

به نام خدا
تمرین سری اول درس یادگیری عمیق
دانشگاه شهید بهشتی
مهلت تحویل: ۱۳۹۸/۰۱/۱۷

در حل تمارین به نکات زیر توجه فرمایید:

- تمام تمارین به صورت حضوری ارائه خواهد شد. فرد ارائه دهنده به صورت تصادفی از میان اعضای گروه انتخاب شده و بقیه اعضای گروه حق شرکت در جلسه ارائه را نخواهند داشت. بنابراین همه اعضای تیم باید به نحوه حل تمام بخش‌های تمارین مسلط باشند.
- گروه‌ها باید به صورت مستقل به حل تمارین بپردازند. اگر مسجل شود پاسخ حتی یک تمرین میان دو گروه مشابه است به عنوان تقلب محسوب شده و نمره کل تمرین برای تمام اعضای دو گروه صفر منظور خواهد شد.
- حل و نتایج تمام تمرین‌های تئوری و عملی باید در قالب یک گزارش و تا ساعت ۲۳:۵۹ روز ۱۷ فروردین ۱۳۹۸ ارسال شود. به ازای هر روز تأخیر در ارسال موارد ۱۰ درصد از نمره کسر شده و آن تیم از شرکت در بخش رقابتی محروم می‌شود.
- تمرین ۲ نمره داشته و به بهترین تیم در بخش رقابتی ۰.۵ نمره اضافه خواهد شد.

۱. توابع \tanh ، sigmoid و ReLU از پرکاربردترین توابع فعال‌ساز برای شبکه‌های عصبی هستند. سوال‌های زیر را در مورد این توابع پاسخ دهید:
(الف) فرمول‌های مربوط به این توابع و مشتق آن‌ها را به صورت ریاضیاتی نوشته و آن‌ها را رسم کنید.
(ب) با توجه به شکل‌های این توابع، نقاط ضعف و قوت آن‌ها را تحلیل کرده و با هم مقایسه کنید.
(پ) به نظر شما چه ویژگی مهمی برای تابع sigmoid با توجه به فرمول مشتق آن وجود دارد؟
(ت) نقاط ضعف تابع ReLU را مشخص کنید. چه توابعی برای حل این مشکلات پیشنهاد شده است؟ حداقل دو تابع دیگر را معرفی کرده و فرمول آن‌ها و مشتق‌هایشان و شکل‌های آن‌ها را رسم کنید.
۲. فرض کنید می‌خواهیم با استفاده از مدل $\text{logistic regression}$ یک دسته‌بند دو کلاسه طراحی کنیم.
این دسته‌بند قرار است در یک فضای با ۳ ویژگی عمل کند.
(الف) گراف محاسباتی مربوط به این مدل را رسم کنید.
(ب) پارامترهای این مدل را مشخص کنید.

پ) فرض کنید تمام پارامترها با مقدار ۰.۱ مقداردهی اولیه می‌شوند. فرض کنید داده‌های زیر به ترتیب مشاهده می‌شود. نحوه به روزرسانی پارامترها بعد از مشاهده هر کدام از این داده‌ها با توجه به الگوریتم backpropagation را مشخص کنید. مراحل اجرا و مقادیر نهایی پارامترها را دقیق نشان دهید.

X_1	X_2	X_3	Class
2	1	4	Class1
3	2	1	Class2

۳. نکته مهم: برای انجام این تمرین تنها مجاز به استفاده از پایتون و کتابخانه Numpy هستید. اگر از زبان دیگری می‌خواهید استفاده کنید از کتابخانه‌هایی معادل Numpy در آن زبان که به تأیید استاد درس رسیده باشد می‌توانید استفاده کنید.

در این تمرین می‌خواهیم صفر تا صد یک شبکه عصبی عمیق را پیاده‌سازی کنیم. هدف این شبکه عصبی دسته‌بندی داده‌ها در دادگان cifar10 است.

