

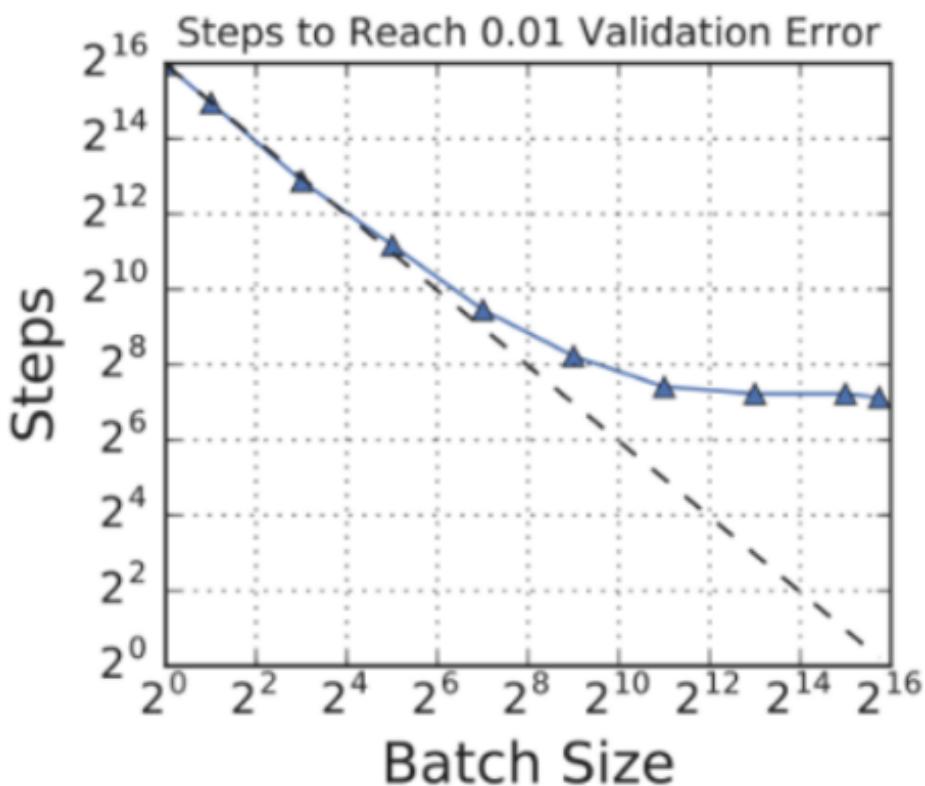
تکلیف اول درس شبکه های عصبی

وحید ملکی
شماره دانشجویی: ۴۰۳۱۳۰۰۴

۱۴۰۴ آبان ۲۵

سوال ۱

به نودار زیر توجه کنید که تعداد iteration های گرادیان 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 گرادیان کاهشی تصادفی SGD لازم برای رسیدن به یک مقدار مشخص از تابع زیان

الف: برای size های batch کوچک، هرچه اندازهی batch افزایش می یابد، تعداد تکرارهای لازم برای رسیدن به مقدار هدف تابع زیان کاهش می یابد. چرا؟

ب: برای size های batch بزرگ، با افزایش اندازهی batch تعداد تکرارها تغییر چندانی نمی کند. چرا؟

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 می دهد، پاسخ به سوالات به شرح زیر است:

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

در اندازه های بچ کوچک، با افزایش اندازه بچ، تعداد تکرارهای لازم برای رسیدن به یک مقدار خطای مشخص کاهش می یابد. دلیل اصلی این پدیده، افزایش دقت تخمین گردنیان است:

- نویز در گردانی: در الگوریتم SGD گردانی در هر مرحله تنها بر اساس یک زیرمجموعه کوچک (بچ) از داده ها محاسبه می شود. این امر باعث ایجاد نویز در تخمین گردانیان نسبت به گردانیان واقعی (محاسبه شده بر کل داده ها) می گردد.

- اثر اندازه بچ کوچک: وقتی اندازه بچ بسیار کوچک باشد (مثالاً ۱ یا ۲)، نویز بسیار بالاست. این نویز باعث نوسانات شدید در جهت به روزرسانی پارامترها و حرکت پر پیچ و خم به سمت نقطه بهینه می شود. در نتیجه، الگوریتم مسیر طولانی تری را طی کرده و به تکرارهای بیشتری نیاز دارد.

- اثر افزایش اندازه بچ: با افزایش اندازه بچ، تعداد نمونه های دخیل در محاسبه گردانیان افزایش می یابد. این امر میانگین گیری بهتری انجام داده و تخمین گردانیان را به گردانیان واقعی نزدیک تر می کند. گردانیان دقیق تر منجر به حرکت مستقیم تر و پایدار تر به سمت مینیمم تابع زیان شده و در نتیجه تعداد تکرارهای لازم کاهش می یابد.

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

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

- بازده تزویی (Diminishing Returns): پس از رسیدن به یک اندازه بچ کافی، گردانیان تخمینی به اندازه های دقیق می شود که افزودن نمونه های بیشتر، بهبود قابل توجهی در دقت آن ایجاد نمی کند. اطلاعات اضافی حاصل از نمونه های بیشتر، تأثیر ناچیزی بر جهت کلی گردانیان دارد.

- اشباع کیفیت گردانیان: وقتی اندازه بچ به اندازه کافی بزرگ باشد، گردانیان محاسبه شده نماینده بسیار خوبی از گردانیان واقعی است. در این حالت، الگوریتم در هر تکرار نزدیک به بهترین جهت ممکن را طی می کند و افزایش بیشتر اندازه بچ نمی تواند این مسیر را بهینه تر کند. بنابراین، تعداد گام های لازم برای همگرای ثابت می ماند.

- افزایش هزینه محاسباتی: هرچند تعداد تکرارها تغییر نمی کند، اما افزایش اندازه بچ باعث افزایش حجم محاسبات در هر تکرار می شود، این امر زمان کل آموزش را افزایش می دهد، بدون بهبود سرعت همگرای (از نظر تعداد تکرار). از این رو، انتخاب اندازه بچ مناسب یک مصالحه بین دقت گردانیان و کارایی محاسباتی است.

در نهایت، نمودار نشان می دهد که یک اندازه بچ بهینه وجود دارد که تعادل مناسبی بین نویز گردانیان و هزینه محاسباتی برقرار می کند.