

به نام خدا



دانشگاه تهران  
پردیس دانشکده‌های فنی  
دانشکده برق و کامپیوتر



## درس یادگیری عمیق با کاربرد در بینایی ماشین و پردازش صوت

تمرین شماره ۳

اردیبهشت ۱۴۰۱

## فهرست سؤالات

مقدمه.....	۳
سوال ۱ - تفکیک بافت توموری به وسیله شبکه کانولوشنی.....	۴
دیتاست.....	۴
مدل U-Net.....	۴
تابع هزینه و معیارهای ارزیابی.....	۶
بخش الف - چرایی عدم استفاده از Padding.....	۶
بخش ب - آموزش مدل U-Net.....	۷
بخش ج - تأثیر تقویت دادگان.....	۷
بخش د - تغییر به مدل باقیمانده.....	۷
توضیحات مربوط به سؤال.....	۷
سوال ۲- تشخیص ادات سخن با استفاده از شبکه های بازگشتی.....	۹
بخش الف - پیش پردازش.....	۱۰
بخش ب - تبدیل متن به بردار.....	۱۰
بخش پ - پیاده سازی و ارزیابی شبکه ی بازگشتی.....	۱۰
بخش ت- انواع شبکه های بازگشتی.....	۱۱
بخش ث - شبکه دو طرفه.....	۱۱

در سؤال اول هدف آشنایی با مسئله‌ی تقطیع<sup>۱</sup> تصویر و کاربردهای صنعتی/تجاری آن در مواردی مانند خدمات درمانی است. مسئله‌ی تقطیع در کنار مسئله‌ی تشخیص اشیاء یکی از بهترین و پراستفاده‌ترین موارد کاربردی یادگیری عمیق در حوزه‌های مختلف هستند که آشنایی با روش‌ها و رویکردهای ارزیابی این گونه مسائل، می‌تواند تجربه‌ای ارزشمند برای دانشجویان باشد.

هدف بخش دوم این تمرین آشنایی بیشتر شما با شبکه‌های بازگشتی<sup>۲</sup> می‌باشد. این شبکه‌ها در موارد بسیار زیادی که ترتیب در آن‌ها اهمیت دارد کاربرد دارند. به‌دلیل آن‌که این شبکه‌ها یک توالی را در نظر می‌گیرند در پردازش زبان طبیعی بشدت مورد استقبال قرار گرفته است. در این بخش قصد داریم یکی از کاربردهای متن را با استفاده از این شبکه‌ها پیاده سازی کنیم.

---

<sup>۱</sup> Segmentation

<sup>۲</sup> Recurrent

## سوال ۱ – تفکیک بافت توموری به وسیله شبکه کانولوشنی

یکی از موفق‌ترین کاربردهای یادگیری عمیق در مسائل پزشکی، استفاده از این روش‌ها برای حل مسائل تقطیع عکس‌های پزشکی است. استفاده از مدل‌هایی مانند U-Net، FPN و Masked RCNN برای حل مسائلی از قبیل تقطیع بافت‌های سرطانی در ریه، تقطیع بافت تومور در مغز و تقطیع دهلیزهای مختلف قلبی به یک پدیده‌ی بسیار رایج در صنعت درمان هوشمند تبدیل شده است. در این سؤال، قصد داریم با استفاده از پیاده‌سازی مدل U-Net برای تفکیک بافت توموری مغزی در تصاویر MRI استفاده کنیم.

### دیتاست

دیتاست مورد استفاده در این سؤال، یکی از معروف‌ترین دیتاست‌های تومورهای مغزی است که بارها در Kaggle مورد بحث قرار گرفته است. این دیتاست عکس‌های MRI مغزی ۱۱۰ بیمار را نشان می‌دهد، به‌صورتی که عکس‌های موجود هر بیمار در یک فولدر خاص قرار گرفته است و این عکس‌ها در سایز ۲۵۶ در ۲۵۶ موجود می‌باشند. هر عکس در دیتاست یک تصویر برچسب نظیر دارد که نام آن به mask ختم می‌شود.

