به نام خدا

دانشگاه تهران

دانشکده مهندسی برق و کامپیوتر

# درس شبکه‌های عصبی و یادگیری عمیق

# تمرین چهارم

| | | |
|---|---|---|
| پرسش ۱ | نام دستیار طراح | کیانا هوشانفر |
| | رایانامه | k.hooshanfar02@gmail.com |
| پرسش ۲ | نام دستیار طراح | احسان فروتن |
| | رایانامه | ehsan.forootan@ut.ac.ir |
| | مهلت ارسال پاسخ | ۱۴۰۴.۰۳.۰۴ |

# فهرست

# قوانین

قبل از پاسخ دادن به پرسش‌ها، موارد زیر را با دقت مطالعه نمایید:

- از پاسخ‌های خود یک گزارش در قالبی که در صفحه‌ی درس در سامانه‌ی Elearn با نام ***REPORTS_TEMPLATE.docx*** قرار داده شده تهیه نمایید.
- پیشنهاد می‌شود تمرین‌ها را در قالب گروه‌های دو نفره انجام دهید. (بیش از دو نفر مجاز نیست و تحویل تک نفره نیز نمره‌ی اضافی ندارد) توجه نمایید الزامی در یکسان ماندن اعضای گروه تا انتهای ترم وجود ندارد. (یعنی، می‌توانید تمرین اول را با شخص A و تمرین دوم را با شخص B و ... انجام دهید)
- **کیفیت گزارش شما در فرآیند تصحیح از اهمیت ویژه‌ای برخوردار است**؛ بنابراین، لطفا تمامی نکات و فرض‌هایی را که در پیاده‌سازی‌ها و محاسبات خود در نظر می‌گیرید در گزارش ذکر کنید.
- در گزارش خود مطابق با آنچه در قالب نمونه قرار داده شده، برای شکل‌ها زیرنویس و برای جدول‌ها بالانویس در نظر بگیرید.
- الزامی به ارائه توضیح جزئیات کد در گزارش نیست، اما باید نتایج بدست آمده از آن را گزارش و تحلیل کنید.
- **تحلیل نتایج الزامی می‌باشد، حتی اگر در صورت پرسش اشاره‌ای به آن نشده باشد.**
- **دستیاران آموزشی ملزم به اجرا کردن کدهای شما نیستند**؛ بنابراین، هرگونه نتیجه و یا تحلیلی که در صورت پرسش از شما خواسته شده را به طور واضح و کامل در گزارش بیاورید. در صورت عدم رعایت این مورد، بدیهی است که از نمره تمرین کسر می‌شود.
- **کدها حتما باید در قالب نوت‌بوک با پسوند .ipynb تهیه شوند، در پایان کار، تمامی کد اجرا شود و خروجی هر سلول حتما در این فایل ارسالی شما ذخیره شده باشد.** بنابراین برای مثال اگر خروجی سلولی یک نمودار است که در گزارش آورده‌اید، این نمودار باید هم در گزارش هم در نوت‌بوک کدها وجود داشته باشد.
- **در صورت مشاهده‌ی تقلب امتیاز تمامی افراد شرکت‌کننده در آن، 100- لحاظ می‌شود.**
- تنها زبان برنامه نویسی مجاز **Python** است.
- **استفاده از کدهای آماده برای تمرین‌ها به هیچ وجه مجاز نیست. در صورتی که دو گروه از یک منبع مشترک استفاده کنند و کدهای مشابه تحویل دهند، تقلب محسوب می‌شود.**
- نحوه محاسبه تاخیر به این شکل است: پس از پایان رسیدن مهلت ارسال گزارش، حداکثر تا یک هفته امکان ارسال با تاخیر وجود دارد، پس از این یک هفته نمره آن تکلیف برای شما صفر خواهد شد.

