



دانشگاه صنعتی شریف

دانشکده مهندسی کامپیوتر

یادگیری ماشین

پاییز ۱۴۰۳

استاد: علی شریفی زارچی

مسئول تمرین: عرفان سلیمان

مهلت ارسال نهایی: ۴ آبان

تمرین اول

مهلت ارسال امتیازی: ۲۷ مهر

- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روزهای مشخص شده است.
- در طول ترم، برای هر تمرین می‌توانید تا ۵ روز تأخیر مجاز داشته باشید و در مجموع حداکثر ۱۵ روز تأخیر مجاز خواهید داشت. توجه داشته باشید که تأخیر در تمرین‌های عملی و تئوری به صورت جداگانه محاسبه می‌شود و مجموع تأخیر هر دو نباید بیشتر از ۱۵ روز شود. پس از اتمام زمان مجاز، دو روز اضافی برای آپلود غیرمجاز در نظر گرفته شده است که در این بازه به ازای هر ساعت تأخیر، ۲ درصد از نمره تمرین کسر خواهد شد.
- اگر بخش عملی یا تئوری تمرین را قبل از مهلت ارسال امتیازی آپلود کنید، ۲۰ درصد نمره اضافی به آن بخش تعلق خواهد گرفت و پس از آن، ویدئویی تحت عنوان راهنمایی برای حل تمرین منتشر خواهد شد.
- حتماً تمرین‌ها را بر اساس موارد ذکر شده در صورت سوالات حل کنید. در صورت وجود هرگونه ابهام، آن را در صفحه تمرین در سایت کوئرا مطرح کنید و به پاسخ‌هایی که از سوی دستیار آموزشی مربوطه ارائه می‌شود، توجه کنید.
- 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.
- فایل پاسخ‌های سوالات نظری را در قالب یک فایل pdf به فرمت `HW1_T_[STD_ID].pdf` آماده کنید و برای سوالات عملی، هریک را در یک فایل zip جداگانه قرار دهید و فایل zip اول را به فرمت `HW1_P1_[STD_ID].zip` و فایل zip دوم را به فرمت `HW1_P2_[STD_ID].zip` نامگذاری کرده و هرکدام را به صورت جداگانه آپلود کنید.
- گردآورندگان تمرین: مبینا سلیمی‌پناه، محمد مولوی، امیرعلی لقمانی، فاطمه السادات موسوی، عرشیا قارونی

سوالات نظری (۱۰۰+۱۰ نمره)

۱. (۲۵ نمره) به سوالات زیر پاسخ کوتاه دهید.
 - ✓ الف) چرا از تابع softmax اغلب برای مسائل دسته‌بندی استفاده می‌شود؟
 - ✓ ب) بالا بودن واریانس در مدل چه معنایی دارد؟ یک روش ممکن برای کاهش واریانس در مدل خود بیان کنید.
 - ✓ پ) چرا در حالتی که تمام ویژگی‌ها تا حد خوبی با خروجی مرتبط هستند، رگرسیون Ridge به رگرسیون Lasso ترجیح داده می‌شود؟
 - ✓ ت) چگونه رگولاریزیشن L_2 در classifierهای خطی بر روی تعادل بایاس-واریانس تأثیر می‌گذارد؟
۲. (۲۵ نمره) در یک مسئله رگرسیون خطی داریم:

underline -> vector

$$y = \underline{w}^T \underline{x}, \quad \underline{x} \in \mathbb{R}^L, \quad y \in \mathbb{R}$$

$$X = [\underline{x}_1, \underline{x}_2, \dots, \underline{x}_N], \quad \underline{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}, \quad \underline{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_L \end{pmatrix}$$

$N \times 1$ $L \times 1$

$X = \begin{bmatrix} & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \end{bmatrix}, \quad L \times N$

$$\omega = (X^T X)^{-1} X^T y$$

الف) اگر رگرسیون را فقط بر روی ویژگی j انجام دهیم، نشان دهید که:

$$w_j = \frac{X_j y}{X_j X_j^T}$$

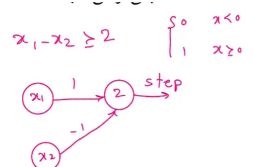
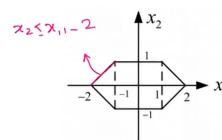
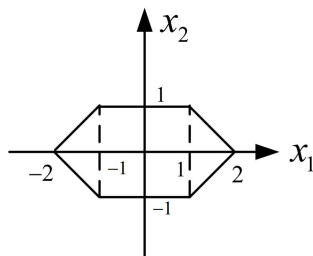
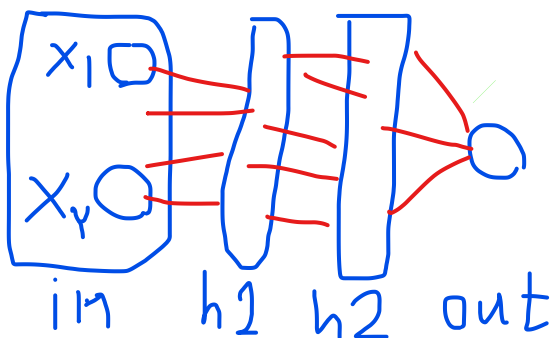
که X_j سطر j ماتریس داده‌ها است.

