

توضیحات مهم:

- تمامی مستندات خود شامل گزارش و کدهای خود را در یک فایل فشرده با فرمت zip ذخیره کرده و با عنوان studentID_HW2.zip بارگذاری نمایید (بعنوان مثال 99131000_HW2.zip).

- مهلت انجام تمرین تا ساعت ۲۳:۵۵ تاریخ ۱۴۰۰/۰۲/۱۶ می باشد و به هیچ وجه تمدید نمی -

شود.

- استفاده از کتابخانه های رایج در یادگیری ماشین بلا مانع است.
- برای سهولت در انجام تمرین می توانید از پلتفرم کولب گوگل استفاده نمایید.
- ملاک اصلی انجام تمرین گزارش آن است و ارسال کد بدون گزارش فاقد ارزش است. یک فایل pdf تهیه کرده و برای هر سوال، ورودی، خروجی و توضیحات مربوطه را بصورت جامع گزارش کنید.
- تا حد ممکن سعی کنید اصول لازم برای گزارش مهندسی را رعایت نمایید. (به بهترین گزارش نمره ی تشویقی تعلق می گیرد).

- مطابق قوانین دانشگاه هر گونه کپی برداری ممنوع می باشد و در صورت مشاهده نمره ی طرفین صفر می شود.

- شما مجاز هستید برای تمامی تمرین ها ۷ روز در کل و با سقف حداکثر ۳ روز برای هر تمرین، تاخیر بدون کسر نمره داشته باشید. به ازای هر روز تاخیر بیشتر، ۱۰٪ از نمره ی تمرین مربوطه کسر می شود.

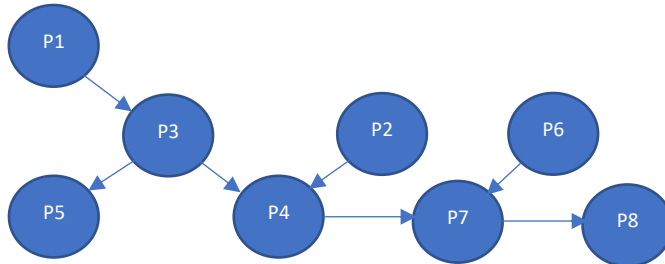
- در صورت داشتن هر گونه ابهام می توانید از طریق ایمیل زیر سوال خود را مطرح نمایید:

fardin.aiar@gmail.com

reza.zadkamali@gmail.com

بخش اول: سوالات تشریحی

(۱) با توجه به شبکه بیزین، درستی یا نادرستی موارد زیر را با ذکر دلیل مشخص کنید.



الف) $P_1, P_5 \perp\!\!\!\perp P_6 | P_8$

ب) $P_1 \perp\!\!\!\perp P_2, P_5 | P_4$

(۲) برای تابع توزیع احتمال $P(A, B, C, D)$ با فرض $P(A, B, C, D) = P(A)P(B)P(C|A, B)P(D|C)$ درستی یا نادرستی موارد زیر مشخص کنید.

الف) $A \perp\!\!\!\perp B | \phi$

ب) $A \perp\!\!\!\perp D | C$

(۳) چرا دسته‌بند بیز ساده‌لوحانه، ((ساده لوحانه)) نامیده می‌شود؟

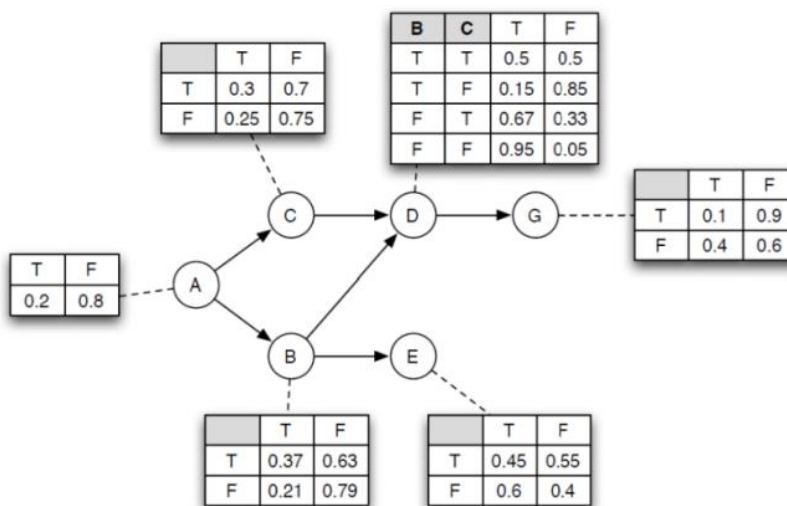
(۴) چرا رگرسیون لاجستیک، ((رگرسیون لاجستیک)) نامیده می‌شود و نه ((دسته‌بند لاجستیک))؟