- سه روز اول: بدون جریمه
- روز چهارم: ۵ درصد
- روز پنجم: ۱۰ درصد
- روز ششم: ۱۵ درصد
- روز هفتم: ۲۰ درصد

- حداکثر نمره‌ای که برای هر سوال می‌توان اخد کرد ۱۰۰ بوده و اگر مجموع بارم یک سوال بیشتر از ۱۰۰ باشد، در صورت اخد نمره بیشتر از ۱۰۰، اعمال نخواهد شد.
  - برای مثال: اگر نمره اخذ شده از سوال ۱ برابر ۱۰۵ و نمره سوال ۲ برابر ۹۵ باشد، نمره نهایی تمرین ۹۷.۵ خواهد بود و نه ۱۰۰.
- لطفا گزارش، کدها و سایر ضمایم را به در یک پوشه با نام زیر قرار داده و آن را فشرده سازید، سپس در سامانه‌ی Elearn بارگذاری نمایید:

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

(مثال: HW1_Ahmadi_810199101_Bagheri_810199102.zip)

- برای گروه‌های دو نفره، بارگذاری تمرین از جانب یکی از اعضا کافی است ولی پیشنهاد می‌شود هر دو نفر بارگذاری نمایند.

# پرسش ۱. توصیف تصویر با شبکه ترکیبی ResNet50 + LSTM–GRU

## ۱-۱. مقدمه

Image-Captioning شاخه‌ای از پردازش تصویر و یادگیری ماشین است که به سیستم‌ها توانایی تولید خودکار توضیحات متنی برای تصاویر را می‌دهد. در این روش، ابتدا مدل‌های عمیق - معمولاً شبکه‌های عصبی پیچشی (CNN)- ویژگی‌های بصری تصویر را استخراج می‌کنند و سپس با کمک شبکه‌های بازگشتی (RNN) یا معماری‌های مبتنی بر ترنسفورمر، این ویژگی‌ها را به جملاتی روان و قابل‌فهم تبدیل می‌نمایند. کاربردهای این فناوری شامل دسترسی بهتر افراد نابینا به محتوای تصویری، جستجوی هوشمند تصاویر و تحلیل محتوا در شبکه‌های اجتماعی است.

در این سوال، قصد داریم برای تصاویر مجموعه داده‌ی Flickr8k کپشن‌های متنی تولید کنیم. برای این منظور از ساختارهای مختلف Encoder–Decoder استفاده خواهیم کرد که در مرحله‌ی اول ویژگی‌های بصری تصویر را با مدل‌های پیش‌آموزش‌دیده استخراج می‌کنند و در مرحله‌ی دوم این بردارهای ویژگی را به شرح‌های متنی معنادار تبدیل می‌نمایند.

## ۱-۲. مجموعه داده و پیش‌پردازش (۳۰ نمره)

### ۱. انتخاب مجموعه‌داده

ابتدا مجموعه‌داده‌ی Flickr8k را از وب‌سایت مربوطه دانلود کنید. چند نمونه تصویر همراه با کپشن‌های متناظر آن‌ها را نمایش دهید تا با ساختار داده و قالب فایل‌ها آشنا شوید.

### ۲. پیش‌پردازش تصاویر

تغییر اندازه: تمام تصاویر را به ابعاد ثابتی (مثلاً ۲۲۴×۲۲۴ پیکسل) درآورید تا به ورودی مدل CNN شما سازگار باشد.

نرمال‌سازی: مقادیر پیکسل‌ها را طوری مقیاس‌بندی کنید که توزیع آن‌ها نزدیک به صفر میانگین و انحراف‌معیار یک باشد. این کار باعث می‌شود یادگیری ویژگی‌های بصری توسط شبکه‌ی عصبی پایدارتر و سریع‌تر شود.

### ۳. پیش‌پردازش متن (کپشن‌ها)

حروف کوچک: همه‌ی حروف را به حروف کوچک تبدیل کنید تا مسأله‌ی حساسیت به بزرگی/کوچکی حروف برطرف شود.

حذف اضافات: علائم نگارشی، نمادهای غیرضروری و اعداد بی‌مورد را حذف نمایید.