برای انجام این سؤال، ۱۱۰ بیمار را به سه بخش دادگان آموزشی و اعتبارسنجی و ارزیابی با نسبت ۰/۸، ۰/۱ و ۰/۱ با یک بذر تصادفی<sup>۱</sup> مشخص تقسیم نمایید. دقت کنید که این تقسیم بر روی بیمارانی صورت می‌پذیرد و نباید کل تصاویر را به‌صورت تصادفی تقسیم کنید. دیتاست مورد استفاده در این سؤال در [Kaggle](#) و [Google Drive](#) قابل دریافت است.

### مدل U-Net

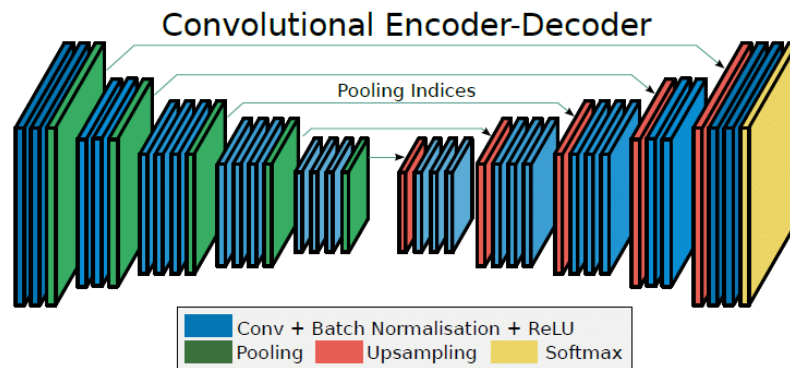
رویکرد رمزگذار<sup>۲</sup>-رمزگشا<sup>۳</sup> یکی از رویکردهای بسیار پرکاربرد برای حل مسائلی است که در آن نگاشتی از یک فضا به فضای دیگر داریم. در اولین روش‌های مبتنی بر این رویکرد، دو بخش شبکه را از هم متمایز و جدا می‌دانستند و این دو بخش برهمکنشی حداقلی داشتند، اما مشاهده شد که در مسائل مختلف ممکن است بتوان از اطلاعات از دست‌رفته‌ی حین فشردن‌سازی شبکه‌ی انکودر در بازسازی در شبکه‌ی دیکودر استفاده نمود. یکی از روش‌های اولیه تقطیع عکس با استفاده از رویکرد رمزگذار-رمزگشای مرتبط،

