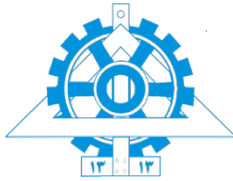


به نام خدا

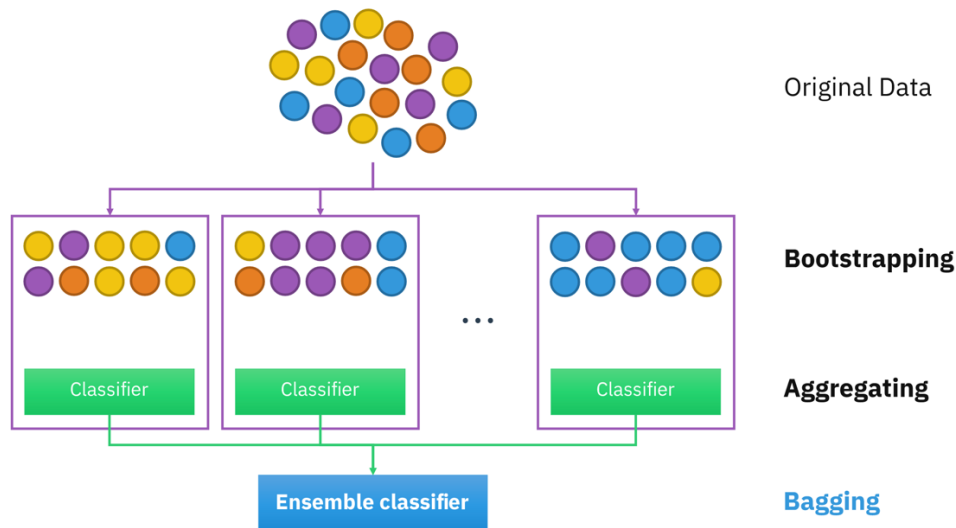


سامانه‌های یادگیری ماشین توزیع شده

تمرین کتبی دوم به همراه پاسخ
موعد تحویل: نیازی به تحویل نیست.

۱. (۱۵ نمره) روش **bagging** یکی از روش‌های یادگیری گروهی (**ensemble learning**) است. توضیح دهید چرا به این روش **bootstrap aggregating** نیز گفته می‌شود. به عبارتی مراحل **bootstrap** و تجمیع آن را توضیح دهید.

پاسخ: نام **bagging** از **bootstrap aggregating** گرفته شده است. در این روش، داده‌ها به طور یکنواخت با جاگذاری نمونه‌برداری شده و به m گروه تقسیم می‌شوند. این نوع نمونه‌برداری **bootstrap** نام دارد که در داده‌های هر بخش مستقل یکدیگر هستند. سپس m مدل روی این داده‌ها آموزش می‌بینند و در نهایت نتیجه این مدل‌ها تجمیع می‌گردند.



۲. (۱۰ نمره) چرا توپولوژی حلقه (ring) برای GPUهای بکار می‌رود؟

پاسخ: به طور سنتی این روش به دلیل سادگی و ارزانی و نیز نحوه‌ی اتصال سخت‌افزاری GPUها پرترفدار بوده است. به علاوه می‌توان نشان داد که عملیات پرهزینه‌ی کاهشی (نظیر **AllReduce**) را می‌توان به کمک توپولوژی حلقه یا مجموعه‌ای از حلقه‌ها پیاده‌سازی کرد. برای اطلاعات بیشتر می‌توانید به این مقاله رجوع کنید:

Pitch Patarasuk and Xin Yuan. "Bandwidth optimal all-reduce algorithms for clusters of workstations." *Journal of Parallel and Distributed Computing* 69.2 (2009): 117-124.

۳. (۲۵ نمره) با یک مثال عددی توضیح دهید روش **sufficient factor broadcasting (SFB)** چگونه کار می‌کند.