**Tokenization:**

- هر کلمه را به یک شناسه‌ی عددی یکتا نگاشت کنید.
- دیکشنری‌ای بسازید که کل واژگان مجموعه‌ی شما را به اعداد متناظرشان اختصاص دهد.
- توکن‌های ویژه مانند <sos> (شروع جمله)، <eos> (پایان جمله)، <pad> (پرکننده) و <unk> (کلمات ناشناخته) را نیز به این دیکشنری اضافه کنید و کاربرد هر یک را توضیح دهید.
- این دیکشنری را در قالب یک فایل JSON ذخیره کنید تا به‌عنوان tokenizer در مراحل بعدی استفاده شود.

**یکنواخت‌سازی طول کپشن‌ها:**

برای تسهیل محاسبات در decoder، طول همه‌ی کپشن‌ها را برابر یک مقدار ثابت (مثلاً ۲۰ کلمه) در نظر بگیرید.

اگر کپشن کوتاه‌تر باشد، با توکن <pad> آن را تا طول موردنظر پر کنید.

توضیح دهید چرا اعمال padding در یادگیری مدل ضروری است (همه‌ی ورودی‌ها باید ابعاد یکسانی داشته باشند).

**۴. تقسیم داده‌ها**

مجموعه‌داده را به سه بخش با نسبت‌های ۸۰٪ آموزش (Train)، ۱۰٪ اعتبارسنجی (Validation) و ۱۰٪ تست (Test) تقسیم کنید. دقت نمایید که تقسیم‌بندی بر اساس تصاویر انجام شود؛ یعنی هیچ تصویری نباید در دو یا سه مجموعه تکرار شود.

## ۱-۳. پیاده‌سازی مدل (۳۰ نمره)

مدل‌های CNN–RNN از جمله متداول‌ترین روش‌های یادگیری عمیق برای مسائل چندرسانه‌ای مانند تولید توضیح متنی برای تصاویر به شمار می‌روند. در این ساختار، ابتدا یک شبکه‌ی پیچشی (CNN) به‌عنوان رمزگذار (Encoder) برای استخراج ویژگی‌های بصری تصویر به کار گرفته می‌شود و سپس یک شبکه‌ی بازگشتی (RNN) یا انواع پیشرفته‌تر آن مانند LSTM یا GRU به‌عنوان رمزگشا (Decoder)، با دریافت بردارهای ویژگی، اقدام به تولید جملات توصیفی می‌کند. در این بخش قصد داریم با الهام از طراحی‌ها و پارامترهای ارائه‌شده در این مقاله، یک مدل برای مسئله‌ی توضیح تصویر پیاده‌سازی کنیم.

**پیاده‌سازی بخش رمزگذار (Encoder):**

از یک مدل CNN پیش‌آموزش‌دیده (ResNet50) استفاده کنید. لایه‌ی Fully Connected نهایی را حذف کنید تا به‌جای کلاس‌بندی، بتوانید بردار ویژگی‌های تصویر را استخراج نمایید. ابعاد بردار خروجی ( feature vector) را بررسی و یادداشت کنید.

**پیاده سازی بخش رمز گشا (Decoder):**

از یک لایه‌ی Embedding برای نگاشت واژگان به بردارهای پیوسته استفاده کنید. (Embeddingها نسبت به نمایش One-hot فضای برداری فشرده‌تری ایجاد کرده و روابط معنایی بین کلمات را بهتر حفظ می‌کنند.)

ساختار decoder را مطابق مقاله (LSTM-GRU) برای تولید توالی کلمات پیاده کنید.

( بردار ویژگی تصویر را به‌عنوان حالت اولیه (initial hidden state) به LSTM بدهید. در انتها از یک لایه‌ی خطی (Linear) همراه با تابع Softmax برای پیش‌بینی توکن (کلمه) بعدی بهره بگیرید.)

**ادغام Encoder و Decoder (مدل End-to-End):**

یک کلاس سفارشی به‌نام HybridModel تعریف کنید که همزمان Encoder و Decoder را در خود جای دهد. پیاده‌سازی به‌گونه‌ای باشد که ابتدا تصویر را از طریق Encoder عبور دهد و سپس با استفاده از خروجی آن، دنباله‌ی متنی را با Decoder تولید کند. چگونه رمز گذار و رمز گشا را به یک مدل End-to-End تبدیل کنیم که قابلیت آموزش داشته باشد؟

## ۱-۴. آموزش و ارزیابی مدل

**آموزش (۱۵ نمره):**

از تابع هزینه مناسب برای محاسبه خطا استفاده کنید و مدل را با توجه به پارامترهای آموزش (نرخ یادگیری، اندازه‌ی batch، تعداد epoch و سایر تنظیمات) که در مقاله پیشنهاد شده‌اند، آموزش دهید.

