

به نام خدا



دانشگاه تهران

پردیس دانشکده‌های فنی

دانشکده مهندسی برق و کامپیوتر



درس یادگیری عمیق با کاربرد در بینایی ماشین و پردازش صوت

تمرین شماره ۱

اسفند ۱۳۹۹

❖ مقدمه

در این تمرین هدف پیاده‌سازی یک شبکه عصبی ساده Fully Connected و برآورده کردن نیازهای مسئله می‌باشد. به همین منظور در سوالات پیاده‌سازی به هیچ عنوان اجازه استفاده از کد یا ابزارهای آماده جهت طراحی شبکه‌های عصبی عمیق مانند Pytorch و Tensorflow را نخواهید داشت و تمامی کدها را باید با زبان برنامه‌نویسی Python و با استفاده از کتابخانه‌های متداول آن مانند Numpy انجام دهید.

❖ سوالات

۱. در این سوال می‌خواهیم با پیاده‌سازی ساختار پایه‌ای از شبکه‌های عصبی تجربه عملی خود در این زمینه را تقویت و با مفاهیم و اصول آن آشنایی بهتری پیدا کنید. شبکه‌های عصبی عمیق برای کاربرد های متفاوتی استفاده می‌شوند که در این سوال از آن‌ها در مسائل ساده طبقه بندی، رگرسیون خطی و رگرسیون لجستیک بهره خواهید گرفت. برای این سوال شما از دیتاست UTKFace استفاده خواهید کرد. برای دسترسی به این پایگاه داده می‌توانید از [این لینک](#) استفاده کنید. برای صرفه جویی در حجم اینترنت شما، دیتاست در صفحه درس در سامانه elearn نیز بارگذاری شده است اما برای آشنایی با دیتاست از لینک قرار داده شده استفاده کنید.

این دیتاست دارای ۲۰۰۰۰ تصویر صورت است که از افراد ۰ تا ۱۱۶ ساله جمع شده است و هر تصویر حاوی چهار دسته اطلاعات مربوط به سن، جنسیت و قومیت افراد و تاریخ ثبت آن است. تصاویر اصلی دارای بک گراند و در ابعاد متفاوت هستند اما برای این تمرین شما تنها از تصاویر تراز و بریده شده استفاده خواهید کرد که با عنوان Aligned&Cropped Faces در سایت قابل دسترسی هستند. برای آموزش (train) شبکه در مسائل بخش‌های بعد از ۱۸۰۰۰ و برای ارزیابی آن (test) از ۲۰۰۰ تصویر به صورت تصادفی استفاده کنید. همچنین ۱۰٪ از داده‌های آموزش را به عنوان داده‌های اعتبار سنجی (validation) در نظر بگیرید.

بخش اول:

در این بخش هدف پیاده‌سازی یک شبکه عصبی چند لایه (Multi-layer Neural Network) برای کاربرد رگرسیون خطی است که بتواند سن افراد را تخمین بزند. همانطور که پیشتر گفته شد سن افراد دیتاست بین ۰ تا ۱۱۶ سال است.

دقت کنید که شبکه پیاده‌سازی شده می‌بایست امکان اضافه کردن تعداد لایه ها و نیز تعداد نرون ها در هر لایه را داشته باشد اما لایه آخر تنها یک نرون خواهد داشت. برای این مسأله از تابع خطا (loss function) L2 استفاده کنید. در لایه‌های میانی و در لایه آخر از تابع فعال ساز (activation function) Leaky ReLU استفاده کنید. برای آموزش شبکه از روش SGD همراه با momentum استفاده کنید و ضریب momentum را برابر با 0.9 فرض کنید. نرخ

یادگیری اولیه را برابر با ۰.۰۱ در نظر گرفته و در طول آموزش آن را با گام های ثابت کاهش دهید. ورودی های شبکه را پیش از ورود به شبکه نرمال کرده (normalization) و سپس آن ها را با تبدیل PCA به 128 بعد کاهش دهید. اندازه دسته ها (batch size) را برابر با 128 بگیرید. برای اعمال PCA میتوانید از کتابخانه های آماده استفاده کنید).

