

تکلیف اول درس شبکه های عصبی

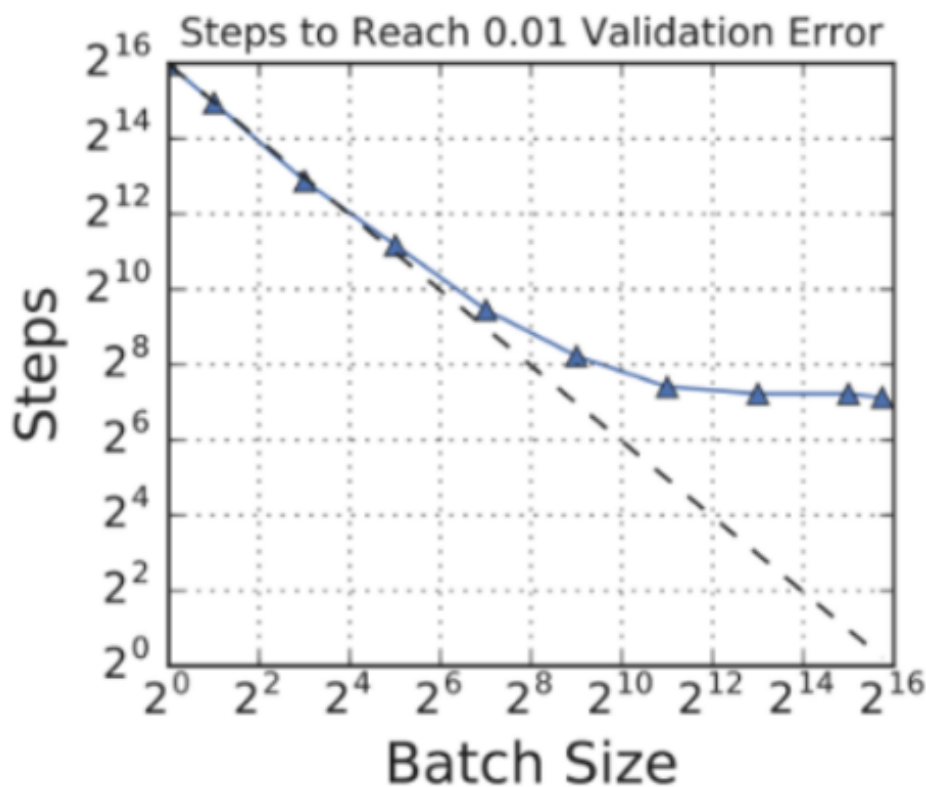
وحید ملکی

شماره دانشجویی: ۴۰۳۱۳۰۰۴

۲۵ آبان ۱۴۰۴

سوال ۱

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



- شکل ۱: تعداد iteration های گرادینان کاهشی تصادفی SGD لازم برای رسیدن به یک مقدار مشخص از تابع زیان
- الف: برای batch های size کوچک، هرچه اندازهی batch افزایش می یابد، تعداد تکرارهای لازم برای رسیدن به مقدار هدف تابع زیان کاهش می یابد. چرا؟
- ب: برای batch های size بزرگ، با افزایش اندازهی batch تعداد تکرارها تغییر چندانی نمی کند. چرا؟

پاسخ

با توجه به نمودار ارائه شده در شکل ۱ که ارتباط بین اندازه بچ (Batch Size) و تعداد تکرارهای (Iterations) لازم برای الگوریتم SGD جهت رسیدن به خطای مشخصی را نشان می‌دهد، پاسخ به سوالات به شرح زیر است:

الف: کاهش تعداد تکرارها با افزایش اندازه بچ (در اندازه‌های کوچک)

در اندازه‌های بچ کوچک، با افزایش اندازه بچ، تعداد تکرارهای لازم برای رسیدن به یک مقدار خطای مشخص کاهش می‌یابد. دلیل اصلی این پدیده، افزایش دقت تخمین گرادیان است:

- نویز در گرادیان: در الگوریتم SGD، گرادیان در هر مرحله تنها بر اساس یک زیرمجموعه کوچک (بچ) از داده‌ها محاسبه می‌شود. این امر باعث ایجاد نویز در تخمین گرادیان نسبت به گرادیان واقعی (محاسبه شده بر کل داده‌ها) می‌گردد.

- اثر اندازه بچ کوچک: وقتی اندازه بچ بسیار کوچک باشد (مثلاً ۱ یا ۲)، نویز بسیار بالاست. این نویز باعث نوسانات شدید در جهت به‌روزرسانی پارامترها و حرکت پرپیچ‌وخم به سمت نقطه بهینه می‌شود. در نتیجه، الگوریتم مسیر طولانی‌تری را طی کرده و به تکرارهای بیشتری نیاز دارد.

- اثر افزایش اندازه بچ: با افزایش اندازه بچ، تعداد نمونه‌های دخیل در محاسبه گرادیان افزایش می‌یابد. این امر میانگین‌گیری بهتری انجام داده و تخمین گرادیان را به گرادیان واقعی نزدیک‌تر می‌کند. گرادیان دقیق‌تر منجر به حرکت مستقیم‌تر و پایدارتر به سمت مینیمم تابع زیان شده و در نتیجه تعداد تکرارهای لازم کاهش می‌یابد.

ب: عدم تغییر محسوس تعداد تکرارها با افزایش اندازه بچ (در اندازه‌های بزرگ)

در اندازه‌های بچ بزرگ، با افزایش بیشتر اندازه بچ، تعداد تکرارها تقریباً ثابت می‌ماند. این پدیده که به نقطه اشباع (Saturation Point) معروف است، به دلایل زیر رخ می‌دهد:

- بازده نزولی (Diminishing Returns): پس از رسیدن به یک اندازه بچ کافی، گرادیان تخمینی به اندازه‌ای دقیق می‌شود که افزودن نمونه‌های بیشتر، بهبود قابل‌توجهی در دقت آن ایجاد نمی‌کند. اطلاعات اضافی حاصل از نمونه‌های بیشتر، تأثیر ناچیزی بر جهت کلی گرادیان دارد.

- اشباع کیفیت گرادیان: وقتی اندازه بچ به اندازه کافی بزرگ باشد، گرادیان محاسبه شده نماینده بسیار خوبی از گرادیان واقعی است. در این حالت، الگوریتم در هر تکرار نزدیک به بهترین جهت ممکن را طی می‌کند و افزایش بیشتر اندازه بچ نمی‌تواند این مسیر را بهینه‌تر کند. بنابراین، تعداد گام‌های لازم برای همگرایی ثابت می‌ماند.

- افزایش هزینه محاسباتی: هرچند تعداد تکرارها تغییر نمی‌کند، اما افزایش اندازه بچ باعث افزایش حجم محاسبات در هر تکرار می‌شود. این امر زمان کل آموزش را افزایش می‌دهد، بدون بهبود سرعت همگرایی (از نظر تعداد تکرار). از این رو، انتخاب اندازه بچ مناسب یک مصالحه بین دقت گرادیان و کارایی محاسباتی است.

در نهایت، نمودار نشان می‌دهد که یک اندازه بچ بهینه وجود دارد که تعادل مناسبی بین نویز گرادیان و هزینه محاسباتی برقرار می‌کند.