



یادگیری عمیق

نیم سال دوم ۰۳-۰۲
مدرس: مهدیه سلیمانی

تمرین اول

- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روز مشخص شده است.
- در طول ترم امکان ارسال با تاخیر تمرین‌های نظری بدون کسر نمره تا سقف ۵ روز و تمرین‌های عملی تا سقف ۱۰ روز وجود دارد. محل بارگزاری جواب تمرین‌های نظری بعد از ۳ روز و تمرین‌های عملی بعد از ۵ روز بسته خواهد شد و پس از گذشت این مدت، پاسخ‌های ارسال شده پذیرفته نخواهند شد.
- هم‌فکری در انجام تمرین مانعی ندارد، فقط توجه داشته باشید که پاسخ تمرین حتما باید توسط خود شخص نوشته شده باشد. همچنین در صورت هم‌فکری در هر تمرین، در ابتدای جواب تمرین نام افرادی که با آن‌ها هم‌فکری کرده اید را حتما ذکر کنید.
- برای پاسخ به سوالات نظری در صورتی که از برگه خود عکس تهیه می‌کنید، حتما توجه داشته باشید که تصویر کاملا واضح و خوانا باشد. در صورتی که خوانایی کافی را نداشته باشد، تصحیح نخواهد شد.
- محل بارگذاری سوالات نظری و عملی در هر تمرین مجزا خواهد بود. به منظور بارگذاری بایستی تمرین تئوری در یک فایل pdf با نام `HW1_[First-Name]_[Last-Name]_[Student-Id].pdf` و تمرین عملی نیز در یک فایل مجزای زیپ با نام `HW1_[First-Name]_[Last-Name]_[Student-Id].zip` بارگذاری شوند.
- در صورت وجود هرگونه ابهام یا مشکل، در کوثرای درس آن مشکل را بیان کنید و از پیغام دادن مستقیم به دستیاران آموزشی خودداری کنید.
- طراحان این تمرین: آقایان حمیدرضا امیرزاده، علی رازقندی، حمیدرضا یعقوبی

بخش نظری (۶۰ نمره)

سوال اول: مشتق جزئی (۱۲ نمره)

فرض کنید یک ماتریس دلخواه $A_{m \times n}$ داریم. همچنین بردارهای x و y که به ترتیب m و n بعدی هستند و به صورت $y = Ax$ به هم مرتبط می‌شوند. مشتق جزئی y نسبت به x به صورت زیر تعریف می‌شود.

$$\frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \dots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \dots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \dots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}$$

از روی فرم باز شده ی یک ضرب ماتریسی یعنی $y_i = \sum_{k=1}^n a_{ik}x_k$ عبارات زیر را بدست آورید.

- $\frac{\partial y}{\partial x} = A$
- اگر x یک تابع از z و A مستقل از z باشد، ثابت کنید $\frac{\partial y}{\partial z} = A \frac{\partial x}{\partial z}$
- اگر تعریف کنیم $\alpha = y^T A x$ ثابت کنید $\frac{\partial \alpha}{\partial x} = y^T A$ و $\frac{\partial \alpha}{\partial y} = x^T A^T$

- اگر x و y دو بردار n بعدی که تابعی از متغیر z اند و $\alpha = y^T x$ ثابت کنید: $\frac{\partial \alpha}{\partial z} = x^T \frac{\partial y}{\partial z} + y^T \frac{\partial x}{\partial z}$
- اگر $A_{m \times m}$ یک ماتریس non singular باشد که درایه های آن تابع هایی از مقدار اسکالر α باشد، ثابت کنید: $\frac{\partial A^{-1}}{\partial \alpha} = -A^{-1} \frac{\partial A}{\partial \alpha} A^{-1}$

سوال دوم: ماتریس Hessian (۸ نمره)

نشان دهید ماتریس Hessian یک تبدیل مثل $y = \psi(u, v, z)$ را می توان به صورت ماتریس ژاکوبی گرادیان این تبدیل نوشت. توجه کنید که متغیرهای u, v, z تک بعدی و y نیز تابعی بر حسب آنها است.

سوال سوم: جلوگیری از نامتقارن شدن (۱۰ نمره)

برخی از انواع مجموعه دادگان، مانند برخی از انواع سری زمانی یا تصاویر صورت، دارای یک شبه تقارن ذاتی هستند. ابتدا تحقیق کنید که منظور از این تقارن چیست و چگونه به آموزش بهتر یک مدل دسته بندی کمک می کنند. حال فرض کنید که یک عکس کوچک با ابعاد 1×2 در اختیار داریم. ترم رگولاریزیشن L_2 به صورت $R(\omega) = \omega^T \omega = \omega^T I \omega$ تعریف می شود. ماتریس S را بیابید به گونه ای که $R(\omega) = \omega^T S \omega$ از نامتقارن شدن وزن ها جلوگیری کند.

