



پردیس دانشکده های فنی

به نام خدا
دانشکده‌ی مهندسی برق و کامپیوتر
تمرین سری دوم یادگیری ماشین



دانشگاه تهران

سلام بر دانشجویان عزیز، چند نکته مهم:

۱. گزارش را حتما با توجه به دستور العمل ارسال شده ، ارسال فرمایید.
 ۲. اگر فرضی برای حل سوال استفاده می کنید حتما آن را ذکر کنید.
 ۳. نام فایل زیپ ارسالی الگوی `ML_HW#_StudentNumber` داشته باشد.
 ۴. از بین سوالات **شبیه سازی** حتما به هر دو مورد پاسخ داده شود.
 ۵. نمره تمرین ۱۰۰ نمره می باشد.
 ۶. هرگونه شباهت در گزارش و کد مربوط به شبیه سازی، به منزله تقلب می باشد و کل تمرین برای طرفین **صفر** خواهد شد.
 ۷. در صورت داشتن سوال، از طریق ایمیل سوال خود را مطرح کنید.
- سوالات ۴ و ۵ و ۶: امیرحسین بنکدار – amirh.bonakdar@ut.ac.ir
- سوالات ۱ و ۲ و ۳: احسان کرمی – ehsan.karamii97@gmail.com

سوال ۱: (۱۵ نمره)

برای یک مسئله‌ی طبقه‌بندی دو کلاسه، برای متغیر ورودی تک بعدی X ، با در نظر گرفتن:

$$P(x|y = 0) = e^{-2x} \quad x \geq 0$$

$$P(x|y = 1) = \frac{2.25}{\sqrt{2\pi}} e^{-\frac{(x-1)^2}{2}} \quad x \geq 0$$

$$\ln \frac{2.25}{\sqrt{2\pi}} \cong -1$$

و فرض برابر بودن احتمال پیشین^۱ هر دو کلاس، بازه‌ی ناحیه‌ی کلاس اول و دوم را بدست آورید.

^۱ Prior

سوال ۲: (۲۰ نمره)

مجموعه داده‌ی جدول ۱ را در نظر بگیرید که در آن متغیر Y برچسب و متغیرهای A, B, C ویژگی‌های باینری هستند.

جدول ۱

A	B	C	Y
0	0	1	0
0	1	0	0
1	1	0	0
0	0	1	0
1	1	1	1
1	0	0	1
1	1	0	1

طبقه‌بند Naive Bayes هر کدام از نمونه‌های زیر را چگونه دسته‌بندی می‌کند.

A	B	C	\hat{Y}
0	0	0	?
1	1	1	?
X	1	0	?
X	0	1	?

سوال ۳: (۱۵ نمره)

برای یک مسئله‌ی طبقه‌بندی دو کلاسه، برای متغیر ورودی تک بعدی X ، با در نظر گرفتن:

$$P(x|y = 0) = N(0, \sigma^2)$$

$$P(x|y = 1) = N(2, \sigma^2)$$

و فرض برابر بودن احتمال پیشین هر دو کلاس، آستانه‌ی به حداقل رسیدن ریسک با فرض اینکه:

$$\begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix} = \begin{pmatrix} 0 & a \\ a^2 & 0 \end{pmatrix}$$

را بدست آوردید. تاثیر افزایش a را در آستانه‌ی بدست آمده تفسیر کنید.

سوال ۴: (۱۵ نمره)

یک مجموعه داده شامل دو کلاس داده‌ی دوبعدی C_1 و C_2 داریم. داده‌های هر کلاس از یک توزیع گاوسی با پارامترهای زیر تولید شده‌اند:

$$C_1: N(\mu_1, \Sigma_1), \quad \mu_1 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad \Sigma_1 = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$$

$$C_2: N(\mu_2, \Sigma_2), \quad \mu_2 = \begin{bmatrix} 5 \\ 1 \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

احتمال پیشین کلاس‌ها نیز به صورت زیر است:

$$P(C_1) = 0.6, \quad P(C_2) = 0.4$$

مرز تصمیم بیزی بین C_1 و C_2 را بدست آوردید.

سوال ۵ (شبيه سازی، ۲۰ نمره):

در یک مسئله طبقه‌بندی ۲ کلاسه، داده‌ها از ۲ توزیع گاوسی متمایز هستند:

$$C_1: p(x|C_1) = N(\mu_1, \sigma_1)$$

$$C_2: p(x|C_2) = N(\mu_2, \sigma_2)$$

احتمال پیشین کلاس‌ها نیز به صورت زیر است:

$$P(C_1) = \pi_1, \quad P(C_2) = \pi_2, \quad \pi_1 + \pi_2 = 1$$

الف) به ازای هر کدام از کلاس‌ها discriminant function را محاسبه کنید. سپس با کم کردن این دو تابع از هم، به یک discriminant function برای مسئله برسید.

$$g(x) = g_1(x) - g_2(x)$$

ب) با استفاده از $g(x)$ ، مرز تصمیم را برای حالت حداقل خطا بدست آورید.

ج) با افزایش احتمال پیشین π_1 و کاهش همزمان π_2 ، مرز تصمیم چگونه تغییر می‌کند؟ به طور شهودی توضیح دهید.

د) ۱۰۰ نمونه داده را از کلاس‌های C_1, C_2 نمونه‌برداری کنید.

$$(\mu_1 = 2, \sigma_1 = 1, \mu_2 = 4, \sigma_2 = 1, \pi_1 = 0.5, \pi_2 = 0.5)$$

و در یک نمودار scatter plot، داده‌ها را نمایش دهید. مرز تصمیمی را نیز در نمودار رسم کنید.

ه) بر اساس مرز تصمیم، لیبل داده‌ها را مشخص کنید. سپس confusion matrix, accuracy, precision, recall, F1-score را برای این طبقه‌بند محاسبه و گزارش کنید.

و) نمودار ROC را محاسبه و رسم کنید. مساحت زیر این نمودار (AUROC) را نیز محاسبه و گزارش کنید. برای به دست آوردن مقادیر لازم برای رسم این نمودار، می‌توانید مرز تصمیم را به صورت $g(x) = \tau$ تعریف کنید و با تغییر دادن مقدار آستانه τ از $-\infty$ تا $+\infty$ به ازای τ ‌های مختلف true positive rate و false positive rate را محاسبه کنید.

و) موارد "د" و "ه" و "و" را به ازای $\pi_1 = 0.9, \pi_2 = 0.1$ تکرار کنید و تغییرات معیارهای ارزیابی را تحلیل کنید. کدام معیارها در این شرایط مناسب‌تر هستند؟

سوال ۶: (۱۵ نمره)

مسئله بهینه‌سازی زیر را در نظر بگیرید:

$$\min_{x \in \mathbb{R}^2} f(x) = \frac{1}{2} x^T Q x + c^T x$$

subject to $Ax \leq b$

$$Q \in \mathbb{R}^{2 \times 2}, c \in \mathbb{R}^2, A \in \mathbb{R}^{3 \times 2}, b \in \mathbb{R}^3$$

با استفاده از ضرایب لاگرانژ، مسئله دوگان این مسئله را بدست آوردید و آن را بر حسب ضرایب لاگرانژ بنویسید.