**ارزیابی مدل (۲۵ نمره):**

- نمودار خطای داده آموزش و ارزیابی را در طول هر دوره (Epoch) گزارش کنید.
- الگوریتم Greedy Search و Beam Search را مقایسه کنید. توضیح دهید کدام کیفیت بهتری دارد.
- نتایج تولیدشده توسط الگوریتم‌های Greedy Search و Beam Search را محاسبه و مقایسه کرده، سپس پنج تصویر را انتخاب کنید و برای هر تصویر، کپشن‌های تولیدشده توسط هر دو روش را در کنار آن نمایش دهید.
- چند نمونه از خطاهای مدل را شناسایی و تحلیل کنید (برای مثال مواردی که مدل اشیاء را اشتباه تشخیص داده یا روابط بین آن‌ها را به‌درستی درک نکرده است).

## ۱-۴. امتیازی (۵ نمره)

در مورد معیارهای رایج ارزیابی مدل‌های تولید کپشن تحقیق کنید، سپس به‌طور مختصر عملکرد BLEU Score (BLEU-1 تا BLEU-4) را بر اساس n-gramها شرح دهید؛ در ادامه مقادیر BLEU-1 تا BLEU-4 را روی مجموعه تست محاسبه کرده و نتایج را در یک جدول گزارش و با تنظیمات مختلف مدل (Greedy vs. Beam Search) مقایسه کنید.

.

# پرسش ۲ – پیش بینی سری زمانی برای Clinical Event

## ۲-۱. مقدمه

مدلسازی موفقیت آمیز سریهای زمانی رویدادهای چندمتغیره پیچیده و توانایی آنها در پیشبینی رویدادهای آینده، برای کاربردهای مختلف در حوزه های علوم، مهندسی و کسب وکار حائز اهمیت است. در محیط‌های بالینی، توانایی در پیش‌بینی رویدادهای آینده برای یک بیمار بر اساس رویدادهای بالینی مشاهده‌شده در گذشته مانند نسخه‌های دارویی قبلی، آزمایش‌های انجام‌شده و نتایج آن ها، یا سیگنال های فیزیولوژیک گذشته می‌تواند به پیش‌بینی طیف گسترده‌ای از رویدادهای آینده کمک کند. این امر به متخصصان مراقبت‌های سلامت امکان می‌دهد تا پیش از وقوع رویداد مداخله کنند یا منابع لازم را برای مقابله با آن آماده نماید. در این سوال، با کمک روش های مختلف، سعی خواهد شد تا مدلی خوب برای پیش بینی مارکر های خونی انجام بپذیرد. به همین خاطر، دیتا MIMIC-III و یک قسمت خاص از آن، مربوط به آزمایشات خون بیماران در https://www.kaggle.com/datasets/salikhussaini49/prediction-of-sepsis انتخاب گردیده است. همچنین برای متدولوژی کلی پیش بینی زمانی، از مقاله https://people.cs.pitt.edu/~milos/research/2019/Lee_Hauskrecht_AIME_2019.pdf بهره برده شده است.

## ۲-۲. متدولوژی

الف)مساله پیش بینی زمانی به طور کلی چگونه انجام می پذیرد؟مدل های مختلف در پیش بینی زمانی را نام ببرید و در حد یک خط، برای هر کدام روش کلی را با کمک فرمول یا رسم شکل توضیح دهید. (۵ نمره)