با توجه به موارد زیر پیاده سازی خود را در هر مرحله تکمیل و به پرسش ها پاسخ دهید. در هر بخش نمودار loss را در طول آموزش بر حسب epoch برای داده های validation و train رسم کنید و آنها را تحلیل کنید (تمام مقایسه ها باید از روی نمودار صورت گیرند). در تحلیل های خود مفهوم overfit را نیز بررسی کنید. همچنین معیار جذر میانگین مربعات خطا (Root Mean Square Error) را نیز برای داده های test گزارش کنید.

الف: وزن دهی اولیه شبکه را ۰ در نظر بگیرید. چه نتیجه ای مشاهده میکنید. با شرح مختصری از نتایج خود روش وزن دهی اولیه مناسبی را انتخاب کرده و با استفاده از آن شبکه را آموزش دهید. دلیل انتخاب خود را به صورت خلاصه توضیح دهید. برای بخش های بعد از این روش انتخابی برای وزن دهی اولیه استفاده کنید.

ب: استفاده از گام های ثابت یکی از راه های کاهش نرخ یادگیریست. یک روش دیگر برای کاهش نرخ یادگیری استفاده از گام های متغیر است. در [این لینک](#) دو روش با طول پله متغیر با عنوان Step Decay و Time-Based Decay قرار داده شده است. یکی از این روش ها را انتخاب و شبکه را با آن آموزش و نتیجه را با حالتی که از گام ثابت استفاده میکردید مقایسه کنید.

بخش دوم :

در این بخش هدف آن است تا از ساختار شبکه ی بخش الف تنها با تغییر در لایه آخر برای کاربرد طبقه بندی افراد بر اساس قومیت استفاده شود. دقت شود که همچنان شبکه باید امکان تغییر در تعداد لایه های میانی و یا تعداد نرون های آن ها را داشته باشد. توجه شود که با توجه به آنکه مسأله این بخش طبقه بندی است، در لایه آخر باید بر اساس تعداد کلاس ها (قومیت های مختلف) نرون خروجی تعریف شود.

در این مسأله از تابع فعال ساز Leaky ReLU برای لایه های میانی و تابع فعال ساز Softmax برای لایه خروجی استفاده کنید. همچنین از تابع خطا Negative Log Likelihood برای آموزش شبکه بهره بگیرید. تمام موارد دیگر مشابه بخش قبل هستند.

بهترین ترکیب از تنظیمات ساختار بخش الف را انتخاب کنید و از آن استفاده کنید. در این بخش علاوه بر loss نمودار accuracy را هم برای داده های آموزش و اعتبار سنجی در طول آموزش رسم و آن ها را تحلیل کنید. پس از آموزش شبکه عمل کرد آن را با سنجش معیار accuracy روی داده های تست بررسی کنید.

بخش سوم (امتیازی: ۱۵ درصد از سؤال ۱):

در این بخش برای کسب امتیاز بیشتر میتوانید از شبکه بخش الف در کاربرد رگرسیون لاجستیک برای تشخیص جنسیت افراد استفاده کنید. تغییراتی که نیاز به اعمال آنها است را شرح داده و پیاده سازی کنید. در این بخش از تابع خطای Logistic Loss برای آموزش شبکه استفاده کنید. همچنین معیار accuracy را نیز در کنار loss برای داده‌های آموزش و اعتبار سنجی در طول یادگیری رسم کنید. در نهایت accuracy برای داده‌های تست را نیز گزارش کنید. در نهایت در رابطه با تفاوت سه مسأله رگرسیون لاجستیک، رگرسیون خطی و طبقه بندی توضیح دهید.

۲. تصویر تک کاناله‌ای با ابعاد 4×4 مانند زیر به یک لایه کانولوشنی وارد شده است.