تغییر در
ماتریس
کوواریانس

ب) فرض کنید ویژگی‌ها مستقل هستند (یعنی سطرهاى ماتریس داده‌ها مستقل هستند). ثابت کنید که پارامترهای بهینه از آموزش رگرسیون بر روی همه ویژگی‌ها با پارامترهای بهینه حاصل از آموزش روی هر ویژگی به طور مستقل یکسان است.

پ) فرض کنید $y = \underline{w}^T \underline{x} + \bar{y}$ و رگرسیون را فقط بر روی ویژگی j انجام دهیم. w_j و w را بدست آورید.

۳. (۲۵ نمره) یک شبکه عصبی با دو گره در لایه ورودی و دو لایه مخفی و تابع فعالسازی پله داریم. وزن‌ها و بایاس‌های این شبکه را به گونه‌ای تعیین کنید که در ناحیه داخل ۶ ضلعی و روی اضلاع خروجی شبکه ۱ و در باقی نواحی ۰ باشد.



۴. (۲۵ نمره) در یک مسئله دسته‌بندی دو کلاسه (binary classification) از رگرسیون لاجستیک با تابع هزینه cross entropy استفاده کرده‌ایم. تابع هزینه برای یک نقطه از داده‌ها به صورت زیر تعریف می‌شود:

$$L = - \sum_{i=1}^n y_i \log(\hat{y}_i)$$

که در آن y بردار برچسب‌های واقعی و \hat{y} احتمالات پیش‌بینی شده با استفاده از تابع softmax به صورت زیر است:

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^n e^{z_j}}$$

قلمه از خبری

الف) مشتق تابع softmax را نسبت به z_k برای دو حالت $k \neq i$ و $k = i$ به دست آورید.

ب) با استفاده از مشتق تابع softmax، مشتق تابع هزینه cross entropy را نسبت به z_k محاسبه نمایید.

۵. (امتیازی-۱۰ نمره) به سوالات زیر در مورد رگرسیون Ridge پاسخ دهید:

الف) نشان دهید که به ازای مقادیر $\lambda > 0$ واریانس ضرایب Ridge از واریانس ضرایب رگرسیون خطی کوچکتر است:

$$\text{Var}(\hat{\beta}^{LS}) > \text{Var}(\hat{\beta}^{Ridge}(\lambda))$$

$$\omega = (X^T X)^{-1} X^T y \quad \text{و} \quad (X^T X + \lambda I)^{-1} X^T y$$

tr = trace = sum of main diameter

✓ (ب) نشان دهید به ازای مقادیر $\lambda > 0$ رابطه زیر برقرار است:

$$\text{tr} \left\{ \text{Var}[\hat{Y}(\lambda)] \right\} = \sigma^2 \sum_{j=1}^p (D_x)_{jj}^2 [(D_x)_{jj}^2 + \lambda]^{-2}$$

در این رابطه D_x یک ماتریس قطری است که شامل مقادیر تکین X است. ← singular values

سوالات عملی (۱۰۰ نمره)

۱. (۱۰۰ نمره) برای حل سوالات به *notebook* های ضمیمه شده مراجعه کنید.

- (۱) (۶۰ نمره) برای پاسخ به تمرین عملی اول ابتدا فایل نوت‌بوک قرار گرفته را باز کنید و سپس مراحل را مطابق آنچه که از شما خواسته شده انجام دهید. در نهایت، مقادیر پیش‌بینی شده برای دیتاست `InsuranceData_test.csv` را در یک فایل به نام `submission.csv` که شامل یک ستون به نام `charges` میباشد، ذخیره کنید. فایل خروجی و فایل نوت‌بوک را در یک فایل `zip` قرار دهید و آن را به فرمت `HW\ P\ [STD_ID].zip` نامگذاری کرده و آپلود کنید.
- توجه بفرمایید این سوال دارای داوری خودکار می‌باشد و ۱۵ نمره از ۶۰ نمره به این قسمت تعلق دارد.
- (۲) (۴۰ نمره) برای پاسخ به تمرین عملی دوم تنها کافی است نوت‌بوک `Perceptron.ipynb` را تکمیل کرده و سپس مطابق با فرمت ذکر شده آپلود کنید.