به نام خدا

دانشکده مهندسی کامپیوتر

امنیت و حریم خصوصی در یادگیری ماشین
(۴۰۸۱۶)(نیم‌سال دوم سال تحصیلی ۱۴۰۲-۱۴۰۱)
استاد درس: دکتر امیرمهدی صادق‌زاده
دستیاران آموزشی: مهدی غزنوی، زینب گلگونی، الهه فرشادفر، محمدرضا کاظمی، حمید دشتبانی

تمرین اول — مهلت تحویل: ساعت ۲۳:۵۹ شنبه ۲۰ اسفند ۱۴۰۱

نکات و قواعد

۱. سوالات خود را زیر پیام مربوطه در Quera مطرح نمایید.
۲. محل بارگذاری تمرین تا یک هفته پس از مهلت ارسال باز خواهد بود. در طول ترم، در مجموع می‌توانید از ۲۱ روز تاخیر مجاز به صورت ساعتی استفاده کنید و پس از آن به ازای هر روز ۲۰ درصد جریمه بر روی نمره‌ی کسب شده اعمال خواهد شد.
۳. لطفا مطابق تاکید پیشین، حتما آداب‌نامه‌ی انجام تمرین‌های درسی را رعایت نمایید. در صورت تخطی از آیین‌نامه، در بهترین حالت مجبور به حذف درس خواهید شد.
۴. در صورتی که پاسخ‌های سوالات نظری را به صورت دست‌نویس آماده کرده‌اید، لطفا تصاویر واضحی از پاسخ‌های خود ارسال کنید. در صورت ناخوانا بودن پاسخ ارسالی، نمره‌ای به پاسخ ارسال شده تعلق نمی‌گیرد.
۵. همه‌ی فایل‌های مربوط به پاسخ خود را در یک فایل فشرده و با نام **SPML_HW1_StdNum_FirstName_LastName** ذخیره کرده و ارسال نمائید.

سوال ۱ تابع هزینه (۱۰ نمره)

یکی از توابع هزینه پیشنهادی در آموزش مدل، تابع هزینه Mean Absolute Error یا MAE می‌باشد. این تابع هزینه برای یک مجموعه داده‌ی Nتایی و مدل f_θ به صورت زیر تعریف می‌شود:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N} |y_i - f_\theta(x_i)|$$

در این معادله، N تعداد داده‌ها، x_i iـامین ورودی و y_i برچسب معادل آن است.

(الف) (۳ نمره) فرض کنید یک مدل f_θ به صورت زیر داریم:

$$f_\theta(x_i) = x_i\theta^2 - x_i\theta$$

با فرض $N = 3$، برای دادگان آموزشی $(1,3)$، $(2,0)$ و $(3,6)$ (هر زوج مرتب به صورت (x_i, y_i) است) معادله تابع $\mathcal{L}$ را بر حسب θ به دست آورید سپس نمودار $\mathcal{L}$ نسبت به θ را رسم کنید. نمودار مربوطه چند کمینه محلی و سراسری دارد؟ آیا با فرض انتخاب اندازه‌ی گام بسیار کوچک و با فرض این که زمان کافی برای همگرایی به الگوریتم گرادیان کاهشی بدهیم، می‌توان با اطمینان گفت که به کمک این الگوریتم همواره می‌توان به مقدار بهینه سراسری برای θ رسید؟ توضیح دهید.

(ب) (۳ نمره) این بار فرض کنید از یک نورون برای f_θ استفاده می‌کنیم. تابع فعالساز این نورون را به صورت زیر در نظر بگیرید:

$$a(x) = \ln(1 + e^x)$$

خروجی این نورون به صورت زیر خواهد بود:

$$y = \ln(1 + e^{x\theta})$$

با در نظر گرفتن تابع هزینه Mean Squared Error (MSE) و برای داده‌های $(1,0)$، $(-1,3)$ و $(3,4)$ تابع هزینه را نسبت به پارامتر θ رسم کنید (برای این کار می‌توانید از برنامه‌های رسم نمودار استفاده کنید). سوالات مطرح شده در قسمت الف را با در نظر گرفتن این نمودار پاسخ دهید.

(ج) (۴ نمره) با توجه به دو قسمت قبل، انتظار دارید شرایط کمینه‌های محلی و سراسری برای تابع هزینه در شبکه‌های عصبی ژرف عموما به چه صورت باشد؟ برای مقابله با اثرات منفی این پدیده، چه راهکارهایی در روند بهینه‌سازی مدل‌های ژرف به کار گرفته می‌شود؟

سوال ۲ آموزش مدل (۱۰ نمره)

با در نظر گرفتن استفاده از یک مدل Logistic Regression در یک وظیفه دسته‌بندی به سوالات زیر پاسخ دهید.

(الف) (۳ نمره) نرمال‌سازی ورودی چه اثری و چگونه بر آموزش مدل و دقت نهایی دارد؟

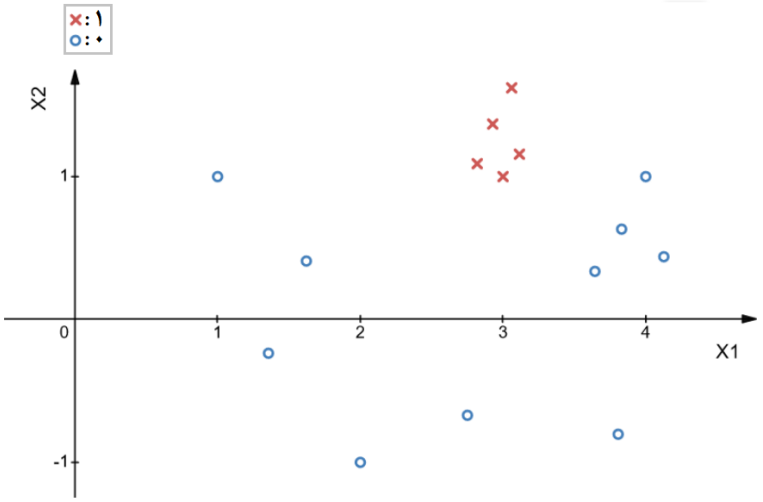
(ب) (۳ نمره) استفاده از منظم‌ساز چه اثری و چگونه بر آموزش و دقت نهایی دارد؟