۲	۱	۳	۳
۱	۶	۲	۵
۱	۴	۵	۲
۶	۶	۵	۳

لایه کانولوشنی از سه فیلتر مختلف به صورت زیر تشکیل شده است.

فیلتر اول

۰	۱	۲
۱	۰	۱
۲	۱	۰

فیلتر دوم

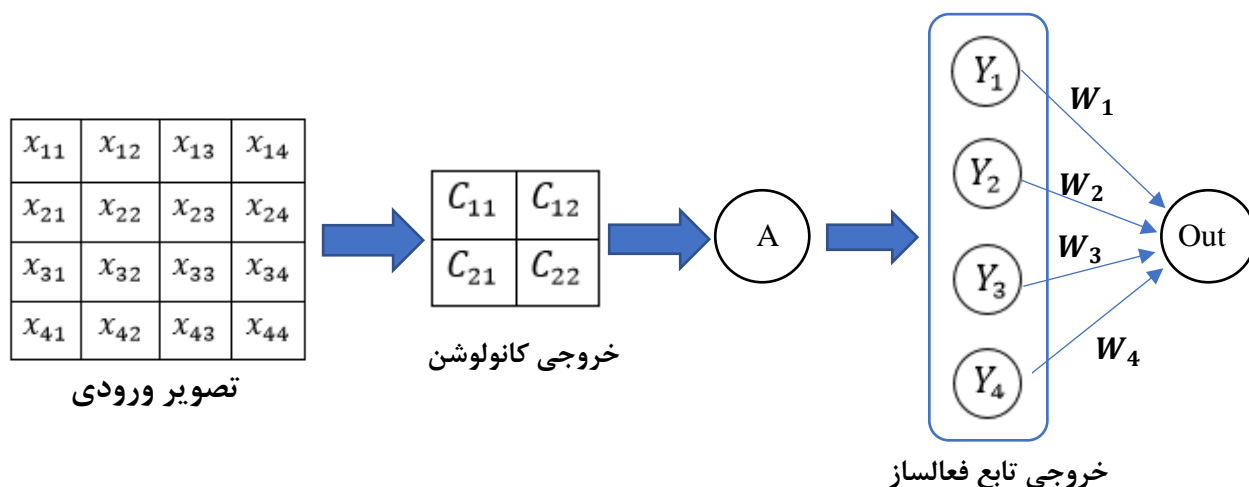
۱	۰	۱
۰	۱	۰
۱	۰	۱

فیلتر سوم

۱	۲	۰
۲	۰	۲
۰	۲	۱

حال با در نظر گرفتن $\text{Stride} = 1$ و $\text{Padding} = 1$ برای لایه کانولوشنی و همچنین با در نظر گرفتن لایه Max Pooling با ابعاد 2×2 و مقدار $\text{Stride} = 2$ و $\text{Padding} = 0$ ، ابعاد و خروجی لایه کانولوشنی و Pooling بعد از اعمال بر تصویر ورودی را با بیان مراحل کامل بدست آورید.

۳. شبکه عصبی ساده زیر را در نظر بگیرید.



شکل ۱: شبکه عصبی شامل یک لایه کانولوشنی و یک لایه خطی در انتها

با توجه به شکل شماره ۱ فرض کنید تصویری تک کاناله با ابعاد 4×4 به یک شبکه مانند بالا وارد شده است و در مرحله اول یک لایه کانولوشنی تک کاناله با فیلتر سایز 2×2 و $Stride$ با مقدار ۲ و بدون $Padding$ به آن اعمال شده است و خروجی آن ماتریس 2×2 نشان داده شده در تصویر بالا می‌باشد و در نهایت از یک تابع فعالساز A عبور کرده است که آن را تابع $g()$ فرض می‌کنیم. حال خروجی آن بعد از تغییر ابعاد به صورت یک بردار ۴ در ۱ می‌باشد که به یک لایه خطی ساده وارد شده است و خروجی out را تولید کرده است. برای سادگی کار ما از لایه $Pooling$ صرفه نظر کرده‌ایم و تابع خطا از نوع MSE می‌باشد که به صورت رابطه ساده شده زیر می‌باشد.

