



به نام خدا

Hardware for AI - بهار ۱۴۰۴



## تمرین چهارم : پیاده‌سازی سطح پایین شبکه‌های عصبی کانولوشنی

طراح: امید صالحی‌نیا، علی علیپور

### هدف پروژه :

هدف این تمرین، پیاده‌سازی نزدیک به سخت افزار و سطح پایین شبکه‌های عصبی کانولوشنی است. در نهایت در این تمرین با زبان C مدل نهایی تمرین قبلی خود را پیاده‌سازی خواهید کرد.

### مقدمه:

شبکه‌های عصبی مبتنی بر کانولوشن (CNN) از انواع مدل‌های یادگیری عمیق هستند. این نوع از شبکه‌ها مناسب برای مسائل پردازشی روی داده‌ها مانند طبقه‌بندی عکس، شناسایی اشیاء و مسائل گوناگون بینایی ماشین هستند. در تمرین قبل با پیاده‌سازی این شبکه‌ها و کوانتیزاسیون آن‌ها در پایتون آشنا شده‌اید. اما زبان برنامه نویسی بسیاری از سیستم‌های نهفته، C است. بنابراین، در این تمرین، شبکه CNN تمرین قبل به زبان C پیاده‌سازی می‌شود.

### پیاده‌سازی به زبان C:

**نکته مهم :** در تمامی کدهای C خواسته شده این تمرین، نیاز است زمان اجرا، مقدار حافظه اشغال شده و تعداد دسترسی به حافظه برای استنتاج یک داده گزارش شود.

1) ابتدا باید شبکه CNN تمرین ۳ را با زبان برنامه نویسی C پیاده‌سازی کنید. یکی از روش‌های مناسب، تبدیل مدل به فرمت ONNX و استفاده از ابزار ONNX2C است. اما شما می‌توانید از هر روش و کتابخانه دلخواهی به شرط توضیح استفاده کنید.

(۲) حال نیاز است عملکرد صحیح شبکه را پس از بازنویسی با زبان C بسنجیم. برای این کار سه داده از داده های تست CIFAR10 انتخاب کنید. logit های شبکه در پایتون و C را برای آن داده ها بسنجید و مقایسه کنید.

برای این بخش دقت کنید که داده های پس از پیش پردازش (در کد پایتون) را به عنوان ورودی مدل در C استفاده کنید. پس برای اجرا در زبان C نیز مانند پایتون به پیش پردازش نیاز است.

(۲.۱) (امتیازی) دقت شبکه را در داده های تست CIFAR10 در پیاده سازی C به دست آورید و با دقت پیاده سازی در پایتون مقایسه کنید. برای این کار نیاز است که داده های تست را در یک فایل با فرمتی مناسب ذخیره کنید تا کد C بتواند از آن فایل داده ها را بخواند.

(۳) در تمرین های قبل به چندی سازی شبکه پرداخته شده است. در این بخش نیز نیاز است چندی سازی انجام شود. دقت کنید توابع چندی سازی باید توسط شما نوشته شود و اگر از ONNX استفاده می کنید، نمی توانید از کوانتایزر های آن استفاده کنید.

در این بخش، اگر از روش اضافه کردن گره چندی سازی در ONNX استفاده کنید، نمره امتیازی به شما تعلق خواهد گرفت (این روش با استفاده از کوانتایزر های خود ONNX متفاوت است).

(۴) چندی سازی برای ۱۶ بیت پیاده سازی کنید و برای ۳ داده از داده های تست CIFAR10 اجرا کنید. این داده ها باید همان داده های سوال ۲ باشد و logit ها مقایسه شود.

(۵) چندی سازی برای ۸ بیت و ۶ بیت پیاده سازی کنید و برای سه داده از داده های تست CIFAR10 اجرا کنید. این داده ها باید همان داده های سوال ۲ و ۴ باشد و logit ها مقایسه شود.

(۶) زمان اجرای هر بخش مدل را در همه موارد گزارش کنید. به نظر شما چه بخش هایی نیاز به بهینه سازی دارند؟ به نظر شما، در صورت فراهم بودن امکان پیاده سازی سخت افزاری، چه بخش هایی بهتر است روی سخت افزار شتابدهی شود؟

توصیه : بخش های امتیازی پیچیده تر هستند هرچند یادگیری آن ها بسیار مفید است. توصیه می شود در صورتی که زمان کافی دارید به آنها بپردازید.

در صورت وجود هرگونه اشکال میتوانید با ایمیل زیر در ارتباط باشید:

[mahdimn2011@yahoo.com](mailto:mahdimn2011@yahoo.com)

[omidsn2@gmail.com](mailto:omidsn2@gmail.com)

[alipoura364@gmail.com](mailto:alipoura364@gmail.com)

## سایر نکات

- - انجام این تمرین به صورت گروه های دونفره خواهد بود.
- فایل ها و گزارش خود را با نام `HWA1_HW4_<SID>.zip` به ترتیب در محل های مربوطه در صفحه درس آپلود کنید.
- نام گذاری صحیح متغیرها، تمیزی کد و توضیحات می تواند تا حدودی کاستی های کد را در بخش های دیگر جبران کند.
- - هدف این تمرین یادگیری شماسست! در صورت کشف تقلب، مطابق با قوانین درس برخورد خواهد شد.
- به ازای هر  $x$  روز تأخیر، ۲ به توان  $x$  (با شمارش از صفر) از نمره شما کسر خواهد شد.

موفق باشید