



سامانه‌های یادگیری ماشین توزیع شده (پاییز ۱۴۰۲)

تمرین کامپیوتری ۲

موعده تحویل: ۱۴۰۲/۹/۱۴

لطفا پیش از شروع کار بر روی تمرین، به نکات زیر توجه فرمایید.

- حتما ویدئوی هندزآن PyTorch DDP را قبل از انجام تمرین مشاهده نمایید.
 - برای راحتی در توسعه و تست کد، از ماشین مجازی لینوکس خود استفاده نمایید تا ترافیک کلاستر (به خصوص در ساعات آخر مهلت تمرین) افزایش نیابد. پس از اطمینان از عملکرد کد، می‌توانید آن را روی کلاستر اجرا کنید.
 - سوالات خود را در گروه تلگرام درس مطرح نمایید. به هیچ وجه کد یا پاسخ سوالات را در گروه به اشتراک نگذارید.
 - میزان تاخیر، تا دو روز مجاز است. تاخیر به صورت ساعتی محاسبه شده و هر روز ۱۰ درصد نمره کم میشود. تحویل تمرین پس از دو روز تاخیر امکانپذیر نخواهد بود.
 - هدف از انجام این تمرین، آشنایی بیشتر با PyTorch DDP و بررسی تعدادی پارامترهای مختلف بر تغییر سرعت و حافظه‌ی مصرفی در آموزش مدل‌های شبکه‌ی عصبی است. در این تمرین، قصد طبقه‌بندی مجموعه داده‌ی Fashion MNIST را داریم. این مجموعه داده شامل ۶۰,۰۰۰ داده‌ی آموزش و ۱۰,۰۰۰ داده‌ی تست است. این مجموعه داده در کتابخانه‌ی torchvision قابل دسترس است. برای طبقه‌بندی تصاویر این مجموعه داده از شبکه‌های کانولوشنی استفاده نمایید.
۱. (۳۵ نمره) در این سوال، قصد بررسی تغییر سرعت آموزش در شرایط استفاده از چند GPU را داریم. ابتدا شبکه‌ی کانولوشنی‌ای طراحی نمایید که بتواند به دقت مناسبی (بالای ۸۰ درصد) برای طبقه‌بندی داده‌های تست برسد. در این سوال از بهینه‌ساز Adam و سایز batch ۳۲ استفاده نمایید.

الف (۵ نمره) این شبکه را به بر روی یک GPU آموزش دهید.

ب (۱۰ نمره) کد بخش الف را به گونه‌ای تغییر دهید که بتوان به کمک دستور python و با استفاده از PyTorch Multiprocessing آن را بر روی ۲ GPU اجرا کرد.

ج (۱۰ نمره) کد بخش ب را به گونه‌ای تغییر دهید که بتوان به کمک torchrun آن را بر روی ۲ GPU اجرا کرد.

(۱۰ نمره) در هر آزمایش دقت نهایی مدل بر روی داده‌های تست، زمان اجرای آموزش و میزان حافظه‌ی مصرفی GPU را به دست آورده، با یکدیگر مقایسه کرده و تحلیل نمایید.

۲. (۱۵ نمره) در این سوال قصد بررسی تغییر سایز batch در سرعت آموزش را داریم. کد بخش (۱-ج) را با سایز batch ۱۶، ۶۴ و ۱۲۸ اجرا نمایید. در هر آزمایش، زمان اجرای آموزش، حافظه‌ی مصرفی GPU و دقت نهایی مدل را به دست آورید. نمودار تغییر زمان آموزش، دقت مدل و حافظه‌ی مصرفی GPU را بر حسب تغییر سایز batch رسم نمایید و نتایج را تحلیل کنید.

۳. (۲۵ نمره) یکی از تکنیک‌های مورد استفاده در آموزش مدل‌های شبکه عصبی، تکنیک Gradient Accumulation است.

الف) (۱۰ نمره) این تکنیک را به طور خلاصه توضیح داده و کد بخش (۱-ج) را به گونه‌ای تغییر دهید تا از این تکنیک پشتیبانی نماید و آن را با بهینه‌ساز Adam و سایز batch ۳۲ و تعداد ۲ گام برای Gradient Accumulation اجرا نماید.

ب) (۵ نمره) کد بخش قبل را با تعداد گام‌های ۴، ۸ و ۱۶ اجرا نمایید.

(۱۰ نمره) در هر آزمایش، دقت نهایی مدل بر روی داده‌های تست، زمان اجرای آموزش و میزان حافظه‌ی مصرفی GPU را به دست آورده و نمودار تغییراتشان را بر حسب تغییر تعداد گام‌های Gradient Accumulation رسم نمایید. نتایج به دست آمده را تحلیل نموده و سناریوهایی که استفاده از این تکنیک پیشنهاد می‌شود را بگویید.

۴. (۱۵ نمره) در این سوال قصد بررسی بهینه‌سازهای مختلف را داریم. در این سوال، سایز batch را برابر ۱۲۸ فرض کنید. کد بخش (۱-ج) را با ۳ بهینه‌ساز SGD، Adagrad و RMSprop اجرا نمایید و بهینه‌سازهای Adam، SGD، Adagrad و RMSprop را بر اساس دقت بر روی داده‌های تست و حافظه‌ی مصرفی GPU مقایسه نمایید.

۵. (۱۰ نمره) در این سوال قصد بررسی استفاده از backend های مختلف communication در PyTorch DDP را داریم. در این سوال با بهینه‌ساز Adam و با سایزهای batch ۳۲ و ۱۲۸ backend های gloo و nccl را از نظر حافظه‌ی مصرفی GPU و زمان آموزش مقایسه نمایید.

نحوه تحویل پروژه

فایل‌ها را به صورت زیر نام گذاری کرده و در آخر همه را در یک فایل zip در سامانه ارسال کنید:

۱- گزارش report.pdf

۲- نام گذاری کدها را به صورت زیر انجام دهید:

نام فایل	بخش	سوال
classifier.py	الف	۱
classifier_mp.py	ب	
classifier_torchrun.py torchrun.sh (اسکرپت اجرای torchrun را در گزارش یا در فایل torchrun.sh بنویسید)	ج	
-	-	۲
classifier_torchrun_acc.py	الف	۳
-	-	۴
-	-	۵