بسمه‌تعالی

مدل‌های گرافی احتمالاتی — **نیمسال دوم ۰۰-۹۹**

امتحان میان ترم — **وقت آزمون: ۱۵۰ دقیقه**

توجه: استفاده از کتاب، جزوه، اسلایدهای درس، اینترنت و مشورت در حین امتحان غیر مجاز است و تقلب محسوب می‌شود. در صورت تشخیص تقلب، نمره کل امتحان **صفر** منظور خواهد شد.

سوال ۱ (۱۵ نمره) روابط استقلال شرطی

گزاره‌های زیر را در صورت درستی اثبات کنید و در صورت نادرست بودن برای آن‌ها مثال نقض یا دلیل بیاورید.

a. (۳ نمره) اگر $X \perp Y, W | Z$ آن‌گاه $X \perp Y | Z$

b. (۳ نمره) اگر $X \perp Y, W | Z$ و $Y \perp W | Z$ آن‌گاه $X, W \perp Y | Z$

c. (۳ نمره) اگر P-map برای یک توزیع وجود داشته باشد، حتما یکتاست.

d. (۳ نمره) در شبکه زیر $A \perp E \mid C$

e. (۳ نمره) در شبکه زیر $E \perp C \mid B$

$$\begin{tikzcd} A \arrow[r] & B \arrow[r] & C \\ D \arrow[ur] \arrow[r] & E & \end{tikzcd}$$

سوال ۲ (۲۵ نمره) استنتاج روی درخت

توزیع زیر را در نظر بگیرید:

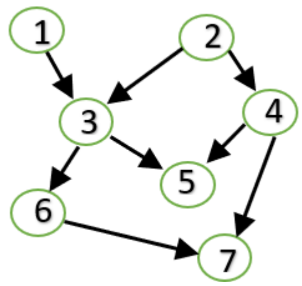
$$P(X_1, \dots, X_T, Y_1, \dots, Y_T) = P(X_1) \prod_{t=1}^{T} P(Y_t | X_t) \prod_{t=2}^{T} P(X_t | X_{t-1})$$

a. (۱۰ نمره) شبکه بیزین و همچنین فاکتور گراف معادل با آن را رسم کنید.

b. (۱۰ نمره) فرض کنید به دنبال پیدا کردن $P(X_t | Y_1, \dots, Y_T)$ برای $1 \le t \le T$ هستیم. در اعمال روش استنتاج sum-product روی شبکه‌ی بیزین بالا، ابتدا توابع ϕ را مشخص کنید و سپس رابطه مربوط به پیامی که X_{t+1} به X_t می‌فرستد را بنویسید.

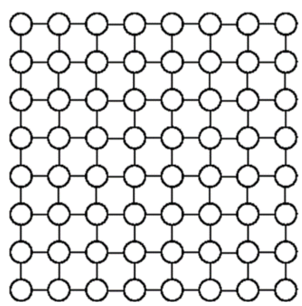
c. (۵ نمره) اگر الگوریتم sum-product را روی یک درخت اعمال کرده باشیم و پیام‌های لازم برای محاسبه‌ی همه احتمال‌های حاشیه‌ای (marginal) روی تک متغیرها به دست آمده باشند. چگونه از روی این پیام‌ها می‌توان توزیع توام تمام متغیرهای یک زیردرخت از درخت اصلی (زیرمجموعه‌ای از گره‌های درخت اصلی که یال‌ها روی آن یک درخت را تشکیل می‌دهند) را پیدا کرد؟ توضیح دهید.

سوال ۳ (۱۸+۵ نمره) Junction tree



a. (۱۵ نمره) در گراف بالا اگر ترتیب حذف متغیرها به صورت 1,7,3,2,4,6,5 باشد (متغیر ۱ اول حذف می‌شود)، junction tree و گراف مثلث‌بندی شده نهایی را بدست آورید.

b. (۳ نمره) ساختار گرافی زیر را که به صورت یک Grid با ابعاد $m \times n$ است، در نظر بگیرید (m: سطر، n: ستون). اگر ترتیب حذف متغیرها به صورت ستون به ستون از چپ به راست باشد، اندازه بزرگترین خوشه‌ی ایجاد شده در junction tree (یا اندازه بزرگ‌ترین خوشه در گراف مثلث‌بندی شده) از چه مرتبه‌ای خواهد بود؟ توضیح دهید.



