



دانشگاه تربیت مدرس شهید رجایی

دانشگاه تربیت مدرس شهید رجایی

تمرین سری 1 درس مباحث ویژه 2 – یادگیری ژرف پیشرفته

تاریخ انتشار تمرین: سه شنبه، ۲۹ آبان ۱۳۹۷

مهلت: چهارشنبه، ۳۰ آبان ۱۳۹۷

1. ایجاد نام دایناسور (35%).

مجموعه داده‌ی زیر را دانلود کنید:

<http://dataset.class.vision/NLP/dinos.txt>

در هر سطر یک نام نوشته شده است.

دانشمندان معمولاً از قواعد نامگذاری مشخصی برای نام گذاری این گونه‌ی جانوری منقرض شده استفاده می‌کنند.

با آموزش یک مدل character level بازگشتی با فریم ورک Keras سعی کنید چند نام جدید تصادفی خروجی بگیرید.

2. تشخیص پایتخت کشورها (15%).

مجموعه داده‌ی زیر را دانلود کنید:

<http://dataset.class.vision/NLP/country-capital.txt>

سپس یک word embedding از قبل آموزش داده شده را لود کنید

از این مجموعه داده 5 درصد را به صورت تصادفی انتخاب کنید و با word analogy پایتخت آن‌ها را حدس زده و کارایی

روشتان را چاپ کنید (چند درصد پایتخت‌ها را درست حدس زده؟)

برای سادگی کار می‌توانید در کدتان اضافه کنید که اگر نام کشور یا پایتخت از حروف انگلیسی تشکیل نشده بود یا فاصله داشت آن

نمونه را بررسی نمی‌کنید و در کارایی محاسبه نخواهد شد.

3. استفاده از GRU (25%).

از شبکه‌ی GRU استفاده کنید و مساله طبقه بندی متن emojiify [نوت بوک 47](#) را طبقه بندی کنید.

4. طبقه بندی با میانگین گیری (25٪)

از Keras استفاده کنید و قسمت نخست نوت بوک 47 مربوط به روش میانگین گیری بردارهای embedding برای طبقه بندی را بازنویسی کنید.