(۵) پیش‌بینی می‌شود کدام یک از مدل‌های بیز ساده‌لوحانه گاوسی و رگرسیون لاجستیک در دیتاست‌های کوچکتر بهتر عمل کند؟ چرا؟

(۶) جدول داروی مورد نیاز بیماران که در تمرین قبل دیدید مجدداً در شکل زیر آورده شده است. با استفاده از روش بیز ساده‌لوحانه داروی مورد نیاز بیماری که در ردیف آخر آورده شده‌است را مشخص کنید.

Patient ID	Age	Sex	BP	Cholesterol	Drug
p1	Young	F	High	Normal	Drug A
p2	Young	F	High	High	Drug A
p3	Middle-age	F	High	Normal	Drug B
p4	Senior	F	Normal	Normal	Drug B
p5	Senior	M	Low	Normal	Drug B
p6	Senior	M	Low	High	Drug A
p7	Middle-age	M	Low	High	Drug B
p8	Young	F	Normal	Normal	Drug A
p9	Young	M	Low	Normal	Drug B
p10	Senior	M	Normal	Normal	Drug B
p11	Young	M	Normal	High	Drug B
p12	Middle-age	F	Normal	High	Drug B
p13	Middle-age	M	High	Normal	Drug B
p14	Senior	F	Normal	High	Drug A
p15	Middle-age	F	Low	Normal	?

۷) احتمال $P(B|D=T)$ را در شبکه بی‌زین زیر حساب کنید.



بخش دوم: پیاده‌سازی

۱) بیز ساده لوحانه گاوسی

کرایوتراپی یا سرما درمانی با استفاده از سرمای شدید تولید شده از گاز نیتروژن مایع، گاز آرگون و یا دی اکسید کربن، برای رفع ضایعات پوستی، تومورهای سرطانی و موارد درمانی دیگر استفاده می‌شود. دیتاست پیوست نتایج اعمال این روش درمانی در ۹۰ بیمار با ضایعات پوستی سرطانی می‌باشد. ستون `Result_of_Treatment` متغیر خروجی و سایر ستون‌ها ویژگی‌ها می‌باشند.

الف) به نظر شما فرض بیز ساده لوحانه در مورد مستقل بودن متغیرها در این دیتاست برقرار است؟ آیا راهی برای بررسی آن وجود دارد؟

ب) از آنجایی که تعداد نمونه‌های این دیتاست نسبتاً کم است، تقسیم داده‌ها به دو دسته آموزش و آزمون روش مناسبی برای ارزیابی مدل‌های یادگیری ماشین نیست؛ بنابراین یک دسته‌بند ساده لوحانه گاوسی را در این دیتاست آموزش داده و به روش `k-fold cross validation` با مقدار مناسب `k` ارزیابی کنید.

۲) رگرسیون لاجستیک

۱-۲) بخش ب سوال قبل را با رگرسیون لاجستیک انجام داده و نتایج را مقایسه کنید. آیا نتایج با جواب شما به سوال ۵ از سوالات تشریحی همخوانی دارد؟

۲-۲) دیتاست finger digit که در پیوست آمده است شامل مجموعه‌ای از عکس‌های سیاه و سفید انگشتان دست است. هدف شما ساخت مدلی برای تشخیص عدد نشان داده شده در هر عکس است. کتابخانه‌های رایج یادگیری ماشین از جمله scikit-learn به صورت پیشفرض تنها با داده‌های یک بعدی کار میکنند. بنابراین لازم است تصاویر را از حالت دو بعدی به بردارهای یک بعدی تبدیل کنید.

الف) یک مدل رگرسیون لاجستیک آموزش داده و ارزیابی کنید. تقسیم داده‌ها به دو دسته آموزش و آزمون به عهده شما می‌باشد.

ب) پیش از آموزش مدل، تصاویر را به ابعاد ۵۰ در ۵۰ کاهش داده و مجدداً بخش الف را تکرار و مقایسه کنید.