ب) روش کلی State-Space Markov Event prediction چگونه انجام می پذیرد؟برای آن یک شبکه Dense طراحی بکنید و شکل خلاصه مدل را بیاورید. آیا استفاده از لایه هایی مانند Normalization مجاز است؟ از کدام Activation Function باید استفاده شود؟ برای تعداد نرون های پنهان از تنظیمات مقاله ( W برابر با ۱۲۸) استفاده بکنید. همچنین Learning Rate را از مقاله انتخاب بکنید. برای Batch-size، از مقادیر ۳۲، ۱۶ و یا ۱ استفاده بکنید. (برای بدست آوردن اندازه ورودی و خروجی ها به قسمت تحلیل آماری رجوع کنید) (۵ نمره)

ج) روش های LSTM-based event prediction چگونه کار می کنند؟روش و Loss-function آن را توصیف کنید و یک Forward pass از پیش بینی را توضیح دهید. (۵ نمره)

د) ساختار کلی Bi-Directional LSTM چگونه است؟ این مدل لایه چه مزیتی نسبت به لایه عادی LSTM را ارائه می کند؟۴ شبکه با به ترتیب GRU،Bi-Directional GRU، LSTM و Bi-Directional LSTM به وجود آورید و خلاصه مدل ها را ارائه بکنید. دقت بفرمایید که همه مدل ها برای پیش بینی لحظات بعد، نیازمند حداقل یک لایه Fully-Connected در Head خود هستند. از تنظیمات هایپرپارمتر مشابه قسمت ب بهره ببرید. (۵ نمره)

## ۲-۳. آماده سازی داده ها و تحلیل آماری

الف) داده های Train، Test و Validation برای شما در غالب فایل CSV قرار داده شده اند. آنها را بخوانید و ۵ ردیف اول را نشان بدهید. ستون 'Hour'، نشان گر ساعت است و نشان گر پیشرفت سری زمانی است. ماتریس Coorelation برای داده های Validation را بدست بیاورید. سپس ستون ها با پایین ترین Coorelation ها ( برای مثال زیر ۵ درصد) را انتخاب کنید و آنها را در یک شکل نشان بدهید. در انتها ۲۰ ستون با پایین ترین Cooreltaion را انتخاب بکنید. چرا باید پایین ترین Coorealtion انتخاب شود؟ در انتها، بر اساس شماره مریض، داده ها را گروه بندی کنید. به این معنا که هر بیمار، یک سری زمانی خواهد بود. (۵ نمره)

ب) سری زمانی یک مریض در ستون 'HR'را در طول Train، Validation و Test نشان بدهید. این سری زمانی را یکبار متشق گسسته بگیرید. تست ADF از کتابخانه Statmodels روی آن اجرا بکنید. آیا احتمال کمتر از ۵ درصد شد؟ انقدر عملیات تست و مشتق را ادامه بدهید تا احتمال کمتر از ۵ درصد شود. این تعداد برابر با d شما خواهد بود. (۵ نمره)

ج) سپس دو نمودار Autocorreltaion و Partial AutoCorrelation را رسم کنید. اولین محل کاهش شدید در دو نمودار را برابر با AR و MA در نظر بگیرید. سپس یک مدل SARIMAX با ورودی بیرونی به عنوان ستون 'O2Sat' در نظر بگیرید. مدل را فیت کرده، پیش بینی زمانی را انجام داده، آن را روی نمودار اصلی 'HR' رسم بکنید. (۵ نمره)

د)مقادر R2-score برای تست را گزارش بکنید. همچنین نمودار پیش بینی و اصلی قسمت قبل را مقایسه و تحلیل کنید. در انتها، همه داده ها را با MinMax نرمالایز بکنید. دقت کنید ستون ساعت باید جداگانه نرمالایز شود. همچنین دقت کنید که ستون های Validation و Test را با استفاده از نرمالایزری

که برای آموزش استفاده کردید، غیر از ستون ساعت، نرمال بکنید و ساعت را بر روی خود این داده ها MinMax بکنید. در اینجا چرا از نرمالایزر های بر پایه متوسط و واریانس استفاده نشده است؟ (۵ نمره)