<sup>۱</sup> Random Seed

<sup>۲</sup> Encoder

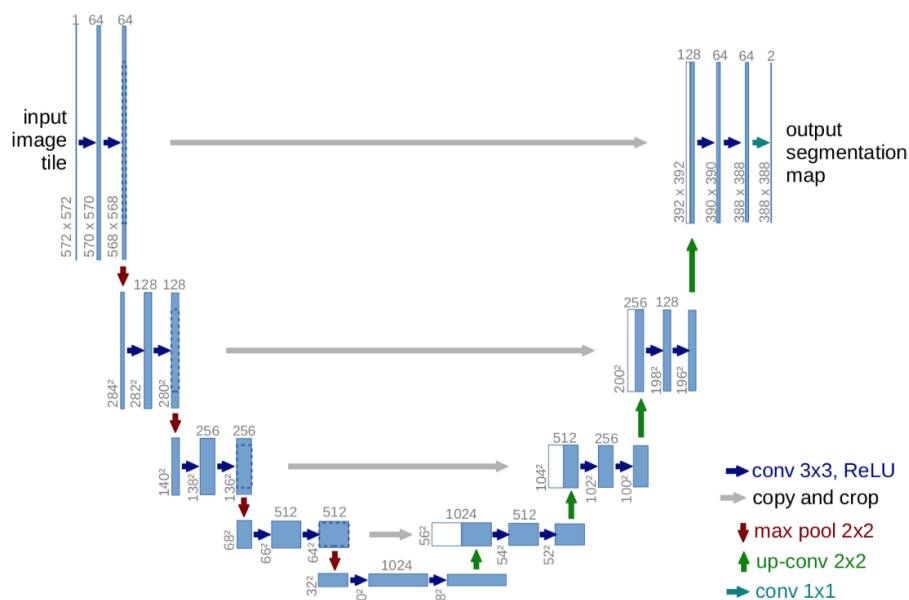
<sup>۳</sup> Decoder

SegNet بود که با استفاده از حفظ کردن اندیس‌های از دست رفته حین کاهش ابعاد به وسیله‌ی MaxPooling و استفاده از آن‌ها برای بازسازی تصویر حین افزایش ابعاد به روش MaxUnpooling سعی می‌کند که بین اطلاعات رمزگذار و رمزگشا ارتباط برقرار کند. (شکل ۱)



شکل ۱ - مدل SegNet یک مدل رمزگذار-رمزگشای کانولوشنی برای حل مسئله‌ی تقطیع که با استفاده از جا به جایی اندیس‌های خروجی Maxpooling سعی به بازسازی تصویر در رمزگشا می‌کند.

مدل U-Net (شکل ۲) با بسط دادن این رویکرد، به‌طور مستقیم از فضای ویژگی رمزگذار در بازسازی تصویر پس از افزایش ابعاد در شبکه‌ی رمزگشا استفاده می‌کند. در این روش فضای ویژگی خروجی هر بلوک رمزگذار، به فضای ورودی یک بلوک رمزگشای نظیر چسبانده می‌شود.



شکل ۲ شبکه‌ی کانولوشنی U-Net. این شبکه با چسباندن خروجی هر بلوک رمزگذار به ورودی هر بلوک رمزگشا، سعی می‌کند که اطلاعات موجود در ابعاد بالاتر را نیز حفظ کند و برخلاف مدل SegNet فقط به اندیس‌های خروجی لایه‌های توجه نکند. با این کار می‌توان ویژگی‌های پیچیده‌تری را در ابعاد بالاتر شبکه‌ی رمزگشا استخراج کرد که به ساخت ماسک‌های خروجی بهتری نسبت به مدل SegNet منجر می‌شود.

## تابع هزینه و معیارهای ارزیابی

در تمامی بخش‌های این سؤال تابع هزینه را در طول یادگیری برای دادگان آموزشی و اعتبارسنجی گزارش نموده و در نهایت برای دادگان ارزیابی نیز مقدار هزینه‌ی به دست آمده را در یک جدول گزارش نمایید. معیارهای ارزیابی ما برای سنجش عملکرد مدل، معیارهای Jaccard و Dice می‌باشند. در تمامی بخش‌های این تمرین این دو معیار را به صورت نمودار برای دادگان اعتبارسنجی و آموزش گزارش نمایید. در نهایت تابع هزینه و معیارهای ارزیابی را برای دادگان ارزیابی نیز گزارش نمایید. برای محاسبه‌ی این معیارها می‌توانید از کتابخانه‌ی <sup>۱</sup>torchmetrics استفاده نمایید.

جدول ۱ – معیارهای ارزیابی معروف برای مسائل تقطیع و تعبیر برداری آنان در فضای برداری. در برخی موارد از بعضی از این معیارها به عنوان تابع هزینه‌ی اصلی و یا تابع هزینه‌ی تنظیم‌کننده<sup>۲</sup> استفاده می‌شوند.

