



- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روز مشخص شده است.
- همکاری و هم‌فکری شما در انجام تمرین مانعی ندارد اما پاسخ ارسالی هر کس حتما باید توسط خود او نوشته شده باشد.
- در صورت هم‌فکری و یا استفاده از هر منابع خارج درسی، نام هم‌فکران و آدرس منابع مورد استفاده برای حل سوال مورد نظر را ذکر کنید.
- لطفا تصویری واضح از پاسخ سوالات نظری بارگذاری کنید. در غیر این صورت پاسخ شما تصحیح نخواهد شد.
- موضوعات تمرین: رگرسیون خطی، دسته‌بندی خطی، رگرسیون و دسته‌بندی با دیدگاه احتمالاتی

سوالات نظری (۷۰ نمره)

۱. (۱۰ نمره) **رگرسیون خطی** - یک مدل رگرسیون خطی با p پارامتر را در نظر بگیرید. به کمک روش کمترین مربعات، مدل را با مجموعه داده‌های تمرینی $(x_1, y_1), \dots, (x_N, y_N)$ که به صورت تصادفی از جمعیت انتخاب شده‌اند، آموزش می‌دهیم. فرض کنید $\hat{\beta}$ تخمین کمترین مربعات باشد. فرض کنید یک سری داده تست $(\bar{x}_1, \bar{y}_1), \dots, (\bar{x}_M, \bar{y}_M)$ به صورت تصادفی از همان جمعیتی که داده‌های تمرین را انتخاب کرده بودیم، انتخاب کرده‌ایم. اگر داشته باشیم: $R_{tr}(\beta) = \frac{1}{N} \sum_1^N (y_i - \beta^T x_i)^2$ و $R_{te}(\beta) = \frac{1}{M} \sum_1^M (\bar{y}_i - \beta^T \bar{x}_i)^2$ نشان دهید:

$$E[R_{tr}(\hat{\beta})] \leq E[R_{te}(\hat{\beta})]$$

۲. (۱۰ نمره) **رگرسیون خطی** - در رگرسیون لاسو، بردار وزن اپتیمال به صورت زیر بدست می‌آید:

$$\omega^* = \operatorname{argmin}_{\omega} J_{\lambda}(\omega)$$

به طوری که:

$$J_{\lambda} = \frac{1}{2} \|y - X\omega\|_2^2 + \lambda \|\omega\|_1$$

که در آن $X \in R^{n \times d}$ و روی داده‌ها whitening انجام داده باشیم. (یعنی $X^T X = I$) نشان دهید که عمل whitening روی داده‌ها باعث می‌شود که ویژگی‌ها از هم مستقل شوند به طوری که ω_i^* به تنهایی از i امین ویژگی نتیجه شود. برای اثبات این، ابتدا نشان دهید که J_{λ} می‌تواند به صورت زیر نوشته شود:

$$J_{\lambda}(\omega) = g(y) + \sum_{i=1}^d f(X_{:,i}; y; \omega_i; \lambda)$$

که $X_{:,i}$ نشان دهنده i امین ستون ماتریس X است.

(ب) اگر $\omega_i \geq 0$ باشد، ω_i را پیدا کنید.

پ) اگر $\omega_i < 0$ باشد، ω_i را پیدا کنید.

ت) با توجه به قسمت‌های قبل، در چه شرایطی ω_i صفر می‌شود؟ این شرایط چگونه قابل اعمال است؟
 ث) همانطور که می‌دانید، در رگرسیون ریب، عبارت نرمال‌سازی در تابع هزینه به صورت $\frac{1}{2} \lambda ||\omega||^2$ ظاهر می‌شود. در این حالت، چه زمانی ω_i صفر می‌شود؟ تفاوت این حالت و حالت قبلی چیست؟

۳. (۱۰ نمره) **دسته‌بندی خطی** - قانون به روزرسانی بردار وزن در پرسپترون را در نظر بگیرید:

$$\text{if } x^{(t)} \text{ is misclassified then : } \omega^{(t+1)} = \omega^{(t)} + \eta x^{(t)} y^{(t)}$$

نشان دهید که در دسته‌بند پرسپترون، بردار وزن ω را می‌توان به صورت ترکیب خطی داده‌ها $x^{(i)}$ نوشت. ضرایب α_i را در ترکیب خطی $\omega = \sum_{i=1}^N \alpha_i x^{(i)}$ مشخص کنید.

۴. (۱۰ نمره) **دسته‌بندی خطی** - یک مدل Multinomial Naive Bayes را برای مسئله دسته‌بندی دو کلاس روی داده‌های متنی در نظر بگیرید. فرض کنید تعداد کل کلمات در دیکشنری ما (تعداد کل ویژگی‌های مدل) برابر d باشد. برای یک نمونه ورودی متنی x مقادیر c_1, \dots, c_d بردار ویژگی‌ها را می‌سازند. به عبارتی هر c_i نشان‌دهنده تعداد دفعاتی است که کلمه i ام در عبارت ما آمده است. پارامترهای این مدل به شکل زیر می‌باشند. (y خروجی مدل یا همان کلاس نمونه است.)

$$P_y = P(y = 1)$$

$$P_{i|y=1} = P(\text{word } i \text{ appears in a specific document position} | y = 1)$$

$$P_{i|y=0} = P(\text{word } i \text{ appears in a specific document position} | y = 0)$$