$$L = \frac{1}{2} (t - out)^2$$

با توضیحات داده شده در بالا عملیات $Back Propagation$ را بر روی این شبکه اعمال کنید و گرادیان L را نسبت به تمامی وزن‌ها که شامل وزن‌های لایه خطی و کانولوشنی می‌باشد بدست آورید. از مقدار بایاس صرفه نظر کنید. وزن‌های مربوط به لایه کانولوشنی را که در تصویر شماره یک نشان داده نشده است به دلخواه نام‌گذاری کنید.

نکات:

- مهلت تحویل این تمرین، تا پایان روز پنجشنبه ۵ فروردین ماه می‌باشد.
- انجام این تمرین به صورت **انفرادی** می‌باشد.
- برای انجام این تمرین فقط مجاز به استفاده از زبان برنامه‌نویسی **Python** خواهید بود. برای **پیاده‌سازی مدل شبکه اجازه استفاده از Pytorch را ندارید و تمامی موارد را باید خودتان پیاده‌سازی نمایید**. استفاده از کتابخانه‌های آماده فقط در مواردی که در صورت سوال اجازه استفاده از آن‌ها به شما داده شده باشد، مقدور خواهد بود.
- داخل کدها کامنت‌های لازم را قرار دهید و تمامی موارد مورد نیاز برای اجرای صحیح کد را ارسال کنید.
- **در صورت مشاهده موارد تشابه بین دو یا چند فرد در گزارش کار و یا کد، به طرفین تقلب نمره صفر داده خواهد شد و هیچ گونه عذر و بهانه‌ای از جمله ارسال کد به دوست خود و عدم آگاهی از کپی برداری کد شما پذیرفته نخواهد شد، بنابراین به هیچ عنوان کدهای خود را در اختیار دیگران قرار ندهید در غیر این صورت مسئولیت تقلب بر عهده شما نیز می‌باشد. همچنین کپی برداری از کدهای آماده موجود در اینترنت و یا استفاده از کدهای افراد ترم‌های گذشته تفاوت چندانی با تقلب ندارد و در چنین مواردی نیز نمره صفر به فرد تعلق می‌گیرد و جای هیچ‌گونه اعتراضی وجود ندارد.**
- اگر بخشی از کد را از کدهای آماده اینترنتی استفاده می‌کنید که جزء قسمت‌های اصلی تمرین نمی‌باشد، حتماً باید لینک آن در گزارش و کد ارجاع داده شود.
- گزارش شما در فرآیند تصحیح از **اهمیت ویژه‌ای** برخوردار است و نیمی از نمره شما را دربر خواهد گرفت. لطفاً تمامی نکات و فرض‌هایی که برای پیاده‌سازی‌ها و محاسبات خود در نظر می‌گیرید را در گزارش ذکر کنید و تمامی اصول نگارشی را مطابق با فایل ارسالی در صفحه درس رعایت بفرمایید.
- الزامی به ارائه توضیح جزئیات کد در گزارش نیست. اما باید نتایج بدست آمده را گزارش و تحلیل کنید.
- برای پیاده‌سازی می‌توانید از محیط **Colab** استفاده نمایید.
- لطفاً گزارش (در قالب PDF)، فایل کدها و سایر ضامین مورد نیاز را با فرمت زیر در صفحه درس در سامانه یادگیری الکترونیکی بارگذاری نمایید.

HW#_[Lastname]_[StudentNumber].zip

- در صورت وجود هرگونه ابهام یا مشکل می‌توانید از طریق رایانامه‌های زیر با دستیاران آموزشی مربوطه در تماس باشید.

○ پیمان باقرشاهی : p.baghershahi@ut.ac.ir

○ حسین پورمهرانی : h.pourmehrani@gmail.com