Similarity Measure	Binary Term Vectors	Weighted Term Vectors
Dice Coefficient	$\frac{2  X \cap Y }{ X  +  Y }$	$\frac{2 \sum x_i y_i}{\sum x_i^2 + \sum y_i^2}$
Cosine Coefficient	$\frac{ X \cap Y }{ X ^{\frac{1}{2}} \cdot  Y ^{\frac{1}{2}}}$	$\frac{\sum x_i y_i}{\sqrt{(\sum x_i^2 \cdot \sum y_i^2)}}$
Jaccard Coefficient	$\frac{ X \cap Y }{ X  +  Y  -  X \cap Y }$	$\frac{\sum x_i y_i}{\sum x_i^2 + \sum y_i^2 - \sum x_i y_i}$

## بخش الف – چرایی عدم استفاده از Padding

توضیح دهید که چرا در مدل اصلی یونت از کانولوشن‌های ۲ در ۲ استفاده می‌شود و چرا به مرز تصویر پیکسلی اضافه<sup>۳</sup> نمی‌شود. اهمیت این موضوع در کاربرد پزشکی چه می‌تواند باشد؟ (۵ نمره)

<sup>۱</sup> <https://torchmetrics.rtfd.io/en/latest/>

<sup>۲</sup> Regularization Loss

<sup>۳</sup> Pad

## بخش ب – آموزش مدل U-Net

با استفاده از این [لینک](#)، مدل آماده را دریافت کنید و بر روی دادگان آموزش دهید. نتایج مربوطه را گزارش کنید. همچنین به صورت خیلی کوتاه تفاوت‌های پیاده‌سازی این مدل با مقاله‌ی اصلی اشاره کنید. (۳۰ نمره)

## بخش ج – تأثیر تقویت دادگان

در این مرحله با استفاده از روش‌های تقویت دادگان سعی می‌کنیم تا عمل کرد مدل را بهبود ببخشیم. با استفاده از روش‌های ذکر شده در این [لینک](#) توضیح دهید که به نظر شما چه مدل تقویت دادگان برای این مسأله می‌تواند بامعنا باشد. با استفاده از این روش‌های تقویت دادگان بر روی دادگان آموزشی، بخش ب را تکرار نمایید. در مورد تفاوت نتیجه استدلال نمایید و در گزارش ذکر کنید. (۳۰ نمره)

## بخش د – تغییر به مدل باقیمانده

در این بخش سعی می‌کنیم که عمل کرد مدل را با تغییراتی، به یک مدل باقی‌مانده<sup>۱</sup> تبدیل نماییم. در ادامه توضیح داده می‌شود که این کار چگونه ممکن است. با استفاده از توضیحات آمده در ادامه سؤال، ساختار جدید را با روش‌های تقویت دادگان به دست آمده از بخش ج آموزش دهید و موارد لازم را گزارش کنید. خروجی به دست آمده را تحلیل کنید و در مورد تأثیر تغییرات اعمالی در ساختار مدل جدید استدلال نمایید. (۳۵ نمره)

### توضیحات مربوط به سؤال:

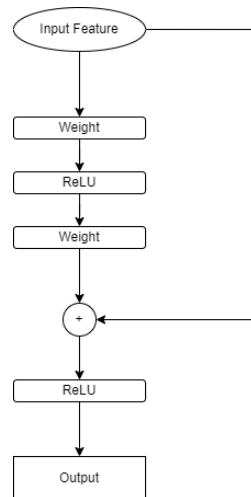
شکل ۳ یک مدل خاص از اتصال پس‌خور<sup>۲</sup> در بلوک باقی‌مانده را نشان می‌دهد که اصطلاحاً به آن اتصال پرشی<sup>۳</sup> می‌گویند. این نوع اتصال در مواردی استفاده می‌شود که کانال‌های ورودی با کانال‌های خروجی برابری می‌کند و می‌توان به راحتی فضای ویژگی ورودی بلوک را با فضای ویژگی خروجی بلوک جمع نمود. با توجه به اینکه در این سؤال کانال‌های ورودی و خروجی هر بلوک U-Net متفاوت است، لازم است که در اتصال پس‌خور، یک کانولوشن با کرنل ۱ در ۱ صورت پذیرد تا فضای ویژگی ورودی به خروجی نگاشت

