



یادگیری ماشین

نیم سال دوم ۱۳۹۹

مدرس: سیدعباس حسینی

زمان: ۳ ساعت

پایان ترم

۲ تیر ۱۴۰۰

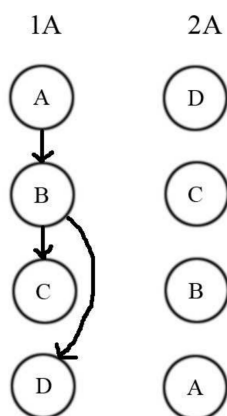
۱. تعداد سوالات ۶ عدد است و جمع امتیاز سوالات ۱۲۰ نمره است.
 ۲. تمام پاسخ‌ها باید همراه با استدلال باشد. هر یک از سوالات را در یک یا چند صفحه جداگانه بنویسید و از هر کدام عکس گرفته و در قالب یک فایل در بخش مربوطه در کوئرا آپلود کنید.
 ۳. زمان امتحان ۳ ساعت است و نیم ساعت نیز برای آپلود پاسخ‌ها در نظر گرفته شده است. بنابراین، بخش آپلود پاسخ‌ها به صورت اتوماتیک ساعت ۱۲:۳۰ بسته می‌شود.
 ۴. شما باید در ابتدا تمامی سوالات را به دقت بخوانید و در صورتی که سوالی دارید در نیم ساعت اولیه آن را بپرسید. برای پرسیدن این سوال می‌توانید از بخش گفتگوی خصوصی vclass استفاده کنید. پس از این ساعت به سوالات شما پاسخ داده نخواهد شد.
 ۵. شما برای پاسخ به سوالات مجاز به استفاده از کتاب، جزوات درس یا جست‌جو در اینترنت یا کمک گرفتن از سایرین نیستید. خواهشمندم با احترام به قوانین، سعی کنید جامعه‌ی آکادمیکی مبتنی بر اعتماد ایجاد کنید.
۱. (۲۵ نمره) درست یا غلط بودن گزاره‌های زیر را با بیان دلیلی کوتاه مشخص کنید.
- (آ) (۵ نمره) با مقداردهی اولیه وزن‌های شبکه عصبی با مقادیر بزرگ در یک شبکه عصبی که از تابع سیگموئید به عنوان تابع فعال‌ساز (Activation Function) استفاده می‌کند، می‌توان از محوشدگی گرادیان (Gradient Vanishing) جلوگیری کرد.
 - (ب) (۵ نمره) محوشدگی گرادیان باعث می‌شود تا یادگیری لایه‌های انتهایی شبکه نسبت به لایه‌های ابتدایی شبکه با سرعت کمتری انجام شود.
 - (ج) (۵ نمره) اگر از توابع خطی به عنوان تابع فعال‌ساز استفاده شود، شبکه‌ی عصبی ۱۰۰ لایه و شبکه‌ی عصبی تک‌لایه توانایی یکسانی دارند.
 - (د) (۵ نمره) در ابتدا یک شبکه‌ی عصبی را با تعداد کمی از دادگان آموزش می‌دهیم. پس از آموزش شبکه، میزان هزینه‌ی آموزش (Training Loss) و هزینه‌ی آزمایش (Test Loss) بالا است. در این حالت با اضافه کردن تعداد داده‌های آموزش می‌توان عملکرد شبکه‌ی عصبی را بهبود داد.
 - () (۵ نمره) در شبکه‌های عصبی معمولاً تعداد بهینه‌های محلی بسیار بیشتر از تعداد نقاط زینی (Saddle Points) است.
۲. (۲۰ نمره) می‌دانیم مسئله‌ی رگرسیون خطی با تابع هزینه‌ی SSE (مجموع مربع خطاها) معادل با در نظر گرفتن تولید داده‌ها (y) ، بر اساس یک رابطه‌ی خطی از ورودی (x) ، به اضافه‌ی نویز گوسی با میانگین صفر و واریانس ثابت است. در برخی از سیستم‌ها واریانس نویز به شکل یک تابع خطی از ورودی در نظر گرفته می‌شود.
- (آ) (۱۰ نمره) مدل حاصل از این فرضیه به چه شکل خواهد بود؟ $P(y|x)$ را بدست آورید.
 - (ب) (۱۰ نمره) تخمین‌گر بیشینه درست‌نمایی ضرایب را در مدلی که در قسمت ۱ ارائه دادید، بدست بیاورید.
۳. (۱۵ نمره) در حالت کلی پیدا کردن محتمل‌ترین (Most Likely) ساختار شبکه‌های بیزین بر اساس داده‌ها حل‌نشدنی (Intractable) است. اما با گذاشتن محدودیت‌هایی روی ساختار شبکه، می‌توان محتمل‌ترین ساختار را پیدا کرد. یکی از این محدودیت‌ها قرارداد کردن ترتیب خاص برای متغیرهاست. به این معنی که تنها یال‌هایی در شبکه حضور دارند که در جهت ترتیب داده شده هستند. مثلاً اگر در ترتیب داده شده X قبل از Y آمده باشد تنها یال $X \rightarrow Y$ می‌تواند وجود داشته باشد.

