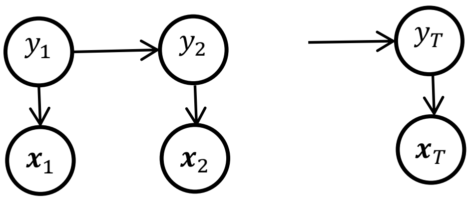
بسمه تعالی

مدل‌های گرافی احتمالاتی | **تیر ماه ۱۴۰۰**

امتحان پایان 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 (۲۰ نمره) مدل‌های جهت‌دار و یادگیری



مدل HMM به صورت روبرو را در نظر بگیرید که مقدار y_t (که state را نشان می‌دهد) عضو $\{1, \dots, K\}$ است و احتمال مشاهده‌ی بردار $\boldsymbol{x}_t$ در هر state به صورت یک گاوسی مدل می‌شود: $P(\boldsymbol{x}_t|y_t = k) = \mathcal{N}(\boldsymbol{x}_t|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$

چنان‌چه مجموعه آموزش $\mathcal{D} = \left\{\left(\boldsymbol{X}^{(n)}, \boldsymbol{y}^{(n)}\right)\right\}_{n=1}^{N}$ را در اختیار داشته باشیم ($\boldsymbol{X}^{(n)} = \left[\boldsymbol{x}_1^{(n)}, \dots, \boldsymbol{x}_T^{(n)}\right]$ بردار مشاهدات در لحظات مختلف و $\boldsymbol{y}^{(n)} = \left[y_1^{(n)}, \dots, y_T^{(n)}\right]$ شماره state در لحظه‌های مختلف را برای داده‌ی آموزش n-ام شامل می‌شود و T طول دنباله‌ی مربوط به این داده).

a. (۵ نمره) کلیه‌ی پارامترهای مدل بالا را مشخص کنید و محل آن‌ها را در شکل مربوطه مشخص کنید.

b. (۱۵ نمره) روابط مربوط به یادگیری پارامترهای $\boldsymbol{\mu}_k$ را با استفاده از داده‌های آموزش و تخمین ML به دست آورید.

سوال ۲ (۲۵ نمره) مدلهای مشهور

۲.۱. (۱۵ نمره) مساله پیدا کردن دنباله‌ی برچسب برای دنباله‌ای از ورودی‌ها (که هر ورودی با بردار $\boldsymbol{x} = [x_1, \dots, x_T]$ مشخص می‌شود) را در نظر بگیرید. می‌خواهیم از یک مدل CRF با توزیع شرطی زیر برای این منظور استفاده کنیم:

$$P(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{w}) = \frac{1}{Z(\boldsymbol{x}, \boldsymbol{w})} \exp\left\{\boldsymbol{w}^T \sum_{t=1}^{T} \boldsymbol{f}(\boldsymbol{x}, y_t, y_{t-1})\right\}$$

که $\boldsymbol{f}$ برداری از خواص را نشان می‌دهد و $\boldsymbol{w}$ بردار وزن متناظر را مشخص می‌کند. چنان‌چه مجموعه داده‌های آموزش $\mathcal{D} = \left\{\left(\boldsymbol{x}^{(n)}, \boldsymbol{y}^{(n)}\right)\right\}_{n=1}^{N}$ را در اختیار داشته باشیم، نشان دهید برای بیشینه کردن conditional likelihood با استفاده از تکنیک gradient descent، گرادیان مربوطه برابر است با:

$$\sum_{n=1}^{N}\left(\sum_{t=1}^{length(\boldsymbol{x}^{(n)})} f\left(\boldsymbol{x}^{(n)}, y_t^{(n)}, y_{t-1}^{(n)}\right) - E_{P(\boldsymbol{y}|\boldsymbol{x}^{(n)})}\left[f\left(\boldsymbol{x}^{(n)}, y_t, y_{t-1}\right)\right]\right)$$

۲.۲. (۵ نمره) اگر N داده برچسب نخورده در یک فضای D بعدی داشته باشیم و بخواهیم از مدل Factor Analysis برای کاهش ابعاد داده‌ها به فضای $L < D$ استفاده کنیم و در نهایت داده‌های کاهش ابعاد یافته را به عنوان خروجی داشته باشیم، مراحل کلی کار را توضیح دهید (ذکر جزئیات روابط لازم نیست).

۲.۳. (۵ نمره) در یک مدل LDS که تمام متغیرها گاوسی هستند، توضیح دهید که برای پیدا کردن

$$z_1^*, \dots, z_T^* = \arg\max_{z_1, \dots, z_T} P(z_1, \dots, z_T|x_1, \dots, x_T)$$

می‌توان از دنباله‌ی مقادیر متغیرهای latent که هر کدام تک‌تک به صورت $z_t^* = \underset{z_t}{\operatorname{argmax}} P(z_t|x_1, \dots, x_T)$ به دست آمده‌اند، استفاده کرد، به عبارت دیگر نیاز به الگوریتمی شبیه Viterbi نیست. (راهنمایی: جواب این سوال نیازی به هیچ رابطه‌ی ریاضی ندارد، کافیست از ویژگی‌های توزیع گاوسی استفاده کنید و دلیل خود را توضیح دهید).