<sup>۱</sup> Residual

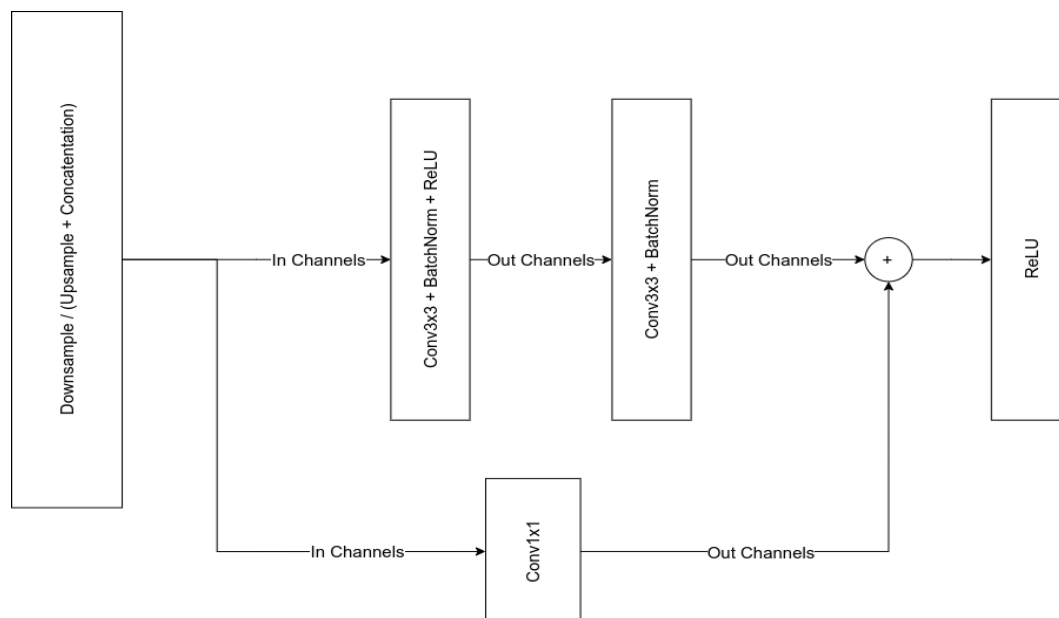
<sup>۲</sup> Feedforward

<sup>۳</sup> Skip Connection

شود. در نهایت بلوک‌های U-Net به صورت شکل ۴ تغییر خواهند کرد. شما باید در این سؤال با پیاده‌سازی این بلوک در هر بخش رمزگشا و رمزگذار، شبکه را به یک شبکه‌ی باقی‌مانده‌ای تغییر دهید.



شکل ۳- یک مدل از بلوک باقی‌مانده با اتصال پرشی. در این مدل سائز و تعداد کانال‌های تصویر ورودی و خروجی بلوک برابر است و به همین دلیل امکان جمع کردن بدون هیچ گونه تغییری بین ورودی پس‌خور و خروجی لایه‌های غیرخطی امکان‌پذیر است.



شکل ۴ بلوک باقی‌مانده یونت – دقت شود که شاکله‌ی اصلی بلوک برای رمزگذار و رمزگشا یکسان است و تنها ورودی هر لایه که کاهش‌دهنده ابعاد یا افزایش‌دهنده ابعاد است بسته به نوع لایه متغیر است. با توجه به اینکه کانال‌های ورودی با خروجی برابری نمی‌کند در این بلوک از کانولوشن‌های یک در یک استفاده می‌شود.



## سوال ۲- تشخیص ادات سخن<sup>۱</sup> با استفاده از شبکه‌های بازگشتی

تشخیص ادات سخن، یکی از مسائلی است که در پردازش متن به‌عنوان گام اولیه برای سایر کارها، کاربرد بسیاری دارد. تشخیص ادات سخن در واقع به‌منظور شناسایی نقش‌های کلمات در جمله به کار می‌رود.