در این مسئله قصد داریم تا تاثیر این محدودیت را بر روی تعداد پارامترها و تعداد استقلال‌های شبکه‌ی بی‌زین بررسی کنیم. در شکل ۱ به شما یک شبکه‌ی بی‌زنی داده شده که از ترتیب ABCD پیروی می‌کنند. (1A)

(آ) (۵ نمره) در بخش (2A) یک شبکه‌ی بی‌زین بکشید که از ترتیب DCBA پیروی کند. این شبکه باید دقیقاً همان توزیع شبکه‌ی 1A را مدل کند همچنین این شبکه نباید هیچ فرض استقلال‌ی داشته باشد که در شبکه‌ی 1A نیست. ضمناً این شبکه نباید هیچ یال اضافی‌ای هم داشته باشد.

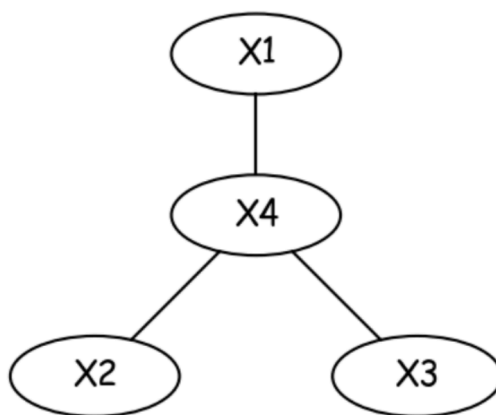
(ب) (۵ نمره) تعداد پارامترهای هر یک از شبکه‌ی 1A, 2A را پیدا کنید. متغیرها را دودویی در نظر بگیرید.

(ج) (۵ نمره) در صورت وجود، استقلال‌هایی که در شبکه‌ی 1A وجود دارد اما در شبکه‌ی 2A وجود ندارد را مشخص کنید.



شکل ۱: شبکه‌های بی‌زین مربوط به سوال ۳

۴. (۱۵ نمره) گراف بدون جهت زیر را در نظر بگیرید که شامل ۴ متغیر تصادفی باینری است.



شکل ۲: گراف مربوط به سوال ۲۳

حال فرض کنید که توزیع توأم این گراف به فرم زیر باشد.

$$P[X_{1:4}] = \frac{1}{Z} \phi_{1,4}(x_1, x_4) \phi_{2,4}(x_2, x_4) \phi_{3,4}(x_3, x_4) \phi_4(x_4)$$

تمام توابع پتانسیل $\phi_{1,4}, \phi_{2,4}, \phi_{3,4}$ مطابق جدول زیر مقداردهی می‌شوند.

