مدل‌های مولد

نیم‌سال اول ۱۴۰۳-۱۴۰۴

مدرس: دکتر سیدصالحی

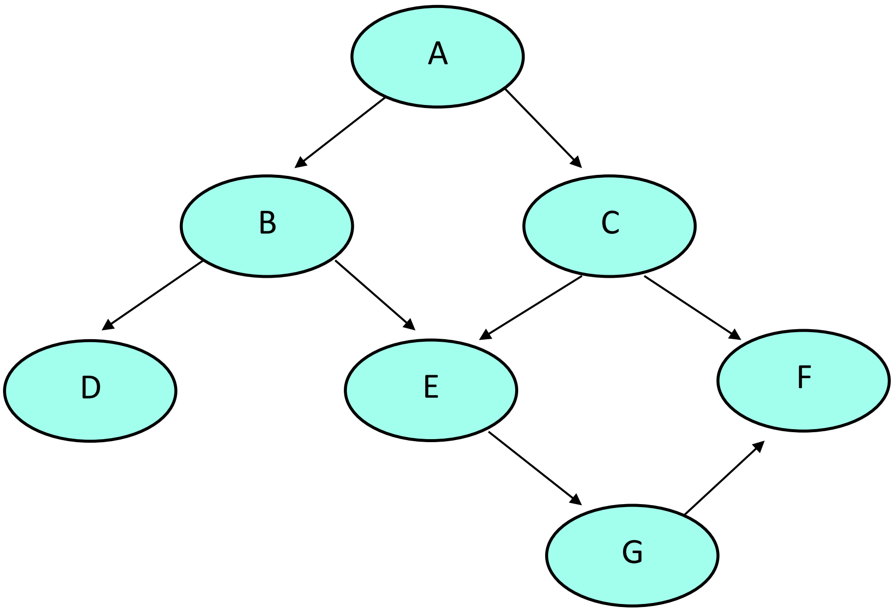
دانشکده علوم ریاضی

تمرین سری اول | موعد تحویل: ۱۱ آبان

- مهلت ارسال پاسخ تا ساعت ۲۳ : ۵۹ روز مشخص‌شده است.
- در طول ترم امکان ارسال با تاخیر پاسخ همه‌ی تمارین تا سقف ۹ روز وجود دارد. پس از گذشت این مدت، پاسخ‌های ارسال‌شده با سیاست با تاخیر خواهند بود.
- هم‌کاری و هم‌فکری شما در انجام تمرین مانعی ندارد اما پاسخ ارسالی هر کس حتما باید توسط خود او نوشته شده باشد.
- در صورت هم‌فکری و یا استفاده از هر منابع خارج درسی، نام هم‌فکران و آدرس منابع مورد استفاده برای حل سوال مورد نظر را ذکر کنید.
- لطفا تصویری واضح از پاسخ سوالات نظری بارگذاری کنید. در غیر این صورت پاسخ شما تصحیح نخواهد شد.

مسئله‌ی ۱. (۱۰ نمره)

با در نظر گرفتن گراف زیر، درست و نادرست بودن هر یک از عبارات زیر را با دلیل مشخص کنید.



شکل ۱: شبکه بیزین

۱. $A \perp D$

۲. $A \perp E | B$

۳. $B \perp C | A$

۴. $B \perp C | A, E$

۵. $D \perp F | E$

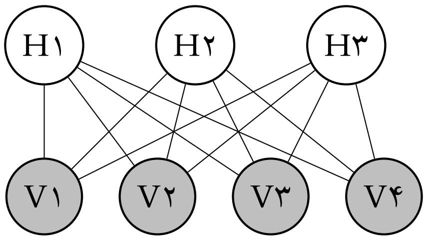
مسئله‌ی ۲. (۵ نمره)

فرض کنید G یک شبکه بیزین و H یک شبکه مارکوف روی مجموعه متغیرهای X باشند به گونه‌ای که اسکلت دو گراف مانند هم باشد. در چه شرایطی $I(G) = I(H)$ خواهد بود؟

