به نام خدا

درس یادگیری ماشین
نیم‌سال دوم ۰۳-۰۴
استاد: ربیعی- دهقانیان

دانشکده مهندسی کامپیوتر

آزمونک سوم زمان آزمون: ۳۰ دقیقه

۱. (۴۰ نمره) ادعاهای زیر در مورد **رگرسیون لجستیک** مطرح شده است، هر ادعا را به صورت جداگانه تحلیل کنید و برای هر مورد

- مشخص کنید که **درست یا نادرست** است.
- پاسخ خود را با **دلایل مفهومی یا ریاضی** توجیه کنید.

ادعا ۱: «تابع هزینه غیر محدب است، بنابراین ممکن است به کمینه‌های محلی گیر کند.»
ادعا ۲: «ماتریس هِسین تابع هزینه ثابت است و به داده‌های ورودی وابسته نیست.»
ادعا ۳: «تابع هزینه مشتق‌پذیر نیست، بنابراین نمی‌توان از روش‌های گرادیان استفاده کرد.»
ادعا ۴: «تابع هزینه به‌صورت قطعی محدب است چون هسین همیشه مثبت معین است.»

پاسخ:

ادعا اول: این ادعا نادرست است. تابع هزینه در رگرسیون لجستیک که بر پایه‌ی منفی درست‌نمایی (likelihood) یا همان خطای آنتروپی متقاطع تعریف می‌شود، تابعی محدب است. بنابراین فقط یک کمینه‌ی سراسری وجود دارد و گرادیان نزولی می‌تواند به خوبی به جواب بهینه برسد.

ادعا دوم: این ادعا نادرست است. هسین تابع هزینه در رگرسیون لجستیک برابر است با $H = X^\top SX$ که در آن S یک ماتریس قطری است و عناصر آن به صورت $h_\theta(x^{(i)})(1 - h_\theta(x^{(i)}))$ تعریف می‌شوند. این مقادیر به پارامترهای مدل و داده‌ها وابسته‌اند، بنابراین هسین ثابت نیست.

ادعا سوم: این ادعا نادرست است. تابع هزینه رگرسیون لجستیک از ترکیب توابع سیگموید، لگاریتمی و ضرب داخلی تشکیل شده که همگی مشتق‌پذیر هستند. بنابراین کل تابع نیز مشتق‌پذیر است و استفاده از روش‌هایی مثل گرادیان نزولی کاملاً امکان‌پذیر است.

ادعا چهارم: این ادعا درست است **به‌شرطی‌که** ماتریس ویژگی‌ها (طراحی) یعنی X دارای رتبه کامل باشد. در این صورت، هسین تابع هزینه مثبت معین خواهد بود و این موضوع تضمین‌کننده‌ی محدب بودن قطعی تابع هزینه و وجود کمینه‌ی یکتا است.

۲. (۳۰ نمره) فرض کنید مجموعه‌دادهٔ زیر به شما داده شده است:

شماره نمونه	X_1	X_2	Y
1	5	1	-1
2	-2	-2	$+1$
3	1	-1	$+1$
4	-3	1	-1

می‌خواهیم الگوریتم پرسپترون (Perceptron) را روی این داده اجرا کنید. فرض کنید وزن‌های اولیه به صورت $\theta^T = [0, 0]$، بایاس $b = 0$ می‌باشند و ما تمام نمونه‌ها را به ترتیب شماره‌شان بررسی می‌کنیم.

الگوریتم را تا اولین مرحله ای که تمام نمونه ها به درستی طبقه بندی شوند ادامه دهید و وزن ها را در هر مرحله مشخص کنید.

پاسخ:

$$\begin{aligned}
\theta_1 &= [-5, -1], \quad b_1 = -1 \\
\theta_2 &= [-5, -1], \quad b_2 = -1 \\
\theta_3 &= [-4, -2], \quad b_3 = 0 \\
\theta_4 &= [-1, -3], \quad b_4 = -1
\end{aligned}$$

۳. (۳۰ نمره) در این سوال میخواهیم ثابت کنیم که طبقه بند بیز بهینه است.
فرض کنید مجموعه کلاس ها $Y = \{0, 1\}$ باشد و تابع احتمال پسینِ تعلق یک نقطه به کلاس ۱ را با

$$\eta(x) = \Pr(Y = 1 \mid X = x)$$

نشان می‌دهیم.
خطا را بصورت زیر تعریف می‌کنیم:

$$R(h) = \mathbb{E}_{XY}\left(\mathbb{I}_{h(X)\neq Y}\right)$$

که h یک طبقه بند دلخواه است. ثابت کنید که h_{bayes} این عبارت را کمینه می‌کند.
راهنمایی: از تساوی زیر کمک بگیرید:

$$\mathbb{E}_{XY}\left(\mathbb{I}_{h(X)\neq Y}\right) = \mathbb{E}_X \mathbb{E}_{Y|X}\left(\mathbb{I}_{h(X)\neq Y}\right)$$

پاسخ:
طبقه‌بند بیز به صورت زیر تعریف می‌شود:

$$h_{\text{Bayes}}(x) = \begin{cases} 1 & \text{if} \eta(x) = \Pr(Y = 1 \mid X = x) \geq \frac{1}{2} \\ 0 & \text{else} \end{cases}$$

هدف ما کمینه کردن خطای طبقه‌بندی است:

$$R(h) = \mathbb{E}_{XY}\left(\mathbb{I}_{h(X)\neq Y}\right)$$

با استفاده از راهنمایی داده شده:

$$R(h) = \mathbb{E}_X \mathbb{E}_{Y|X}\left(\mathbb{I}_{h(X)\neq Y}\right)$$

برای هر مقدار مشخص از $X = x$، خطای شرطی برابر است با:

$$\mathbb{E}_{Y|X=x}\left(\mathbb{I}_{h(x)\neq Y}\right) = \begin{cases} \eta(x) & \text{if} h(x) = 0 \\ 1 - \eta(x) & \text{if} h(x) = 1 \end{cases}$$

پس طبقه‌بند بهینه آن است که برچسبی را انتخاب کند که احتمال بیشتری داشته باشد:

$$h_{\text{Bayes}}(x) = \arg\min_{y\in\{0,1\}} \Pr(Y \neq y \mid X = x)$$

در نتیجه این طبقه‌بند در هر نقطه x خطای شرطی را کمینه می‌کند، و با میانگین‌گیری روی تمام مقادیر x، کل خطا $R(h)$ را کمینه می‌کند.