل‌ها در یادگیری مخصوصا یادگیری عمیق هستند و معمولا در بخش‌های مدل‌های مولد به کار می‌روند. به صورت خیلی ساده یک RBM یک مدل گرافی غیرجهت دار است که دارای متغیرهای $(\mathbf{v}, \mathbf{h})\,,\;\; \mathbf{v} \in \{0,1\}^m \text{ and } \mathbf{h} \in \{0,1\}^n$ هستند. توزیع توام این متغیرها از رابطه‌ی $P(\mathbf{v}, \mathbf{h}) = \frac{1}{Z}\exp(\phi(\mathbf{v}, \mathbf{h}))$ پیروی می‌کند به صورتی که $\phi(\mathbf{v}, \mathbf{h}) = \alpha^T\mathbf{v} + \beta^T\mathbf{h} + \mathbf{v}^T W \mathbf{h}$ است. دقت کنید که هاRBM گراف دوبخشی کامل هستند به این معنا که دارای دو لایه هستند. لایه‌ی اول متغیرهای قابل رویت یعنی $\mathbf{v}$ و لایه‌ی دوم متغیرهای پنهان یعنی $\mathbf{h}$.

الف. رابطه‌ی $p\left(h_i = 1 \mid \mathbf{v}\right)$ را محاسبه کنید. (۱۰ نمره)

ب. استدلال کنید $p(\mathbf{h} \mid \mathbf{v}) = \prod_{i=1}^{n} p\left(h_i \mid \mathbf{v}\right)$ برقرار است یا نه. (۵ نمره)

پرسش ۶

سوال عملی

در این سوال می‌خواهیم یکی از اصلی‌ترین کاربردهای مدل‌های گرافی را در تئوری اطلاعات بررسی کنیم. فرض کنید قصد داریم پیامی را از طریق یک کانال مخابراتی انتقال دهیم. این پیام می‌تواند فرمت‌های متنوعی داشته باشد (مثلا عکس باشد، یا یک متن ساده و یا ...). بنابراین برای سادگی انتقال پیام از یک Source encoder استفاده می‌کنیم و پیام را به رشته‌هایی از ۰ و ۱ تبدیل می‌کنیم. برای سادگی فرض کنید طول این رشته‌ها ثابت و برابر N بیت است. حال پیام باید از طریق کانال مخابراتی، مخابره شود و به مقصد برسد. همانطور که می‌دانید کانال‌ها دارای نویز هستند و پیامی که به مقصد می‌رسد از این کانال نویزی عبور کرده و ممکن است تعدادی از بیت‌های آن complement شده باشد. بنابراین باید مکانیزمی طراحی کنیم که بتوانیم بهترین تخمین ممکن از پیام اصلی را با استفاده از این پیام نویزی به دست آوریم. بنابراین علاوه‌بر Source encoder یک Channel encoder نیز در نظر می‌گیریم که به صورت $Y = GX$ ساخته می‌شود. فرض کنید $G \in \mathbf{B}^{2N \times N}$ است و درنتیجه $Y \in \mathbf{B}^{2N}$ است. درواقع X رشته‌ی اصلی است و Y رشته‌ای است که وارد کانال می‌شود. ماتریس G به گونه‌ای است که N بیت اول Y با X برابر است و N بیت بعدی برای parity check لحاظ می‌شود. ماتریس $H \in \mathbf{R}^{N \times 2N}$ را درنظر بگیرید که از روی G ساخته می‌شود. هر رشته‌ی $2N$ بیتی لزوما نمی‌تواند کدی باشد که از G بتواند ساخته شود. اگر $HY = 0$ برقرار باشد، آنگاه کد Y می‌تواند از G حاصل شده است. کانال را نیز با عدد ϵ مدلسازی می کنیم. به این صورت که هر بیت، به صورت مستقل نسبت به بیت‌های دیگر با احتمال ϵ، مقدار خود را complement می‌کند.

