

به نام خدا



دانشکده‌ی مهندسی کامپیوتر

سررسید نظری: ۸ بهمن ماه

سررسید عملی: ۱۰ بهمن ماه

پاییز ۱۴۰۲

یادگیری ماشین

مدرس: مهدی جعفری سیاوشانی

تمرین ۶: خوشه‌بندی و کاهش بعد

- سررسید بخش نظری این تمرین ۸ بهمن ماه ساعت ۵۹: ۲۳ است.
- سررسید بخش عملی این تمرین ۱۰ بهمن ماه ساعت ۵۹: ۲۳ است.
- در صورت کشف تقلب، بار اول برای افراد درگیر تقلب، نمره‌ی همان سوال(های) خاص صفر در نظر گرفته می‌شوند.
- در صورت تکرار، نمره کل تمرین صفر در نظر گرفته می‌شود و در صورت تکرار، درس برای افراد حذف خواهد شد.
- تمامی پاسخ‌های خود را در یک فایل با فرمت (pdf) HW6-[SID]-[Fullname].zip روی کوئرا قرار دهید.

۱ بخش تئوری

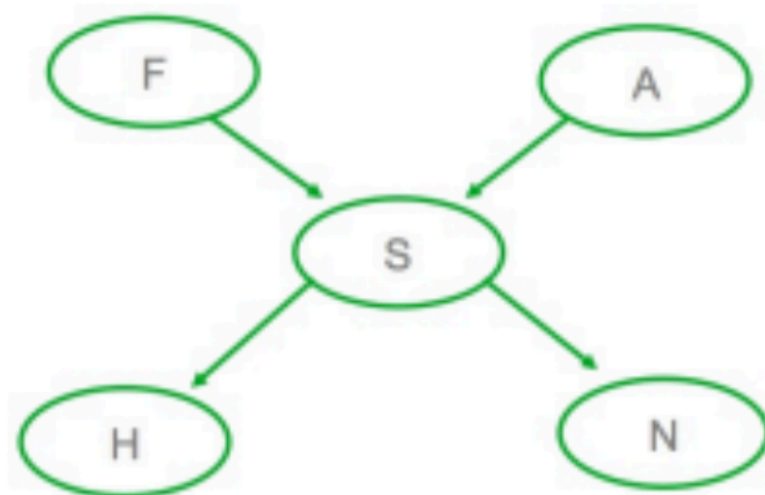
۱.۱ تحلیل مؤلفه‌های اصلی (۲۰ نمره)

سه نقطه $(-۱, -۱)$ ، $(۰, ۰)$ و $(۱, ۱)$ را در یک فضای دو بعدی در نظر بگیرید:

۱. بردار مؤلفه‌های اصلی دادگان را بنویسید (۶)
۲. برای هر یک از مؤلفه‌های اصلی، مختصات تصویر دادگان در زیرفضای یک بعدی متناظر و همچنان واریانس تصویر دادگان در هر زیرفضا را بدست آورید. (۷)
۳. برای هر یک از تصاویر دادگان در مرحله قبل، اگر تصاویر را به عنوان بازنمایی دادگان در فضای دوبعدی در نظر بگیریم، میانگین خطای بازسازی را برای هر یک از تصاویر محاسبه کنید. (۷)

۲.۱ بیشینه سازی امید ریاض (۳۰ نمره)

مدل گرافی زیر که یک توزیع مشترک بر روی ۵ متغیر بولی تعریف می کند را در نظر بگیرید. در این سوال شما به آموزش شبکه بیزین داده شده به کمک روش بیشینه سازی امید ریاضی بر روی دادگان آموزش خواهید پرداخت. در دادگان آموزش ۴ متغیر F و H ، S ، N کاملاً قابل مشاهده هستند و متغیر A گاهی غیر قابل مشاهده است.



۱. توزیع شرطی متناظر با هر یک از متغیرهای این شبکه را بدست آورید. (۵)
۲. در مرحله امید ریاضی الگوریتم، به تخمین توزیع مقادیر متغیرهای مشاهده نشده هر داده آموزش خواهیم پرداخت. از آنجایی که تنها A می تواند متغیر مشاهده نشده باشد، مرحله امید ریاضی معادل محاسبه توزیع شرطی متغیر A خواهد بود. عبارت متناظر برای توزیع $P(A = ۱ | F, S, H, N)$ را بر اساس توزیع های شرطی دیگر متغیرهای شبکه بنویسید. (راهنمایی: با نوشتن تعریف توزیع شرطی شروع کنید). (۵)

۳. کدام متغیر ها در مجموعه ی پتوی مارکوف متغیر A قرار دارند ؟ (۵)

۴. با توجه به اینکه تمامی متغیر های مجموعه پتوی مارکوف متغیر A قابل مشاهده هستند، توزیع متغیر A به تنهایی بر اساس این متغیر ها نیز قابل محاسبه است. عبارت بدست آمده در قسمت (ب) را طوری بازنویسی کنید که تنها شامل متغیر های مجموعه پتوی مارکوف متغیر A باشد. (۵)

۵. در مرحله پیشینه سازی، پارامتر های شبکه بر اساس دادگان آموزش و توزیع متغیر های غیر قابل مشاهده که در مرحله امید ریاضی محاسبه شده اند دوباره محاسبه می شوند. بعضی از پارامتر های شبکه را می توان تنها بر اساس متغیر های قابل مشاهده محاسبه کرد. برخی دیگر از پارامتر های به توزیع متغیر های غیر قابل مشاهده که در مرحله امید ریاضی محاسبه شدند وابسته اند. تمامی پارامتر هایی که به توزیع های محاسبه شده در مرحله امید ریاضی وابسته اند را مشخص کنید.

برای ذکر هر پارامتر توزیعی که آن را نمایندگی میکند را مشخص کنید. (مثلا $P(S = 1 | N = 0)$) (۵)

Example	F	S	H	N	A	$P(A = 1 F, S, H, N)$
1.	0	1	0	1	1	—
2.	0	0	1	0	1	—
3.	1	0	1	1	0	—
4.	0	0	0	1	?	0.8
5.	0	1	0	0	?	0.4

Table 1: Training examples, and E-step results for unobserved values of A

۶. جدول دادگان فوق را در نظر بگیرید. این جدول مقادیر متغیر های قابل مشاهده در دادگان آموزش و توزیع متغیر های غیر قابل مشاهده محاسبه شده در مرحله امید ریاضی سومین تکرار الگوریتم را نشان می دهد. چه تخمینی در گام پیشینه سازی بعدی برای پارامتر تعریف کننده $P(A = 1)$ محاسبه خواهد شد ؟ چه مقداری برای پارامتر $P(S = 1 | F = 0, A = 1)$ بدست خواهد آمد ؟ (۵)

۳.۱ الگوریتم Kmeans (۲۰ نمره)



۱. اگر از الگوریتم های مبتنی بر فاصله مانند Kmeans برای خوشه بندی داده هایی که ابعاد بالایی دارند استفاده کنیم چه مشکلی رخ می دهد؟ (۱۰)
۲. هنگام استفاده از الگوریتم ذکر شده، چگونه استفاده از نرمال سازی بر روی داده ها می تواند موثر باشد؟ مثال بزنید. (۵)
۳. مشکل الگوریتم Kmeans را در خوشه بندی دادگان عکس بالا توضیح دهید. (۵)

موفق باشید