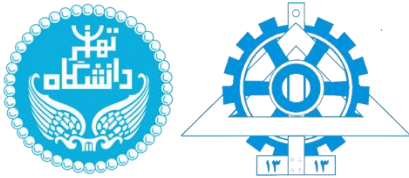


به نام خدا



سامانه‌های یادگیری ماشین توزیع شده (پاییز ۱۴۰۳)

تمرین نوشتاری ۱

موعد تحویل: ۱۴۰۳/۹/۱

۱. عملیات مورد استفاده در محاسبات یادگیری ماشین عموماً مبتنی بر محاسبات جبر خطی هستند. دیدیم که سخت افزارهای خاص منظوره‌ای برای این محاسبات وجود دارند. با این حال، این سخت افزارها پیش از رواج یادگیری ماشین وجود داشتند. ۴ نمونه از سایر استفاده‌های این سخت افزارها را نام برده و توضیح دهید.
۲. آیا افزونه‌های برداری که در پردازنده‌ها وجود دارند و برای تسک‌های SIMD استفاده می‌شوند واحد محاسبه و منطق جدا از pipeline مربوط به تسک‌های SISD دارند؟ با ذکر مثال برای نوع خاصی از پردازنده توضیح دهید.
۳. کتابخانه‌های MPI, GLOO و NCCL را مقایسه کرده و مشخص کنید هر کدام در چه سناریویی کاربرد دارند.
۴. NVLink و NVSwitch را مقایسه کنید. مزیت استفاده از آنها نسبت به PCIe چیست؟ آیا هنگام استفاده از NVLink همچنان PCIe برای ارتباط استفاده می‌شود؟
۵. آیا تمام الگوریتم‌های یادگیری ماشین را می‌توان به صورت data parallel پیاده سازی کرد؟ اگر محدودیتی وجود دارد، آن را بیان کنید.
۶. دقت مدلی که با روش data parallel آموزش داده شده نسبت به مدلی که به صورت عادی آموزش دیده است چه قدر کاهش می‌یابد؟ توضیح دهید.
۷. تصور کنید K پردازنده در اختیار دارید و می‌خواهید مدلی با N پارامتر را بر روی M داده به روش data parallel آموزش دهید. قدرت پردازشی پردازنده نام 2^I است. برای آن که آموزش سریع‌تر شود به هر پردازنده چه مقدار اختصاص می‌دهید؟
۸. الف) ماتریس‌های تنک چه ویژگی یا ویژگی‌هایی دارند؟
ب) ماتریس تنک زیر را در نظر بگیرید. آن را با فرمت اندیس‌گذاری زبان FORTRAN نشان دهید. با همین فرمت مقدار $2A+5$ را محاسبه کنید و بنویسید.

$$A = \begin{bmatrix} 0 & 5 & 0 & 0 & 0 \\ 12 & 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 71 & 0 \\ 0 & 0 & 4 & 0 & 1 \end{bmatrix}$$

۹. دیتاستی شامل ۳ میلیارد داده آموزشی برای آموزش یک شبکه عصبی داریم، یک GPU برای آموزش داریم و حافظه‌ی آن ۸ گیگابایت است. اگر داده‌ها با فرمت fp32 استفاده شوند، برای آموزش این شبکه‌ی عصبی از چه روش یا روش‌هایی استفاده می‌شود؟ مزایا و معایب آن‌ها را نسبت به سایر حالت بنویسید.
۱۰. عملیات زیر و ساین ماتریس و بردارها را در نظر بگیرید. برای انجام عملیات زیر کدام یک از سطوح BLAS را پیشنهاد می‌کنید؟ چرا؟

$$\alpha A + \beta$$

$$\alpha = 1 \times m \text{ vector}$$

$$A = m \times n \text{ matrix}$$

$$\beta = 1 \times n \text{ vector}$$

$$m = 10^{10}$$

$$n = 10^6$$

۱۱. فرض کنید که یک شبکه‌ی عصبی برای آموزش دارید که اندازه‌ی آن ۳۶ گیگابایت است. تعدادی GPU داریم که حافظه‌ی آن‌ها هشت گیگابایت است. همچنین حداقل حافظه‌ی مصرفی برای داده‌ها حین آموزش ۲ گیگابایت است. الف) توضیح دهید چرا استفاده از روش موازی‌سازی داده در این سناریو امکان‌پذیر نیست؟
 ب) حداقل تعداد GPUهای مورد نیاز برای آموزش این مدل با استفاده از روش Model Parallelism را محاسبه کنید.
 ج) یک عیب روش data parallelism را از نظر ارتباطات بین GPUها توضیح دهید.

۱۲. تفاوت میان fp32 و fp16 را توضیح دهید و مزایا و معایب استفاده از هر یک را حین و یا پس از آموزش توضیح دهید. آیا تمام سخت‌افزارهای مرسوم برای شتابدهی محاسبات جبر خطی از هر دو پشتیبانی می‌کنند؟