(ج) (۴ نمره) ضعف‌ها و قوت‌های نسبی دو شبکه‌ی زیر چیست؟

- ۲۰ لایه‌ی خطی
- ۳ لایه با فعال‌ساز غیرخطی

سوال ۳ شبکه Perceptron (۱۰ نمره)

شکل زیر را در نظر بگیرید. این شکل نمودار مربوط به یک مجموعه از داده‌ها را نشان می‌دهد که در دو کلاس دسته‌بندی شده‌اند. هر داده با دو بعد (x_1, x_2) نمایش داده می‌شود و نقاط به شکل ضربدر مربوط به داده‌های کلاس ۱ و دایره‌های توخالی مربوط به کلاس ۰ می‌شوند. با توجه به شکل به سوالات زیر پاسخ دهید.



(الف) (۲ نمره) آیا می‌توان به صورت خطی داده‌های موجود در شکل را دسته‌بندی کرد؟

(ب) (۸ نمره) یک شبکه Perceptron با حداقل تعداد لایه‌های ممکن و وزن‌های دلخواه برای دسته‌بندی داده‌های موجود در شکل ارائه دهید.

سوال ۴ Cross-Entropy و softmax (۱۲ نمره)

فرض کنید یک شبکه عصبی برای دسته‌بندی با M دسته داریم. لایه انتهایی این شبکه عصبی از فعالساز softmax استفاده می‌کند که به این صورت تعریف می‌شود:

$$\text{softmax}\,(x_j) = \frac{e^{x_j}}{\sum_{m=1}^{M} e^{x_m}}$$

در این رابطه، x_j ها ویژگی‌های استخراج شده در شبکه هستند که در انتهای شبکه، به لایه فعالساز به عنوان ورودی داده شده‌اند. با در نظر گرفتن تابع هزینه زیر، رابطه $\frac{\partial \mathcal{L}(\hat{y},y)}{\partial x_j}$ را به دست آورید:

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

سوال ۵ انتشار گرادیان به عقب (۱۵ نمره)

یک شبکه عصبی با دو لایه برای دسته‌بندی دو کلاسه به شکل زیر داریم:

$$
\begin{aligned}
z_1 &= W_1 x^{(i)} + b_1 \\
a_1 &= \text{Leaky-ReLU}\,(z_1) \\
z_2 &= W_2 a_1 + b_2 \\
\hat{y}^{(i)} &= \sigma\,(z_2) \\
\mathcal{L} &= \tfrac{1}{m}\left[\textstyle\sum_{i=1}^{m} -y^{(i)} * \log\left(\hat{y}^{(i)}\right) - \left(1-y^{(i)}\right) * \log\left(1-\hat{y}^{(i)}\right)\right]
\end{aligned}
$$

در عبارات بالا، $x^{(i)}$ iـامین داده و $y^{(i)}$ برچسب معادل آن است. با این فرض که $x^{(i)} \in \mathbb{R}^{n\times 1}$ باشد به سوالات زیر پاسخ دهید:

(الف) (۳ نمره) ابعاد تمام وزن‌ها و اریبی‌های شبکه را مشخص کنید.

(ب) (۲ نمره) اگر ۲۰۰۰ داده‌ی ورودی را به صورت یکجا و در قالب ماتریس‌های X و Y به شبکه بدهیم، ابعاد وزن‌ها و اریبی‌ها را بازنویسی کنید.

(ج) (۱۰ نمره) عبارات $\frac{\partial \mathcal{L}}{\partial \hat{y}^{(i)}}$, $\frac{\partial \hat{y}^{(i)}}{\partial z_2}$, $\frac{\partial z_2}{\partial a_1}$, $\frac{\partial a_1}{\partial z_1}$, $\frac{\partial z_1}{\partial W_1}$ را به دست آورید.

سوال ۶ منظم‌سازی (۱۳ نمره)

یکی از منظم‌سازهای پر استفاده در آموزش مدل‌ها، منظم‌ساز L_2 می‌باشد که به صورت زیر تعریف می‌شود:

$$\Omega(\mathbf{w}) = \lambda \|\mathbf{w}\|_2^2$$

(الف) (۵ نمره) چرا از منظم‌ساز L_2 با عنوان weight decay نیز یاد می‌شود؟

(ب) (۳ نمره) با در نظر گرفتن تابع هزینه MSE و استفاده از منظم‌ساز L_2 وزن بهینه را (به فرم بسته) به دست آورید:

$$\mathcal{L} = \frac{1}{N}\|X\mathbf{w} - y\|^2 + \lambda \|\mathbf{w}\|_2^2$$

(ج) (۵ نمره) آیا عبارت به دست آمده در قسمت الف برای وزن بهینه، همواره قابل محاسبه خواهد بود؟ در صورت مثبت بودن پاسخ آن را اثبات کنید و در صورت منفی بودن، برای شرایطی که وزن بهینه قابل محاسبه نخواهد بود یک مثال بزنید.

سوال ۷ (عملی) شبکه عصبی با NumPy (۳۰ نمره)

در این سوال قرار است یک مدل شبکه عصبی را با استفاده از NumPy پیاده‌سازی کنید. برای حل این سوال به فایل nn_numpy.ipynb مراجعه کنید.

موفق باشید