بخش اول: آماده‌سازی اجزای مورد نیاز برای گراف محاسباتی

الف) گرهی Layer را طراحی کنید که ورودی آن ماتریس X_{in} و خروجی آن ماتریس X_{out} می‌باشد. این گره یک ماتریس W به عنوان پارامتر دارد. داریم:

$$X_{in}: N * D_{in}$$

$$X_{out}: N * D_{out}$$

$$W: D_{in} * D_{out}$$

$$X_{out} = f(X_{in} * W)$$

تابع f می‌تواند از یک لیست انتخاب شود (سعی کنید توابع فعال‌ساز معروف را در لیست داشته باشید). فاز رو به جلو و فاز رو به عقب را برای این گره بنویسید. در این گره N تعداد داده‌های داخل batch خواهد بود.

ب) گره softmax را طراحی کنید که ورودی آن ماتریس X_{in} و خروجی آن ماتریس X_{out} می‌باشد. داریم:

$$X_{in}: N * D$$

$$X_{out}: N * 1$$

$$X_{out} = \text{softmax}(X_{in})$$

فاز رو به جلو و فاز رو به عقب را برای این گره بنویسید.

پ) گره Loss را طراحی کنید که ورودی آن ماتریس YP و بردار Y و خروجی آن یک اسکالر L است. داریم:

$$Y_P: N \times 1$$

$$Y: N \times 1$$

$$L = \text{CrossEntropy}(Y_P, Y)$$

فاز رو به جلو و فاز رو به عقب را برای این گره بنویسید.

بخش دوم: مدل سازی دسته بند با استفاده از گراف محاسباتی

گراف محاسباتی یک شبکه عصبی چند لایه را با استفاده از گره های تعریف شده در بخش قبل کشیده و با استفاده از آن و توابع رو به جلو و رو به عقب نوشته شده برای آن گره ها به سؤالات زیر پاسخ دهید.

الف) دادگان را دقیق مطالعه کرده و نکات مهم آن را بنویسید. تعداد داده ها؟ تعداد کلاس ها؟ ابعاد داده های ورودی؟ ...

ب) تابعی بنویسید که عکس های این دادگان را به یک بردار ویژگی تبدیل کند. بردار ویژگی برای هر عکس میانگین وزن رنگ های مختلف هر پیکسل است. بنابراین طول بردار ویژگی برابر ۱۰۲۴ خواهد بود.

پ) تابعی بنویسید که با توجه به مقادیر مشخصی از پارامترها مقدار خطای سیستم را محاسبه کند.
ت) تابعی بنویسید که با استفاده از الگوریتم backpropagation مقدار گرادیان را برای تمام پارامترهای مدل محاسبه کند.

ث) تابعی بنویسید که مقدار گرادیان را برای یک نقطه و به صورت عددی محاسبه کند. سپس میزان تفاوت گرادیان های مشاهده شده در این بخش و بخش قبل با هم مقایسه کرده و نتایج را تحلیل کنید.

ج) تابعی بنویسید که با استفاده از الگوریتم mini-batch gradient descent و با استفاده از گرادیان های محاسبه شده در بخش های قبل پارامترهای مدل را آپدیت کند.

چ) تابعی بنویسید که با استفاده از دو الگوریتم دلخواه دیگر (غیر از انواع گرادیان کاهشی) پارامترهای مدل را آپدیت کند.

ح) تمام توابع بالا را در یک روند آموزشی برای آموزش پارامترهای مدل به کار برده و مدل را آموزش دهید.

خ) کارایی روش های پیاده سازی برای آپدیت کردن پارامترها را با هم مقایسه کرده و نتایج را تحلیل کنید.

د) مقدار loss مدل را در انتهای هر بار آپدیت کردن پارامترها به صورت یک نمودار نمایش دهید. این نمودار را تحلیل کنید.

ذ) کارایی مدل را روی پیکره تست امتحان کرده و نتایج را گزارش دهید.

ر) بخش رقابتی: تا حد امکان مدل را بهینه کنید تا دقت آن روی پیکره تست تا حد امکان بیشینه شود. از هر تکنیکی در این بخش می توانید استفاده کنید. حتی می توانید مدل را تغییر دهید. فقط کماکان نمی توانید از هیچ کتابخانه دیگری غیر از Numpy استفاده کنید.