یادگیری ماشین
بهار ۱۴۰۳

استاد: فاطمه سیدصالحی

دانشگاه صنعتی شریف

مهلت ارسال: ۲۲ اسفند

تمرین اول

- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روز مشخص‌شده 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، دسته بندی خطی، رگرسیون و دسته بندی 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 $(x_1, y_1), ..., (x_N, y_N)$ که به صورت تصادفی از جمعیت انتخاب شده‌اند، آموزش می‌دهیم. فرض کنید $\hat{\beta}$ تخمین کمترین مربعات باشد. فرض کنید یک سری داده تست $(\bar{x_1}, \bar{y_1}), ..., (\bar{x_M}, \bar{y_M})$ به صورت تصادفی از همان جمعیتی که داده‌های تمرین را انتخاب کرده بودیم، انتخاب کرده‌ایم. اگر داشته باشیم: $R_{tr}(\beta) = \frac{1}{N}\Sigma_1^N(y_i - \beta^T x_i)^2$ و $R_{te}(\beta) = \frac{1}{M}\Sigma_1^M(\bar{y_i} - \beta^T \bar{x_i})^2$. 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:

$$\omega^* = argmin J_\lambda(\omega)$$

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 $X^T X = I$)

آ) 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 $\omega_i > ۰$ 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 ریج، عبارت نرمال‌سازی در تابع هزینه به صورت $\frac{۱}{۲}\lambda||\omega||^۲_۲$ 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:

$$if\ x^{(t)}\ is\ misclassified\ then : \omega^{(t+۱)} = \omega^{(t)} + \eta x^{(t)} y^{(t)}$$

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 $\omega = \sum_{i=۱}^{N} \alpha_i x^{(i)}$ مشخص کنید.

۴. (۱۰ نمره) **دسته‌بندی خطی -** یک مدل Multinomial Naive Bayes را برای مسئله دسته‌بندی دو کلاسه 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 $c_۱, ..., c_d$ 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.)

$$P_y = P(y = ۱)$$

$$P_{i|y=۱} = P(word\ i\ appears\ in\ a\ specific\ document\ position|y = ۱)$$

$$P_{i|y=۰} = P(word\ i\ appears\ in\ a\ specific\ document\ position|y = ۰)$$

الف) عبارتی برای احتمال شرطی $P(y = ۱|x)$ برای نمونه متنی x برحسب پارامترهای مدل بنویسید.

ب) نشان دهید که مرز تصمیم مدل آموزش دیده‌شده خطی است.

ج) نشان دهید احتمال شرطی نوشته شده در بخش الف یک تابع logistic است:

$$P(y = ۱|x) = \frac{۱}{۱ + e^{-(\theta^T x + \theta_۰)}}$$

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

$$P(x|\omega_۱) = N(۰, I)$$

$$P(x|\omega_۲) = N([۱ \quad ۱]^T, I)$$

$$P(x|\omega_۳) = ۰/۵ \times N([۰/۵ \quad ۰/۵]^T, I) + ۰/۵ \times N([-۰/۵ \quad ۰/۵]^T, I)$$

$$P(\omega_۱) = P(\omega_۲) = p(\omega_۳)$$

الف) با محاسبه‌ی احتمال پسین، نقطه‌ی $x = [۰/۳ \quad ۰/۳]^T$ را برای حالت کمترین احتمال خطا کلاسه‌بندی نمایید.

ب) فرض کنید برای یک نقطه‌ی خاص، ویژگی اول را نداریم (یعنی $x = [* \quad ۰/۳]^T$) 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:

$$J(\theta) = \frac{۱}{۲}\sum_{i=۱}^{m}\omega(i)(\theta^T x^{(i)} - y^{(i)})^۲$$

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 قراردادن آن با صفر، مقدار θ ای را که $J(\theta)$ را کمینه می‌کند، بیابید.

(توجه: در حالتی که همه وزن‌ها یکسان باشند، می‌دانیم $\theta^* = (X^TX)^{-۱}X^Ty$. جوابتان برای این قسمت باید یک فرم بسته باشد که تابعی از W X، و y است).

ج) فرض کنید مجموعه داده $\{(x^{(i)}, y^{(i)}) : i = ۱, ۲, ..., m\}$ 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:

$$p(y^{(i)}|x^{(i)};\theta) = \frac{۱}{\sqrt{۲\pi}\sigma^{(i)}} exp(-\frac{(y^{(i)} - \theta^T x^{(i)})^۲}{۲(\sigma^{(i)})^۲})$$

به بیان دیگر، $y^{(i)}$ از یک توزیع گاوسی با میانگین $\theta^T x^{(i)}$ و واریانس $(\sigma^{(i)})^۲$ 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 $\mathbf{X} = [X_۱, X_۲]^T$ و کلاس با Y تمایش داده میشود.

توزیع ابتدایی به صورت زیر میباشد:

$$P[Y = ۰] = ۰/۵, P[Y = ۱] = ۰/۲۵, P[Y = ۲] = ۰/۲۵$$

توزیع ویژگی‌ها به صورت زیر است:

$$p_{X_۱|Y}(x_۱|Y = c) = Ber(x_۱;\theta_c),$$
$$p_{X_۲|Y}(x_۲|Y = c) = Normal(x_۲;\mu_c, \sigma_c^۲),$$

همچنین فرض کنید:

$$\sigma_c^۲ = \begin{cases} ۱ & \text{if } c = ۰ \\ ۱ & \text{if } c = ۱ \\ ۱ & \text{if } c = ۲ \end{cases} ،\ \mu_c = \begin{cases} -۱ & \text{if } c = ۰ \\ ۰ & \text{if } c = ۱ \\ ۱ & \text{if } c = ۲ \end{cases} ،\ \theta_c = \begin{cases} ۰/۵ & \text{if } c = ۰ \\ ۰/۵ & \text{if } c = ۱ \\ ۰/۵ & \text{if } c = ۲ \end{cases}$$

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.