درنهایت رشته‌ی نویزی $\widetilde{Y}$ به مقصد رسیده و قصد داریم با استفاده از این رشته، تخمین مناسبی از رشته‌ی اصلی به دست آوریم. برای اینکار از یک مدل گرافی غیرجهت‌دار استفاده می‌کنیم. این گراف را می‌توان دوبخشی و یا غیردو بخشی درنظر گرفت که ما در این سوال برای سادگی آن را غیر دو بخشی درنظر می‌گیریم. به اندازه $2N$ گره در این گراف داریم که هرکدام نماینده‌ی بیت‌های پیام هستند. دو نوع فاکتور برای این مدل گرافی تعریف می‌کنیم. مدل اول، unary factors هستند که برای هر بیت تعریف می‌شود و مقدار آن به صورت زیر تنظیم می‌شود.

$$\phi[i] = \begin{cases} 1-\varepsilon & Y_i = \widetilde{Y}_i \\ \varepsilon & Y_i \neq \widetilde{Y}_i \end{cases}$$

دومین نوع فاکتور در این گراف از ماتریس H ساخته می‌شود. درواقع هر سطر این ماتریس مشخص کننده‌ی یک فاکتور در این گراف است. به عبارت بهتر، در هر سطر این ماتریس درایه‌هایی که ناصفر هستند یک فاکتور می‌سازند. مقداردهی این فاکتورها نیز مشخص است. دقت کنید قبلا اشاره کردیم که برای هر کد معتبر Y رابطه‌ی $HY = \mathbf{0}$ برقرار است. پس اگر مقداردهی فاکتور، شامل زوج مقدار ۱ باشد، مقدار آن فاکتور ۱ و درغیر این صورت ۰ است.

نکته‌ی دیگری که باید درمورد ماتریس H درنظر داشته باشید این است که این ماتریس رو می‌توان جزو خانواده‌ی ماتریس‌های Sparse قرار داد و هر ردیف شامل تعداد محدود بسیار کمی درایه‌ی ۱ است و باقی درایه‌ها ۰ است. پس هر فاکتور شامل تعداد کم و محدودی از assignment هایی است که می‌تواند منجر به مقدار ۱ شود.

الف. تابعی پیاده‌سازی کنید که با دریافت پیام نویزی $\widetilde{Y}$ و همچنین ماتریس G گراف گفته‌شده و فاکتورهای آن را بسازد. (۱۰ نمره)

ب. تابعی پیاده‌سازی کنید که با دریافت مقداردهی Y بتواند توزیع توام آن را محاسبه کند. (۴ نمره)

پ. ماتریس H زیر را درنظر بگیرید.

$$H = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 \end{bmatrix}$$

توزیع توام را برای هرکدام از پیام‌های زیر محاسبه کنید و آیا پاسخ با شهودتان مطابقت دارد؟

- $Y = [0, 1, 1, 0, 1, 0]^T$
- $Y = [1, 0, 1, 1, 1, 1]^T$
- $Y = [1, 0, 1, 1, 0, 1]^T$

(۶ نمره)

ت. درفایل GH.mat دو ماتریس G و H برای حالت $N = 128bit$ قرار گرفته‌است. فرض کنید پیام اصلی شامل بیت‌های تماما ۱ است. حال پیام نویزی را مطابق فرایندی که گفتیم، با فرض $\epsilon = 0.05$ شبیه‌سازی کنید. حال می‌خواهیم با استفاده از این پیام نویزی به‌دست آمده، پیام اصلی را تخمین بزنیم. روشی که پیشنهاد می دهیم این است که توزیع حاشیه‌ای (marginal) هر بیت را محاسبه کنید و مقداری را انتخاب کنید که آن را بیشینه می‌کند. دقت کنید که هر بیت تنها مقدار ۰ یا ۱ می‌تواند بگیرد. پس از پیدا کردن assinment ای که بیشترین احتمال را به شیوه‌ی گفته شده حاصل می‌کند، ۱۲۸ بیت اول آن تخمینی است که از پیام اولیه به دست آورده‌ایم. آیا این پیام تخمین زده شده با پیام اصلی (بیت‌های تماما یک) برابر است؟ (۱۶ نمره)

برای محاسبه‌ی توزیع حاشیه‌ای می‌توانید از روش هایی که بلد هستید (مثل Variable Elimination و یا ...) استفاده کنید. اما از آنجایی که مقیاس این سوال نسبتا زیاد است شاید استفاده از این الگوریتم‌ها از نظر زمانی به صرفه نباشد. بنابراین پیشنهاد می‌شود از الگوریتم‌های Massege Passing مثل Belief propagation (sum product) و یا حالت‌های تقریبی این الگوریتم‌ها استفاده کنید.

برای خواندن فایل گفته شده نیز پیشنهاد می‌شود از ماژول scipy.io و تابع sio.loadmat استفاده کنید. همچنین تنها مجاز به استفاده از numpy و scipy هستید.