سوال ۳ (۱۵ نمره) یادگیری ساختار

۱٫۳. (۸ نمره) الگوریتم Chow-Liu را توضیح دهید.

۲٫۳. (۷ نمره) دلیل ایجاد مشکل overfitting در روش likelihood-score را توضیح دهید.

یادآوری: $score_L = N\sum_{i=1}^{M} I_{\hat{P}_{\mathcal{D}}}\left(X_i; Pa_{X_i}\right) - N\sum_{i=1}^{M} H_{\hat{P}_{\mathcal{D}}}(X_i)$ که M تعداد گره‌ها و N تعداد نمونه‌ها را نشان می‌دهد.

سوال ۴ (۲۰ نمره) استنتاج وردشی (variational inference)

مدل مخلوط پواسون (Poisson Mixture Model) زیر را در نظر بگیرید:

$$P\left(z^{(n)}\right) = Multi(\pi_1, \dots, \pi_k)$$

$$P\left(x^{(n)}|z_k^{(n)} = 1\right) = Poisson\left(x^{(n)}|\lambda_k\right)$$

$$P(\lambda_k) = Gamma(\lambda_k|\alpha, \beta)$$

چنان‌چه مطابق روش mean-field، $Q(\boldsymbol{Z}, \boldsymbol{\lambda}) = Q(\boldsymbol{Z})Q(\boldsymbol{\lambda})$ در نظر گرفته شود، نشان دهید برای $Q(\boldsymbol{\lambda})$ داریم $Q(\boldsymbol{\lambda}) = \prod_k Q(\lambda_k)$ که به‌روزرسانی $Q(\lambda_k)$ به صورت زیر انجام می‌شود:

$$Q(\lambda_k) = Gamma\left(\alpha + \sum_{n=1}^{N} E\left[z_k^{(n)}\right] x^{(n)}, \beta\right)$$

یادآوری: برای کمینه کردن معیار $D(Q(Z)||P(Z|X))$ بر حسب $Q_i(Z_i)$ که در Mean Field (با فرض $Q(Z) = \prod_i Q_i(Z_i)$) به‌کار گرفته می‌شود، باید داشته باشیم $Q_i(Z_i) \propto \exp\left(E_{-z_i}[\log P(X, Z)]\right)$.

توزیع گاما: $Gam(\lambda|\alpha, \beta) \propto \lambda^{\alpha-1}\exp(-\beta\lambda)$

توزیع پواسون: $Poisson(x|\lambda) = \frac{\lambda^x \exp(-x)}{x!}$

سوال ۵ (۲۰ نمره) نمونه برداری (sampling)

فرض کنید d متغیر تصادفی $X_1, \dots, X_d$ با توزیع برنولی به صورت زیر داریم:

$$P(X_i|\theta_i) = Bern(X_i|\theta_i) \qquad\qquad i = 1, \dots, d$$

توزیع پیشین با پارامترهای مشترک روی همه‌ی θ_i ها در نظر گرفته‌ایم:

$$p(\theta_i|\alpha, \beta) = Beta(\theta_i|\alpha, \beta)$$

در این سوال قصد داریم با مشخص بودن مقادیر $X_1, \ldots, X_d$ از متغیرهای $\theta_1, \ldots, \theta_d$ نمونه برداریم. برای این منظور می‌خواهیم از Gibbs sampling استفاده کنیم.

۱٫۵. (۱۰ نمره) در این قسمت فرض کنید خود زوج α, β هم متغیر تصادفی فرض شده و توزیع پیشین $p(\alpha, \beta)$ را روی آن در نظر گرفته‌ایم. ابتدا توزیع توام همه متغیرها را بنویسید و با توجه به آن نشان دهید توزیع‌های شرطی کامل به صورت زیر درمی‌آیند:

$$P(\theta_i | rest) = Beta(\theta_i | \alpha + x_i, \beta + 1 - x_i)$$

$$p(\alpha | rest) \propto \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)}\right]^d \prod_{i=1}^{d} \theta_i^{\alpha} p(\alpha, \beta)$$

$$p(\beta | rest) \propto \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\beta)}\right]^d \prod_{i=1}^{d} (1-\theta_i)^{\beta} p(\alpha, \beta)$$

یادآوری: توزیع بتا $Beta(\theta | \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}$

۲٫۵. (۱۰ نمره) فرض کنید قصد داریم توزیع $p(x_1, \ldots, x_d | \alpha, \beta)$ را با استفاده از نمونه‌هایی که از θ_i ها برمی‌داریم تخمین بزنیم. مشخص کنید که به چه شکل می‌توان این تخمین را به دست آورد. (توجه داشته باشید که متغیر x_i به شرط θ_i از سایر متغیرها مستقل است).

موفق باشید.