الف) عبارتی برای احتمال شرطی $P(y = 1 | x)$ برای نمونه متنی x برحسب پارامترهای مدل بنویسید.
 ب) نشان دهید که مرز تصمیم مدل آموزش دیده شده خطی است.
 ج) نشان دهید احتمال شرطی نوشته شده در بخش الف یک تابع logistic است:

$$P(y = 1 | x) = \frac{1}{1 + e^{-(\theta^T x + \theta_0)}}$$

۵. (۱۰ نمره) **دسته‌بندی خطی** - یک مسئله دسته‌بندی سه کلاس در دو بعد با توزیع‌های زیر در نظر بگیرید:

$$P(x | \omega_1) = N(0, I)$$

$$P(x | \omega_2) = N([1 \ 1]^T, I)$$

$$P(x | \omega_3) = 0.5 \times N([0.5 \ 0.5]^T, I) + 0.5 \times N([-0.5 \ 0.5]^T, I)$$

$$P(\omega_1) = P(\omega_2) = p(\omega_3)$$

الف) با محاسبه‌ی احتمال پسین، نقطه‌ی $x = [0.3 \ 0.3]^T$ را برای حالت کمترین احتمال خطا کلاسه‌بندی نمایید.

ب) فرض کنید برای یک نقطه‌ی خاص، ویژگی اول را نداریم (یعنی $x = [* \ 0.3]^T$) این نقطه را کلاسه‌بندی نمایید.

۶. (۱۰ نمره) **رگرسیون با دیدگاه احتمالاتی** - در مسئله رگرسیون خطی، قصد داریم به نمونه‌های آموزشی، وزن‌های متفاوتی نسبت دهیم. به بیان دقیق‌تر، می‌خواهیم مقدار $J(\theta)$ را کمینه کنیم که به صورت زیر تعریف می‌گردد:

$$J(\theta) = \frac{1}{2} \sum_{i=1}^m \omega(i) (\theta^T x^{(i)} - y^{(i)})^2$$

(آ) نشان دهید ماتریس W موجود است؛ به طوری که داریم:

$$J(\theta) = (X\theta - y)^T W (X\theta - y)$$

(ب) با محاسبه $\nabla_{\theta} J(\theta)$ و برابر قرار دادن آن با صفر، مقدار θ ای را که $J(\theta)$ را کمینه می‌کند، بیابید. (توجه: در حالتی که همه وزن‌ها یکسان باشند، می‌دانیم $\theta^* = (X^T X)^{-1} X^T y$ جوابتان برای این قسمت باید یک فرم بسته باشد که تابعی از X و y است).

(ج) فرض کنید مجموعه داده $\{(x^{(i)}, y^{(i)}) : i = 1, 2, \dots, m\}$ شامل m نمونه مستقل داده شده است. قصد داریم $y^{(i)}$ ها را گونه‌ای مدل کنیم که گویی از توزیع‌های شرطی با سطوح مختلفی از واریانس گرفته شده‌اند. به طور مشخص، فرض کنید داریم:

$$p(y^{(i)} | x^{(i)}; \theta) = \frac{1}{\sqrt{2\pi}\sigma^{(i)}} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2(\sigma^{(i)})^2}\right)$$

به بیان دیگر، $y^{(i)}$ از یک توزیع گاوسی با میانگین $\theta^T x^{(i)}$ و واریانس $(\sigma^{(i)})^2$ می‌آید؛ $\sigma^{(i)}$ ها ثابت هستند و مقدارشان مشخص است. نشان دهید که یافتن تخمین بیشینه درست‌نمایی برای θ ، معادل است با حل یک مسئله رگرسیون خطی وزن دار. به طور مشخص مقادیر $\omega^{(i)}$ ها را بر حسب $\sigma^{(i)}$ ها به دست آورید.

۷. (۱۰ نمره) **دسته‌بندی با دیدگاه احتمالی** - یک مسئله دسته‌بندی Naive Bayes را با ۳ کلاس و ۲ ویژگی در نظر بگیرید. یکی از این ویژگی‌ها از توزیع برنولی و دیگری از توزیع گاوسی می‌آید. ویژگی‌ها با $\mathbf{X} = [X_1, X_2]^T$ و کلاس با Y تمایش داده می‌شود.

توزیع ابتدایی به صورت زیر می‌باشد:

$$P[Y = 0] = 0.5, P[Y = 1] = 0.25, P[Y = 2] = 0.25$$

توزیع ویژگی‌ها به صورت زیر است:

$$p_{X_1|Y}(x_1|Y=c) = \text{Ber}(x_1; \theta_c),$$

$$p_{X_2|Y}(x_2|Y=c) = \text{Normal}(x_2; \mu_c, \sigma_c^2),$$

همچنین فرض کنید:

$$\sigma_c^2 = \begin{cases} 1 & \text{if } c = 0 \\ 1 & \text{if } c = 1 \\ 1 & \text{if } c = 2 \end{cases}, \mu_c = \begin{cases} -1 & \text{if } c = 0 \\ 0 & \text{if } c = 1 \\ 1 & \text{if } c = 2 \end{cases}, \theta_c = \begin{cases} 0.5 & \text{if } c = 0 \\ 0.5 & \text{if } c = 1 \\ 0.5 & \text{if } c = 2 \end{cases}$$

الف) $(p_{Y|X_1, X_2}(y|x_1 = 0, x_2 = 0))$ را محاسبه کنید. (جواب باید یک بردار در R^3 باشد که جمع اعضای آن ۱ باشد.)

ب) $(p_{Y|X_1}(y|x_1 = 0))$ را محاسبه کنید.

پ) $(p_{Y|X_2}(y|x_2 = 0))$ را محاسبه کنید.

ت) الگوی یافته شده در جواب‌های قسمت‌های قبل را تحلیل کنید.