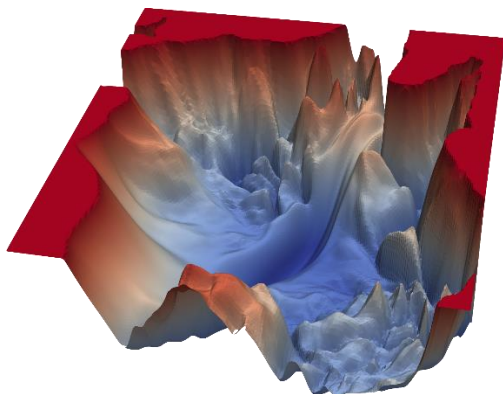


شبکه‌های عصبی و کنترل عصبی

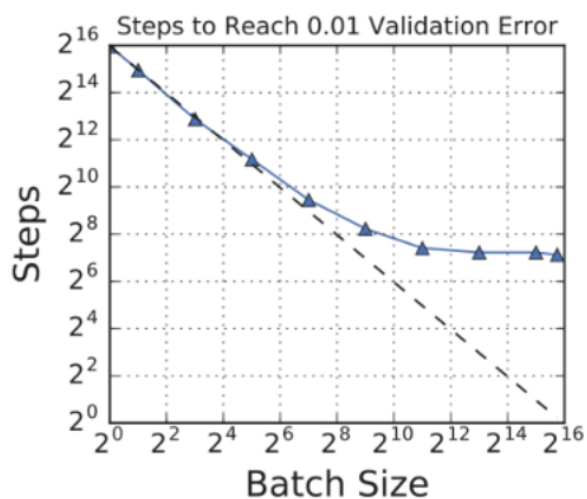
دکتر تشنه لب

تمرین سری اول، مهلت تحویل: ۳۰ آبان



سوال اول

به نمودار زیر توجه کنید که تعداد iteration های گرادیان کاهشی تصادفی SGD لازم برای رسیدن به یک مقدار مشخص از تابع زیان را برحسب Batch Size نشان میدهد.



شکل ۱) تعداد iteration های گرادیان کاهشی تصادفی SGD لازم برای رسیدن به یک مقدار مشخص از تابع زیان

الف: برای batch size های کوچک ، هرچه اندازه ی batch افزایش مییابد، تعداد تکرارهای لازم برای رسیدن به مقدار هدف تابع زیان کاهش مییابد. چرا ؟

ب) برای batch size های بزرگ ، با افزایش اندازه ی batch تعداد تکرارها تغییر چندانی نمیکند. چرا ؟

سوال دوم

مجموعه داده مسکن کالیفرنیا^۱، که داده‌های آن از سرشماری ۱۹۹۰ ایالات متحده استخراج شده است، به خانواده‌ای از مجموعه داده‌های محک^۲ برای رگرسیون اطلاق می‌شود. این مجموعه داده توانایی شبیه‌سازی یک مسئله پیش‌بینی چند متغیره در دنیای واقعی را داراست، که در آن متغیر هدف (Median House Value) توسط مجموعه‌ای از پارامترهای اقتصادی، جمعیتی و جغرافیایی کنترل می‌شود. این دیتاست در ابتدا برای مدلسازی و تحلیل الگوهای قیمت‌گذاری مکانی و اقتصادی در بازار مسکن و ارزیابی مدل‌های آماری پیش‌بینی‌کننده مورد استفاده قرار گرفت.

مستندات و طریقه بارگیری را می‌توانید از لینک زیر مطالعه کنید:

https://scikit-learn.org/stable/modules/generated/sklearn.datasets.fetch_california_housing.html

الف) توزیع (histogram) داده‌ها را برای هر کدام از ویژگی‌ها رسم کرده و آن را تفسیر کنید. بگویید که چه اطلاعاتی از این هیستوگرام دریافت می‌کنید.

ب) روش‌های زیر را برای نرمال‌سازی دیتاست فوق پیاده کنید و با ذکر نمونه (بررسی و مقایسه تاثیر روی دیتاها) عملکرد روش‌های زیر را برای نرمال سازی ویژگی‌های این دیتاست با هم مقایسه کنید. ابتدا فرمول هر کدام از این روش‌ها را بفهمید، تفسیر کنید و سپس به پیاده سازی و بررسی عملکرد بپردازید.

- نرمال سازی بین صفر و یک
- نرمال سازی بین -1 و $+1$
- نرمال سازی به توزیع نرمال استاندارد (standard scaling)
- نرمال سازی مقاوم (Robust scaling)

پ) کدام روش برای داده‌های پرت (outlier) روش بهتری است؟ چرا؟ (از نظر ریاضی بگویید که چرا و به ازای نمونه‌های مورد نظرتان از دیتاست، تفسیر خود را تصدیق کنید)

سوال سوم

الف) هر کدام از توابع فعال‌ساز زیر را به همراه مشتقشان در یک figure رسم کنید (در هر figure هم خود تابع نشان داده شود و هم مشتق آن)

| | | |
|---------|------|----------|
| ReLU | Mish | ELU |
| Sigmoid | Tanh | Softplus |

¹ California Housing Prices

² Benchmark Datasets

ب) یکی از مشکلات تابع فعالسازی sigmoid اشباع شدن است. به طور خلاصه این مشکل را توضیح دهید و بگویید آیا با جایگزین کردن تابع tanh این مشکل برطرف میشود یا خیر؟ دلیل خود را توضیح دهید.