پاسخ: ماتریس وزن‌های M قابل تبدیل به ضرب خارجی دو بردار a و b است. در نتیجه، به جای ارسال آن می‌توان دو بردار a و b را ارسال کرد. در این حالت، به جای ارسال ۱۶ پارامتر، فقط ۸ پارامتر ارسال شده است. در عمل، این روش، پیچیدگی تبادل را از توان دو به خطی تبدیل می‌کند.

$$M = \begin{bmatrix} 20 & 5 & 45 & 10 \\ 8 & 2 & 18 & 4 \\ 4 & 1 & 9 & 2 \\ 12 & 3 & 27 & 6 \end{bmatrix} = a \otimes b \Rightarrow a = \begin{bmatrix} 5 \\ 2 \\ 1 \\ 3 \end{bmatrix}, b = \begin{bmatrix} 4 \\ 1 \\ 9 \\ 2 \end{bmatrix}$$

۴. (۱۰ نمره) پردازنده‌های Intel x۸۶ و ARM کدام یک از روش‌های big-endian یا little-endian را استفاده می‌کنند؟ پروتکل شبکه TCP/IP چگونه؟

پاسخ: پردازنده‌های Intel x۸۶ از روش little-endian استفاده می‌کنند. پردازنده‌های ARM از دو روش big-endian و little-endian پشتیبانی می‌کنند. البته نسخه‌های قدیمی این پردازنده فقط از little-endian پشتیبانی می‌کردند. در پروتکل‌های شبکه، از روش big endian استفاده می‌شود.

۵. (۲۰ نمره) سرعت همگرایی آموزش یک مدل توزیع شده با روش Stale Synchronous Parallel (SSP) را نسبت به Bulk Synchronous Parallel (BSP) مقایسه کنید. توضیح دهید که چرا این سرعت متفاوت است.

پاسخ: در روش SSP بسته به میزان stale بودن ماشین‌ها (که وابسته به میزان مجاز تعریف شده، تفاوت سرعت ماشین‌ها، میزان خرابی و ...) دارد، می‌تواند سریع‌تر از BSP عمل کند چون ماشین‌ها لازم نیست منتظر همدیگر بمانند. از طرفی، در صورت افزایش میزان stale بودن از حدی، باعث کندی آن نسبت به BSP می‌گردد.

۶. (۲۰ نمره) فرض کنید توپولوژی حلقه را با n ماشین پیاده کرده‌اید. ماشین i ام به احتمال p_i ممکن است خراب شود. اگر بتوانید یک ماشین دیگر به عنوان پشتیبان تهیه کنید تا در صورت خرابی یکی از ماشین‌ها با آن جایگزین شود، احتمال خرابی این سیستم چقدر کاهش می‌یابد؟

پاسخ: با توجه به اینکه مسئله احتمال خرابی ماشین پشتیبان را نداده است، احتمال آن را صفر در نظر می‌گیریم. احتمال خرابی بدون پشتیبان:

$$1 - \prod_{i=1}^n (1 - p_i)$$

احتمال خرابی با پشتیبان:

$$1 - \left(\prod_{i=1}^n (1 - p_i) + \sum_{i=1}^n p_i \prod_{j=1, j \neq i}^n (1 - p_j) \right)$$

در نتیجه احتمال خرابی به میزان زیر کاهش می‌یابد:

$$\begin{aligned}
& \sum_{i=1}^n p_i \prod_{j=1, i \neq j}^n (1 - p_j) \\
&= \sum_{i=1}^n p_i \frac{(1 - p_i)}{(1 - p_i)} \prod_{j=1, i \neq j}^n (1 - p_j) = \sum_{i=1}^n \frac{p_i}{(1 - p_i)} \prod_{j=1}^n (1 - p_j) \\
&= \prod_{j=1}^n (1 - p_j) \sum_{i=1}^n \frac{p_i}{(1 - p_i)}
\end{aligned}$$

نحوه تحویل تمرین

پاسخ سوالات را به صورت یک فایل PDF با شماره دانشجویی خود در سایت elearn آپلود کنید. لطفاً از فونت سایز ۱۲ (یا بیشتر) برای پاسخ به سوالات استفاده نمایید.