برای مثال در جمله "ما بسیار کتاب می‌خوانیم" ادات سخن به‌صورت زیر می‌باشد:

('ما', 'PRO'), ('بسیار', 'ADV'), ('کتاب', 'N'), ('می‌خوانیم', 'V')

در مراجع مختلف، تعداد و انواع ادات متفاوت است. در این تمرین ما از [پیکره فارسی بیژن خان](#) استفاده می‌کنیم که دارای ۳۱ تگ متفاوت می‌باشد که می‌توانید انواع آن را در جدول ۲ مشاهده کنید:

جدول ۲ - ادات سخن به کار رفته شده در پیکره بیژن خان

Category	Description
ADJ	Adjective
ADJ_CMPR	Comparative adjective
ADJ_INO	Participle adjective
ADJ_SUP	Superlative adjective
ADJ_VOC	Vocative adjective
ADV	Adverb
ADV_COMP	Adverb of comparision
ADV_I	Adverb of interrogation
ADV_LOC	Adverb of location
ADV_NEG	Adverb of negation
ADV_TIME	Adverb of Time
CLITIC	Accuasative marker
CON	Conjunction
DELM	Delimiter
DET	Determiner
FW	Foreign word
INT	Interjection
N_PL	Plural noun
N_SING	Singular Noun
NUM	Numeral
N_VOC	Vocative noun
P	Preposition
PREV	Preverbal particle
PRO	Pronoun
SYM	symbol
V_AUX	Auxiliary verb
V_IMP	Imperative verb
V_PA	Past tense verb
V_PP	Past participle verb
V_PRS	Present tense verb
V_SUB	Subjective verb

<sup>۱</sup> Part-Of-Speech Tagger (POS Tagger)

## بخش الف – پیش پردازش

در گام اول شما می‌بایست پیش پردازش‌های لازم را بر روی متن اعمال کنید. برای این منظور می‌توانید از کتابخانه <sup>۱</sup> `parsinorm` استفاده نمایید. پیش پردازش‌های مورد استفاده را شرح دهید. دادگان را به سه بخش دادگان آموزشی<sup>۲</sup>، اعتبارسنجی<sup>۳</sup>، ارزیابی<sup>۴</sup> تقسیم کنید و تعداد و درصد داده‌های هر بخش را گزارش کنید.

## بخش ب – تبدیل متن به بردار

از آنجایی که کاربرد مورد نظر ما متن می‌باشد نمی‌توان آن را مستقیماً به شبکه اعمال کرد به دلیل آنکه ورودی شبکه‌های عمیق باید در قالب اعداد و بردارها باشد. باید ابتدا متن را به بردار تبدیل کرد و سپس به شبکه اعمال کنیم. برای این منظور از بردار تعبیه<sup>۵</sup> استفاده می‌کنیم. انواع مختلفی از بردارهای تعبیه وجود دارد. در این سوال می‌توانید از Glove (فایل آن در ضمایم تمرین به شما داده شده‌است) یا هر بردار تعبیه دیگری به غیر از بردارهای تعبیه وابسته به متن<sup>۶</sup> استفاده کنید. ساز و کار بردار تعبیه انتخاب شده خود را توضیح دهید.

## بخش پ – پیاده‌سازی و ارزیابی شبکه‌ی بازگشتی

در این بخش شما می‌بایست یک شبکه‌ی RNN پیاده سازی کنید. جزئیات شبکه طراحی شده را شرح دهید و دلایل انتخاب پارامترهای مورد نظر را توضیح دهید. نمودار هزینه بر حسب تعداد ایپاک‌ها را رسم کنید و همچنین مقادیر `accuracy`، `micro F۱`، `macro F۱` را مطرح کنید.

---

<sup>۱</sup> <https://pypi.org/project/parsinorm>

<sup>۲</sup> Train data

<sup>۳</sup> Validation data

<sup>۴</sup> Test data

<sup>۵</sup> Word embedding

<sup>۶</sup> Contextual word embedding

## بخش ت - انواع شبکه‌های بازگشتی

شبکه پیاده سازی شده را این بار با استفاده از LSTM و GRU پیاده سازی کنید و معیارهای بخش قبل را گزارش کنید. در بین این سه شبکه کدام یک بهترین عملکرد را داشته است؟ دلیل آن را شرح دهید.