ج) توضیح دهید که استفاده از تابع فعالسازی tanh به جای سیگموئید چگونه باعث میشود بهینه سازی آسانتر شود؟

سوال چهارم

منظور از Gradient Vanishing و Gradient Exploding در آموزش شبکه های عصبی چیست؟ یک مقاله علمی از منابعی مانند Scholar Google بیاورید که روشی برای مقابله با این مشکلات پیشنهاد کرده باشد. ایده روش ارائه شده در آن مقاله را توضیح دهید و مزایا و محدودیت های آن را ذکر کنید.

سوال پنجم

در این بخش شما باید با تابع داده شده F ، 100 نمونه دیتا تولید کنید و آنها را رسم کنید. سپس این نمونه دیتاها را در یک فایل اکسل ذخیره کنید. چون قرار است الگوریتمهای بهینه سازی معرفی شده در جدول زیر را روی نمونه هایی که تولید کردید، تست کنید.

$$F = 3\text{logsig}(1.7x) + 4\text{tansig}(3x) + 4\text{Swish}(2.5x) + x^4 - 2x^2 + 0.1x + 1$$

اکنون تابع زیر را در نظر بگیرید. مقدار اولیه تمامی پارامترها را به صورت رندوم بین 0 و 1 انتخاب کنید. سپس با در نظر گرفتن مقدار خطای میانگین مربعات MSE و داده های تولید شده در مرحله قبل، با استفاده از الگوریتمهای بهینه سازی زیر، پارامترهای مربوط به 3 حالت زیر را بهینه سازی کنید و سپس جدول زیر را به طور کامل پر کنید.

- بهینه سازی مقادیر آلفا
- بهینه سازی مقادیر تتا
- بهینه سازی همزمان مقادیر آلفا و تتا

$$F = \alpha_1 \text{logsig}(\theta_1 x) + \alpha_2 \text{tansig}(\theta_2 x) + \alpha_3 \text{Swish}(\theta_3 x) + (\alpha_4 x^2 - 1) + \alpha_5 x$$

| MSE حاصل از آموزش همزمان پارامترهای آلفا و تتا | MSE حاصل از آموزش پارامترهای تتا | MSE حاصل از آموزش پارامترهای آلفا | الگوریتم بهینه سازی |
|--|----------------------------------|-----------------------------------|---------------------|
| | | | SGD |
| | | | Momentum |
| | | | AdaGrad |

| | | | |
|---------------------|--|--|--|
| RMSProp | | | |
| AdaDelta | | | |
| Nestrov | | | |
| Adam | | | |
| AdaMax | | | |
| Na ³ dam | | | |
| AMSGrad | | | |

پس از بررسی الگوریتمها، آنها را مقایسه کنید و خطای میانگین مربعات هر یک را گزارش کنید. عملکرد کدام روش بهتر بوده است؟ به نکات زیر توجه کنید.

بررسی دقیق هر الگوریتم لازم است. (توضیح کامل همراه با بیان دقیق روابط و پارامترها)

پیاده‌سازی جداگانه هر روش بهینه‌سازی در پایتون یا متلب (ترجیحا پایتون)

عدم استفاده از کتابخانه‌های آماده جهت پیاده‌سازی

پلات نمودار داده‌های پیش‌بینی شده و همچنین نمودار Loss.

گزارش تمامی پارامترهای اولیه مثل تعداد اپاک، نرخ یادگیری و...

صرف پیاده‌سازی روشها ملاک نمره نیست، کیفیت کار و تسلط به موضوع و نتایج پیاده‌سازی مهم است

برای درک بهتر این روشها میتوانید به لینک زیر مراجعه کنید:

<https://www.kdnuggets.com/2019/06/gradient-descent-algorithms-cheat-sheet.html>

نکات و تذکرات تحویل تمرین

- استفاده از زبان برنامه نویسی پایتون و متلب مجاز است.
- تعیین مقادیر ذکر نشده مانند نرخ یادگیری، تعداد اپاک و ... به عهده خودتان است تا نتیجه مطلوب حاصل شود.

³ این بهینه ساز را خودتان مطالعه کنید. در pdf آورده شده در ادامه سوال قابل مطالعه است.

- می‌توانید تمرین را به صورت انفرادی یا گروه‌های حداکثر 2 نفره انجام دهید. تمامی اعضا باید در تمامی بخش‌ها مشارکت و تسلط داشته باشند. دانشجویان دکتری باید به صورت انفرادی کار کنند.
- کدها و گزارش را به صورت یک فایل فشرده در سایت آپلود کنید. حتما هر دو نفر اعضای گروه باید هر تمرین را بارگزاری کنند.
- در صورت مشاهده گزارش‌ها یا کدهای مشابه، نمره کسر می‌گردد
- توجه شود که گزارش‌ها باید متفاوت باشد و در صورت وجود تشابه نمره آن بخش کسر می‌گردد