مسئله‌ی ۳. (۲۰ نمره)

ماشین بولتزمن محدود[۱] یک مدل گرافی دارای تابع احتمالاتی به صورت زیر است:

$$P(V, H) = \frac{1}{Z} \exp(\sum_i a_i v_i + \sum_j b_j h_j + \sum_{i,j} w_{ij} v_i h_j)$$



آ) نشان دهید که در این مدل می‌توان توزیع شرطی متغیرهای پنهان (H) به شرط متغیرهای مشاهده شده (V) را به صورت زیر نوشت:

$$P(H|V) = \prod_j P(H_j|V)$$

که:

$$P(H_j = 1|V = v) = \sigma(b_j + \sum_i w_{ij} v_i)$$

ب) فرم تجزیه شده[۲] توزیع $P(V|H)$ را به دست آورید. (در صورت شباهت با قسمت پیش نیازی به محاسبه جزئیات نیست)

پ) توزیع $P(H)$ را به دست آورید. آیا می‌توان توزیع $P(H)$ را بر روی متغیرهای نهان (H) تجزیه کرد؟

ت) آیا در مدل گرافی این توزیع، استقلال شرطی وجود دارد که از فرمول تابع احتمالاتی آن به دست نیاید؟ در مقابل آیا تمام استقلال‌های شرطی که از فرمول تابع توزیع این احتمال به دست می‌آیند، از گراف آن قابل استنتاج هستند؟

مسئله‌ی ۴. (۵ نمره)

با خواندن بخش‌های ۱-۴ این مقاله به صورت مختصر توضیح دهید که چگونه از Structural Causal Model یا SCM در افزایش قدرت تفسیرپذیری مدل‌های ژرف استفاده می‌شود؟

مسئله‌ی ۵. (۵ نمره)

در کلاس پارامترهای مدل HMM را در حالتی که همه متغیرهای تصادفی مدل دیده شده باشند، بررسی کردیم. حال حالتی را فرض کنید که مدل فقط از روی جملات به تنهایی باید آموزش دیده شود و متغیرهای نهان را نداریم. در این حالت برای یادگیری متغیرها، باید از الگوریتم Expectation-Maximization استفاده کنیم. گام‌های

---

[۱] Restricted Boltzman Machine

[۲] factorized

الگوریتم را برای این مدل به دست آورید. برای این کار می‌توانید از این منبع کمک بگیرید. (دقت بفرمایید که فایل برای یادگیری و راهنمایی است و کپی کردن متن آن در جواب تمرین نمره‌ای ندارد و خودتان باید بعد از مطالعه مجدد عبارات را به دست آورید.)

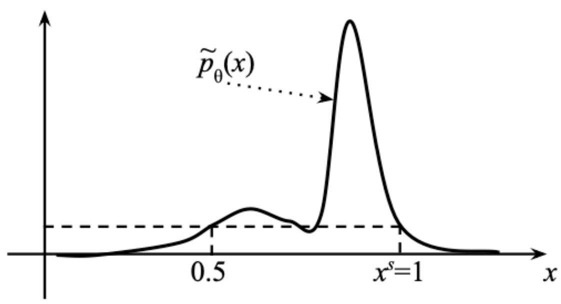
مسئله‌ی ۶. (۵ نمره)

می‌خواهیم با استفاده از Naïve Bayes بررسی کنیم که ایمیل دریافت شده کاری است یا نه. با استفاده از داده‌های زیر کاری بودن ایمیل یا کاری نبودن آن را بررسی کنید. تفاوت این روش با مدل‌های Discriminative چیست؟ (راهنمایی: برای جلوگیری از صفر شدن احتمالات، می‌توانید به صورت همه کسرها عدد یک را اضافه کنید)

email	words used	work related?
۱	schedule – event – schedule	yes
۲	schedule – schedule – meeting	yes
۳	event – deadline	yes
۴	movie – lunch – schedule	no
۵	schedule – movie – lunch – schedule	?