سوال چهارم: Backpropagation Basics (۱۰ نمره)

فرض کنید شبکه عصبی دو لایه مانند زیر داریم

$$\begin{aligned} z_1 &= W_1 x^{(i)} + b_1 \\ a_1 &= \text{ReLU}(z_1) \\ z_2 &= W_2 a_1 + b_2 \\ \hat{y}^{(i)} &= \sigma(z_2) \\ L^{(i)} &= y^{(i)} * \log(\hat{y}^{(i)}) + (1 - y^{(i)}) * \log(1 - \hat{y}^{(i)}) \\ J &= \frac{1}{m} \sum_{i=1}^m L^{(i)} \end{aligned}$$

توجه کنید که $x^{(i)}$ نشان دهنده یک نمونه ورودی با ابعاد $D_x \times 1$ است. همچنین $y^{(i)}$ برچسب یک نمونه است و به صورت اسکالر می باشد. دیتاست شامل m نمونه است. همچنین z_1 ابعاد $D_{a_1} \times 1$ دارد.

- ابعاد W_1, b_1, W_2, b_2 را بنویسید.
- نتیجه $\partial J / \partial \hat{y}^{(i)}$ را بدست آورید و آن را با δ_1 نشان دهید.
- نتیجه $\partial \hat{y}^{(i)} / \partial z_2$ را بدست آورید و آن را با δ_2 نشان دهید.
- نتیجه $\partial z_2 / \partial a_1$ را بدست آورید و آن را با δ_3 نشان دهید.
- نتیجه $\partial a_1 / \partial z_1$ را بدست آورید و آن را با δ_4 نشان دهید.
- نتیجه $\partial z_1 / \partial W_1$ را بدست آورید و آن را با δ_5 نشان دهید.
- نتیجه $\partial J / \partial W_1$ را با استفاده از نتایج قبلی بدست آورید.

سوال پنجم: بهینه سازی (۲۰ نمره)

۱. تابع زیر که Beale نام دارد را در نظر بگیرید:

$$f(\underline{x}) = (1.5 - x_1 + x_1x_2)^2 + (2.25 - x_1 + x_1x_2^2)^2 + (2.625 - x_1 + x_1x_2^3)^2 \quad (1)$$

- با استدلال مناسب بیان کنید که این تابع محدب است یا نامحدب؟ این مسئله در بهینه‌سازی توابع چه اهمیتی دارد؟
- حال با در نظر گرفتن نقطه شروع $(0, 1)$ ، جهت گرادیان را پیدا کرده و سپس با کمک الگوریتم گرادیان کاهشی، مقدار جدید نقطه شروع پس از یک بار به‌روزرسانی را به دست آورید.

۲. تابع هدف محدب درجه‌دو زیر که در آن Q یک ماتریس مثبت معین (Positive definite) است را در نظر بگیرید:

$$h(x) = \frac{1}{2}x^T Q x + x^T c + b \quad (2)$$

ثابت کنید که الگوریتم بهینه‌سازی گرادیان کاهشی برای چنین توابعی معادل بهینه‌سازی درجه‌دو بردارهای ویژه ماتریس درجه‌دو Q است.

۳. با الگوریتم بهینه‌سازی Adam در درس آشنا شدید. در این الگوریتم قانون به‌روزرسانی هر وزن این است که گرادیان آن را به صورت متناسب با معکوس نرم‌دو گرادیان‌های فعلی و قبلی آن مقیاس کنیم.

- با جایگزینی نرم‌دو مذکور با نرم‌بی‌نهایت به الگوریتم بهینه‌سازی جدیدی برسید. (توجه کنید که توان β_2 را برابر p که $p \rightarrow \infty$ در نظر بگیرید.)

- الگوریتم حاصل را با روش Adam مقایسه و بیان کنید در چه شرایطی استفاده از الگوریتم حاصل شده بهتر است؟

بخش عملی (۴۰ نمره)

۱. در این سوال به پیاده‌سازی چندین الگوریتم بهینه‌سازی رایج در حل مسائل یادگیری ژرف پرداخته و عملکرد آن‌ها را برای بهینه‌سازی دو تابع آزمون Ackley و Beale ارزیابی و مقایسه می‌کنیم. برای حل این سوال به نوت‌بوک HW1-PART1.ipynb مراجعه کنید. (۱۵ نمره)

۲. در این نوت‌بوک به پیاده‌سازی شبکه‌های عصبی با استفاده از کتابخانه numpy می‌پردازیم. برای تکمیل نوت‌بوک تنها کافیست بخش‌های مشخص شده در نوت‌بوک را پیاده‌سازی کنید سپس با استفاده از توابعی که برای شما از قبل آماده شده اند آنها را تست کنید و در نهایت نیز پس از تکمیل بخش‌های مختلف شبکه عصبی خود را روی دو مسئله regression و classification تست کنید. (توجه کنید توابع مورد نیاز برای تست گرادیان بخش‌های مختلف شبکه و همچنین آموزش شبکه عصبی در پوشه utils در اختیار شما قرار گرفته اند و نیازی به پیاده‌سازی یا تغییر آنها ندارید.) (۱۵ نمره)

۳. در این نوت‌بوک با مفاهیم ابتدایی PyTorch آشنا می‌شویم. در ابتدا برای بخش‌های ابتدایی فایل -pytorch basic.py را کامل کنید و توابع آن را در نوت‌بوک تست کنید. پس از اتمام بخش ابتدایی نوت‌بوک و کامل کردن فایل پایتون مربوط به آن در بخش دوم در خود نوت‌بوک با استفاده از PyTorch یک MLP ساده ترین می‌کنید. (برای این سوال نیز تنها کافیست بخش‌های مشخص شده در نوت‌بوک و فایل پایتون را پیاده‌سازی کنید.) (۱۰ نمره)