## ۲-۴. آموزش مدل های یادگیری عمیق

الف) داده های قسمت قبل باید به صورت دوگانه Winodw شوند. یک Window، نمایانگر طولی از سری زمانی است که مدل میبیند و Window دیگر، نمایانگر طول زمانی است که باید پیش بینی بشود. این دو طول را چگونه می توان به صورت تقریبی از قسمت قبل متوجه شد؟ دو طول پنجره ها را گزارش کنید. دقت شود که برای آموزش در این بخش، باید از ۲۰ ستون با کمترین Coorelation استفاده شود. همچنین معیار Explained Variance Score را نیز توضیح دهید. چرا این معیار می تواند برای سری های زمانی مناسب باشد؟ (۵ نمره)

ب) مدل های Dense، LSTM و GRU را بر روی داده های آموزش، آموزش بدهید. در یک نمودار برای هر Epoch، مقدار تابع هزینه را برای آموزش و صحت سنجی گزارش بکنید. دقت کنید که سری زمانی هر بیمار خود یک گروه داده به حساب می آید و به تعداد بیماران ، سری زمانی دارید. (۱۰ نمره)

ج) مدل های Bi-Directional را آموزش داده و نمودار قسمت قبل را تکرار کنید. (۱۰ نمره)

د) برای تمامی مدل ها، در یک جدول، مقادیر Loss، MSE، MAE، R2-score و Cosine-Distance را گزارش بکنید. بهترین مدل از نظر R2-Score کدام است؟ آیا بهترین مدل از صرفا گزارش میانگین به عنوان پیش بینی بهتر عمل کرده است؟ (۵ نمره)

## ۲-۵. رسم نتایج و تحلیل جواب ها

برای هر ۵ مدل یادگیری عمیق، پیش بینی ها در طول زمان برای ستون 'HR' را رسم کنید. کدام مدا ها بهتر از SARIMAX عمل کرده اند؟ نتایج را هم روی متریک ها و هم روی روند ها برای هر ۶ مدل تحلیل کنید. (۱۵ نمره)

## ۲-۶. روش Maximum Log-Likelihood Estimation

الف) فرض شود که یک سیستم مدیریت منابع مالی بیمارستان، به اطلاعات آماری بیماران و پیش بینی این آمار نیازمند است. این سیستم، تنها یک تقریب درجه ۲ از تابع توزیع در هر لحظه را نیازمند است. بنابراین، تابع توزیع مورد حدس، به شکل زیر طراحی می شود:

$$\mu(t) = E\{X(t)\}$$

$$\sigma(t,s) = E\{X(t)\,X(s)\}$$
$$P(X(t) = X) = f(\mu(t),\sigma(t))$$

هدف این روش، مینیم کردن اختلاف مقدار P حدس زده شده توسط شبکه و P واقعی پنجره خروجی است. یک شبکه برپایه LSTM طراحی بکنید که از یک Window ورودی، مقادیر میانگین و واریانس در هر لحظه برای هر ستون را دریافت کرده و در خروجی، مقادیر میانگین و واریانس در هر لحظه برای هر ستون را خروجی بدهد. چرا جدا کردن واریانس و میانگین ها برای هر ستون و مستقل فرض کردن آنها، درست است؟ آیا می توان ساختار های قبل را استفاده کرد؟ آیا همچنان می توان از Loss-Function بخش های قبل بهره برد؟ آیا نیازمند تغییر Activation Function در لایه آخر هستیم؟ **(۵ نمره)**

ب) با کمک شبکه، داده ها را آموزش بدهید. نتایج Loss-Function در طول آموزش برای داده های آموزش و Validation را در کنار همدیگر در یک نمودار نمایش دهید. مقدار R2-score برای داده های تست را گزارش بکنید. **(۱۰ نمره)**

ج)به داده ها یک نویز با توزیع Lacplace اضافه بکنید. با استفاده از شبکه Bi-LSTM بخش ۴، سری زمانی برای یک بیمار را پیش بینی بکنید. سپس نمودار میانگین و واریانس را برای این پیش بینی رسم بکنید.(برای ستون 'HR') همچنین از مدل قسمت ۲-۶-ب برای پیش بینی میانگین و واریانس در زمان استفاده کرده و در کنار نتایج Bi-LSTM و همچنین میانگین و واریانس داده اصلی قرار بدهید. کدام روش نسبت به نویز مقاوم تر بوده است؟ چرا؟**( ۱۰ نمره )**