## بخش ث - شبکه دو طرفه<sup>۱</sup>

بهترین شبکه‌ایی را که در بخش قبل به دست آوردید به صورت دوطرفه پیاده سازی کنید. نتایج به دست آمده را با بهترین شبکه مقایسه کنید. آیا بهبود حاصل شده است؟ دلیل آن چیست؟

---

<sup>۱</sup> Bidirectional

## نکات:

- مهلت تحویل این تمرین، پنجشنبه ۲۹ اردیبهشت است
- انجام این تمرین به صورت یک نفره می باشد.
- شما قادر نیستید هیچ تمرینی را با بیش از 7 روز تأخیر بارگذاری کنید (دقیقاً ۷ روز پس از مهلت آپلود، سامانه بسته خواهد شد).
- گزارش شما در فرآیند تصحیح از اهمیت ویژه‌ای برخوردار است. لطفاً تمامی نکات و فرض‌هایی که برای پیاده سازی‌ها و محاسبات خود در نظر می‌گیرید را در گزارش ذکر کنید. دقت داشته باشید ۵۰ درصد از نمره تمرین شما مربوط به گزارش است.
- کدهای خود را به صورت عکس در داخل گزارش کپی نکنید و با فرمتی مناسب آن را در گزارش قرار دهید.
- داخل کدها کامنت‌های لازم را قرار دهید و تمامی موارد مورد نیاز برای اجرای صحیح کد را ارسال کنید.
- الزامی به ارائه توضیح جزئیات کد در گزارش نیست. اما باید نتایج به دست آمده را گزارش و تحلیل کنید.
- گزارش را در قالب تهیه شده که روی صفحه درس در سامانه eLearn بارگذاری شده، بنویسید. در صورت تمایل می‌توانید از Latex نیز برای نوشتن گزارش استفاده نمایید.
- در گزارش خود برای تصاویر زیرنویس و برای جداول هم بالا نویس اضافه کنید.
- از آدرس دهی مطلق در کدهای خود استفاده نکنید و به جای آن از آدرس دهی نسبی استفاده نمایید.
- فایل‌های ارسال شده باید به فرمت py. باشد و از ارسال فایل تمرین‌ها به صورت ipynb خودداری نمایید. همچنین ساختار کلی کدهای شما باید حداقل شامل فایل‌های زیر باشد. همچنین پیشنهاد می‌گردد قسمت‌های بارگذاری داده، توابع مورد استفاده در فایل‌هایی مانند dataloader و utils قرار داده شود.

نام فایل	توضیح
model	ساختار مدل
main	کد آموزش و اجرای مدل

- کد شما باید قابلیت اجرا بر روی قسمت کوچکی از داده‌ها را داشته باشد تا دستیار آموزشی مربوطه بتواند با استفاده از کد شما در مدت زمان کوتاهی مدل شما را آموزش دهد.
  - در صورت مشاهده‌ی موارد تشابه بین دو یا چند فرد در گزارش کار و یا کد، به طرفین تقلب نمره صفر داده خواهد شد. کپی برداری از کدهای آماده موجود در اینترنت و یا استفاده از کدهای افراد ترم‌های گذشته تفاوت چندانی با تقلب 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، در غیر این صورت تقلب محسوب شده و کل نمره تمرین را از دست می‌دهید.
  - لطفاً فایل کدها و سایر ضمائم مورد نیاز را با فرمت زیر در صفحه درس در سامانه eLearn بارگذاری نمایید.
- HW۳\_[Lastname]\_[StudentNumber].zip

- در صورت وجود هرگونه ابهام یا مشکل میتوانید از طریق رایانامه زیر با دستیار آموزشی طراح تمرین فرهود اطاعتی (سوال اول) و رومینا اوجی (سوال دوم) در تماس باشید:

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

[Romina.oji@ut.ac.ir](mailto:Romina.oji@ut.ac.ir)