

# تمرین سری ۳ واحد درسی شبکه عصبی

جناب آقای دکتر خردپیشه

دستیاران آموزشی : نوید کاشی ، احمد رضا شریفیان زاده ، علی شریفی

۳ خرداد ۱۴۰۰

توجه کنید شما میتوانید بر روی کگل یا کولب و یا کامپیوتر های شخصی خود کار کنید .  
به جای دانلود و آپلود دیتاست در گوگل درایو برای استفاده در کولب میتوانید به شیوه زیر عمل کنید .

**چگونه از دیتاست های کگل در کولب استفاده کنیم ؟**

ددلاین تمرین تا تاریخ ۲۰ روز بعد از امتحان درس شبکه عصبی مقطع کارشناسی ارشد میباشد .

## ۱ تمرین

در تمرین جاری قصد داریم با توییت های تویتر کار کنیم . امروز تویتر زندگی ما را تغییر داده است . دیتاست جاری شامل توییت های درخصوص وقایع طبیعی میباشد . هدف ما این است که بفهمیم آیا این توییت در خصوص یک واقعه طبیعی واقعی است و یا نه یک واقعه طبیعی جعلی . در اصل ما یک کلاسیفایر نیاز داریم که ورودی متن توییت ها را دریافت کند و در آخر بیان کند مربوط به کدام یک از دو کلاس ما میباشد .

**لینک دیتاست**

به صورت کلی در هنگام کار با متن نیاز است که پیش پردازش هایی برای استفاده از متن انجام گیرد . این اعمال عبارتند از :

- حذف حروف اضافه که تحت عنوان stopwords شناخته میشوند .
- تمامی حروف به حالت lower case در آیند .
- نشانه هایی همانند هشتگ و ... حذف گردند .

---

• Lemmatization

• Stemming

• استاندارد کردن حالت های مختلف یک کلمه . به طور مثال برای tomrw یا 2mrw حالت استاندارد tomorrow درج گردد . یا به جای کلمه b4 ، کلمه before درج گردد .

• تبدیل کلمات به وکتور جهت کار در مدل ها به این کار Text Vectorization میگوییم که روش های مختلفی از قبیل Word2Vec ، Bag-of-Words و ... وجود دارد .

مراحل میتوان به صورت خلاصه به صورت زیر مطرح کرد .

### 1. Preprocessing:

- Normalize
- Stemming
- Lemmatization
- Stop words

### 2. Text to vector

- Word2vec or bag of words or ...

### 3. Classify

RNN یا LSTM مختلف های معماری از می توانید بندی کلاس بخش برای . کنید استفاده

دو تا از پکیج های مهم در زمینه nlp در پایتون عبارتند از :

• Gensim

• nltk

از منابع زیر میتوانید جهت آشنایی بیشتر در خصوص NLP استفاده کنید .

• [text-vectorization](#)

• [preprocessing-nlp](#)

• [pytorch in lstm](#)

• [beginners for nlp](#)

• [nlp-preprocessing](#) حتما این لینک را بررسی کنید .

## ۲ تمرین های امتیازی

۱. این دیتاست از یک مسابقه فعال کگل استفاده شده است . در صورتی که کد شما جز ۳۰ درصد برتر در جدول امتیازی گرفت لطفا با ذکر گزارش دقیق و جداگانه به ما اطلاع دهید تا نمره امتیازی به شما تعلق گیرد .
۲. در صورت استفاده از ایده transformer و ارتقا جواب ها ، نمره امتیازی به شما تعلق میگیرد . برای آشنایی با ترانسفرمرها میتوانید از [لینک](#) استفاده کنید .
۳. سعی کنید مدلی پیاده سازی کنید که در آن از مکانیزم attention استفاده شده باشد .