مسئله‌ی ۷. (۱۵ نمره)

نمودار زیر توزیع نرمالایز نشده $\tilde{p}_\theta(x)$ را نشان می‌دهد:



می‌خواهیم یک مرحله از اجرای الگوریتم Metropolis-Hastings را از نقطه $x^s = ۱$ اجرا کنیم. اگر توزیع پیشنهادی به صورت یک توزیع نمایی باشد، به پرسش‌های زیر پاسخ دهید:

$$q(X|x^s) = \text{Exp}(X|\lambda = x^s + ۱)$$

آ) اگر نتیجه نمونه‌برداری از توزیع پیشنهادی، میانگین این توزیع باشد در این صورت x^p چیست؟

ب) مقدار $\alpha = \frac{\tilde{p}(x^p)}{\tilde{p}(x^s)}$ را محاسبه کنید.

پ) مقدار Hastings Correction یا همان $\frac{q(x^s|x^p)}{q(x^p|x^s)}$ را حساب کنید.

ت) عبارات زیر و مقادیر آن‌ها را مقایسه کرده و توضیح دهید که Hastings Correction چه تأثیری بر نقطه پیشنهادی برای نمونه‌برداری و توزیع پیشنهادی دارد.

$$\alpha_۱ = \frac{\tilde{p}(x^p)}{\tilde{p}(x^s)}, \quad \alpha_۲ = \frac{\tilde{p}(x^p)q(x^s|x^p)}{\tilde{p}(x^s)q(x^p|x^s)}$$

مسئله‌ی ۸. (۱۰ نمره)

متغیر تصادفی $X_i \sim Normal\left(X_i;\mu,\tau^{-1}\right)$ که $i \in \{1,\cdots,N\}$ را در نظر بگیرید. فرض کنید X_iها i.i.d. هستند. هر یک از پارامترهای این توزیع دارای عدم قطعیت اند، به‌طوری‌که

$$\mu \sim Normal\left(\mu;\mu_0,(\tau\tau_0)^{-1}\right),$$
$$\tau \sim Gamma\left(\tau;\alpha_0,\beta_0\right)$$

که α_0، β_0 و τ_0 مقادیری مشخص و ثابت اند. می‌خواهیم با استفاده از استنتاج وردشی توزیع پسین $p\left(\mu,\tau|x\right)$ را محاسبه کنیم. مطابق روش Mean-Field تقریب توزیع احتمال پسین را به‌صورت $q\left(\mu,\tau\right)=q(\mu)q(\tau)$ درنظر بگیرید و توزیع‌های $q(\mu)$ و $q(\tau)$ را محاسبه کنید.

راهنمایی. توزیع $q(\mu)$ توزیع نرمال و $q(\tau)$ توزیع گاما خواهند بود.

مسئله‌ی ۹. سوال عملی اول (۱۰ نمره)

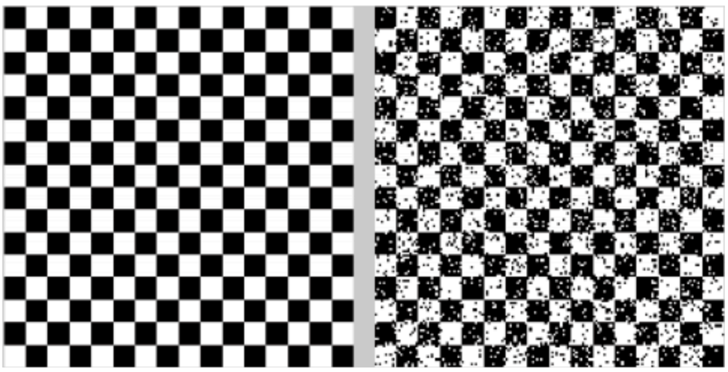
یک تصویر دو بعدی از پیکسل های باینری $x_i \in \{+1,-1\}$ را در نظر بگیرید. حال فرض کنید که یک نسخه نویزی از این تصویر با پیکسل‌های y_i به ما داده شده است به گونه‌ای که در آن هر y_i با احتمال ۰/۱ مقدارش با x_i متفاوت است. در شکل زیر یک نمونه تصویر واقعی و نویزی نمایش داده شده است. در این مساله هدف بازیابی تصویر اصلی با مشاهده تصویر نویزی به کمک یک شبکه مارکوف است. تابع انرژی شبکه روی این مجموعه متغیرها به صورت زیر تعریف می‌شود:

$$E = \alpha\sum_i x_i - \beta\sum_{i,j} x_i x_j - \gamma\sum_i x_i y_i$$

که پارامترهای مدل با مقادیری مثبت هستند که باید تنظیم شوند. براساس تصویر داده شده در فایل HW1_image.mat باید مقدار واقعی هر پیکسل را براساس مقدار نویزی آن که در فایل مزبور نشان داده شده است، به دست آورید. برای این کار باید مقدار بهینه برای تابع انرژی بالا را بیابید. برای بهینه‌سازی در ابتدا مقادیر پیکسل‌های واقعی یعنی x_iها را با مقادیر نویزی آن که در فایل آمده مقداردهی اولیه کنید. سپس، به ترتیب روی تمامی پیسکل‌ها چک کنید که مقداردهی آن به ۱+ و یا ۱− مقدار تابع انرژی را کاهش می‌دهد و یا خیر. این کار را تا جایی که مقدار تابع انرژی همگرا شود تکرار کنید.

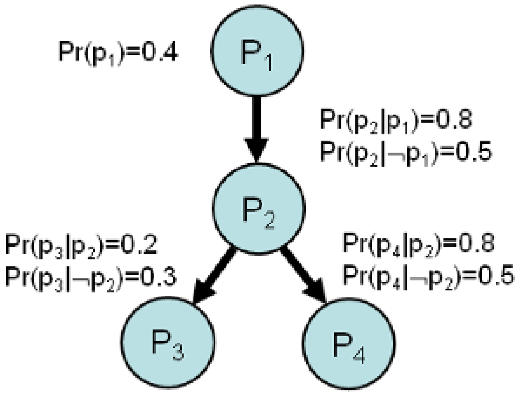
آ) ضمن نمایش تصویر بازیابی شده، مقدار عددی دقت خود را بیان کنید.

ب) در رابطه با اثر کم و زیاد کردن هر یک از سه پارامتر مدل بحث کنید و برای یکی از این سه پارامتر این اثر را با تغییر آن پارامتر و بازیابی مجدد تصویر بررسی کنید.



مسئله‌ی ۱۰. سوال عملی دوم (۱۵ نمره)

در این سوال قرار است که شبکه‌بیزی زیر را پیاده‌سازی کرده و پاسخ پرسش‌های زیر را به صورت عملی بدست آوریم.



آ) $Pr(p_۲, \neg p_۳)$ را محاسبه کنید.

ب) $Pr(p_۱|p_۲, \neg p_۳)$ را محاسبه کنید.

پ) با استفاده از نمونه‌های تصادفی می‌خواهیم نمونه‌هایی را برای تخمین $Pr(p_۱|p_۲, \neg p_۳)$ ایجاد کنیم. این کار را با استفاده از دو روش Rejection Sampling و Gibbs Sampling انجام می‌دهیم. در روش Rejection Sampling به این صورت عمل می‌کنیم که در هر مرحله مقدار یکی از متغیرها را بدست می‌آوریم و در صورتی که sample بدست آمده در لحظه با خواسته‌های ما ناسازگار بود آن sample را reject کرده و به سراغ ایجاد یک sample جدید می‌رویم. در واقع با این کار می‌خواهیم نسبت به prior sampling کمی بهینه‌تر عمل کرده و از ایجاد نمونه‌های غیر دلخواه پیشگیری کنیم. در روش Gibbs Sampling از یک sample اولیه سازگار شروع کرده و در هر مرحله از یک متغیر به شرط سایر مقادیر نمونه‌برداری می‌کنیم ولی به شواهد (مقادیر معلوم) دست نمی‌زنیم.

موفق باشید :)