c. (اختیاری ۵ نمره) اگر برای یک مدل گرافی بدون جهت، junction tree متناظر با آن را داشته باشیم. اگر به این مدل یک یال دلخواه اضافه کنیم و با همان ترتیب حذف قبلی، junction tree را برای این مدل جدید بدست آوریم، آن‌گاه اندازه بزرگترین خوشه درخت جدید حداکثر چقدر بزرگ‌تر اندازه بزرگترین خوشه در درخت قبلی است؟

سوال ۴ (۱۵ نمره) یادگیری

a. (۵ نمره) نشان دهید اگر توزیع $P(\theta|\alpha)$ یک conjugate prior برای پارامتر θ باشد، آن‌گاه توزیع $\sum_{d=1}^{D} \lambda_d P(\theta|\alpha_d)$ هم یک conjugate prior برای θ خواهد بود.

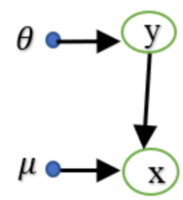
b. (۱۰ نمره) یادگیری در مدلهای بدون جهت

مدلهای بدون جهتی را در نظر بگیرید که توزیع احتمال در آن‌ها به صورت $P(\boldsymbol{x}) = \frac{1}{Z} \exp \sum_{i=1}^{K} \boldsymbol{\theta}_i^T \boldsymbol{f}(\boldsymbol{x}_{c_i})$ تعریف می‌شود که $\boldsymbol{x}_{c_i}$ مجموعه متغیرهای موجود در خوشه c_i را نشان می‌دهد. چنان‌چه مجموعه داده‌های آموزش $\mathcal{D} = \{\boldsymbol{x}^{(n)}\}_{n=1}^{N}$ را در اختیار داشته باشیم، نشان دهید برای بیشینه کردن likelihood با استفاده از تکنیک gradient descent، گرادیان $\nabla_{\boldsymbol{\theta}_i} \ln P(\mathcal{D}|\boldsymbol{\theta})$ برابر است با:

$$\sum_{n=1}^{N} \boldsymbol{f}_i\left(\boldsymbol{x}_{c_i}^{(n)}\right) - N \times E_{P(\boldsymbol{x}|\boldsymbol{\theta})}[\boldsymbol{f}_i(\boldsymbol{x}_{c_i})]$$

سوال ۵ (۲۷ نمره) Expectation-Maximization

مدل گرافی زیر را که برای یک مسأله دسته‌بندی است را درنظر بگیرید. در این مدل، متغیر y یک متغیر باینری است که مشخص کننده کلاس متغیر x است. توزیع‌های شرطی این مدل به صورت زیر است:



$$p(y) = bernoulli(\theta)$$

$$p(x|y=0) = \mathcal{N}(x; \mu_0, 1)$$

$$p(x|y=1) = \mathcal{N}(x; \mu_1, 1)$$

فرض کنید داده‌های آموزشی مجموعه $D = \{x_1, \dots, x_N\}$ باشد. می‌خواهیم با استفاده از روش EM، پارامترهای θ، μ_0 و μ_1 را تخمین بزنیم.

a. (۱۰ نمره) با محاسبه $\mathbb{E}[\log p(x_1, \dots, x_N, y_1, \dots, y_N)]$ نشان دهید که expected sufficient statistics برای این مسأله عبارت $\sum_{i=1}^{N} \mathbb{E}[y_i]$ است.

راهنمایی: توزیع $p(x|y)$ را به صورت زیر بنویسید:

$$p(x|y) = \left(\mathcal{N}(x; \mu_0, 1)\right)^{1-y}\left(\mathcal{N}(x; \mu_1, 1)\right)^{y}$$

b. (۵ نمره) نشان دهید که در گام *E*، امید ریاضی *sufficient statistics* از رابطه‌ی زیر بدست می‌آید:

$$\mathbb{E}_{p(y_i|x_i)}[y_i] = \frac{\mathcal{N}(x_i; \mu_1^t, 1)\,\theta^t}{\theta^t \mathcal{N}(x_i; \mu_1^t, 1) + (1-\theta^t)\mathcal{N}(x_i; \mu_0^t, 1)}$$

c. (۱۲ نمره) گام *M*: روابط به روز رسانی پارامترهای θ و μ_0 و μ_1 را بدست آورید.

موفق باشید.