همچنین تابع ϕ_4 مطابق جدول زیر مقداردهی می‌شود.

$$\phi(X, Y) =$$

	Y	0	1
X	0	1	2
	1	3	4

$$\phi(X) =$$

X	0	1
	1	2

(آ) (۱۰ نمره) فاکتور گراف (Factor graph) حاصل از این شبکه را رسم کنید و با استفاده از گراف به دست آمده در قسمت قبل، مقادیر زیر را محاسبه کنید.

$$M_{X_1 \rightarrow \phi_{1,2}}$$

$$M_{\phi_2 \rightarrow X_2}$$

(ب) (۵ نمره) سپس با استفاده از نتایج قسمت قبل $P[X_2 = 1]$ را محاسبه کنید.

۵. (۳۰ نمره) در کلاس یادگیری ماشین، احتمال این که دانشجویی نمره A بگیرد $P(A) = \frac{1}{4}$ ، احتمال این که نمره B بگیرد $P(B) = \mu$ ، احتمال این که نمره C بگیرد $P(C) = 2\mu$ و احتمال این که نمره D بگیرد $P(D) = \frac{1}{4} - 3\mu$ است. می‌دانیم در این کلاس c دانشجو نمره C و d دانشجو نمره D گرفته‌اند. اما راجع به این که دقیقاً چند دانشجو نمره A یا B گرفته‌اند چیزی نمی‌دانیم. آنچه می‌دانیم این است که تعداد h دانشجو نمره a یا b گرفته‌اند به عبارت دیگر مقدار دقیق a یا b را نمی‌دانیم اما می‌دانیم که $h = a + b$. هدف این مسئله استفاده از EM برای پیدا کردن تخمین بیشینه‌ی درستنمایی از μ است.

(آ) (۱۰ نمره) گام E : مقدار $\mathbb{E}[a], \mathbb{E}[b]$ را محاسبه کنید.

(ب) (۱۰ نمره) گام M : با فرض داشتن $\mathbb{E}[a], \mathbb{E}[b]$ ، تخمین بیشینه‌ی درستنمایی برای μ را محاسبه کنید.

(ج) (۱۰ نمره) فرض کنید $\mu \sim \text{Beta}(\alpha, \beta)$ ، تخمین Map را برای μ محاسبه کنید.

۶. (۱۵ نمره) یک شبکه‌ی عصبی کاملاً متصل سه لایه (دو لایه‌ی نهان و یک لایه‌ی خروجی) در نظر بگیرید که تابع فعالساز همه‌ی نرون‌های آن سیگموئید $\sigma(x) = \frac{1}{1+e^{-x}}$ است و لایه‌ی خروجی تنها یک نرون دارد. داده‌های ورودی n ویژگی دارند و برای بهینه‌سازی تابع هزینه نیز از الگوریتم گرادیان کاهشی تصادفی استفاده می‌شود. تمام وزن‌های شبکه‌ی عصبی و مقدار مجانبی‌ها (Biases) را در ابتدای فرایند آموزش برابر صفر قرار می‌دهیم و داده‌های $x \in \mathbb{R}^{n \times b}$ را به عنوان ورودی به شبکه‌ی عصبی می‌دهیم. به سوالات زیر پاسخ دهید.

(آ) (۵ نمره) پس از تنها یک بار عملیات انتشار به جلو (Forward Propagation) خروجی شبکه چه مقداری دارد؟

(ب) (۱۰ نمره) فرض کنید پس از یک بار عملیات انتشار به جلو یک بار هم عملیات انتشار به عقب (Backward Propagation) را انجام می‌دهیم. درباره‌ی مقادیر وزن‌های لایه‌ی اول شبکه‌ی عصبی $W^{(1)}$ کدام یک از عبارات زیر صحیح است چرا؟

- مقادیر $W^{(1)}$ می‌تواند مثبت و یا منفی باشد.
- همه‌ی مقادیر $W^{(1)}$ مثبت است.
- همه‌ی مقادیر $W^{(1)}$ منفی است.
- همه‌ی مقادیر $W^{(1)}$ صفر است.