

تمرین سری ۴ درس NLP	مدرس: دکتر مرضیه داودآبادی فراهانی دستیاران آموزشی: مرتضی حاجی آبادی، غزل زمانی نژاد
------------------------	--

سوالات تئوری

1. دو جدول زیر تعداد Unigram ها و Bigram ها در یک پیکره فرضی را نشان می‌دهد.

ما	خواندیم	دیروز	امروز	داستان	کتاب
۱۸۷۲	۱۴۹۵	۲۰۲۱	۱۹۴۳	۹۴۵	۱۲۴۵

	ما	خواندیم	دیروز	امروز	داستان	کتاب
ما	۰	۱۵۶	۴۱۱	۴۵۲	۲۳۸	۳۸۷
خواندیم	۲	۰	۶	۱۱	۸۴	۱۱۲
دیروز	۳۴۱	۳۲	۰	۴۸	۶۸	۲۵۴
امروز	۳۲۸	۱۲	۰	۰	۸۷	۲۳۱
داستان	۴	۳۴۵	۸۴	۳۸	۰	۱
کتاب	۳۱	۳۲۰	۳	۰	۴۰۳	۰

احتمال رخداد جملات تست زیر را محاسبه کنید. فرض بر این است که جملات تست در وسط یک

رشته هستند. یعنی در نظر گرفتن احتمال بندهای شروع و پایان جمله لازم نیست.

$$P(w_2 | w_1) = \frac{C(w_1, w_2)}{C(w_1)}$$

جمله تست ۱: ... ما امروز کتاب خواندیم ...

جمله تست ۲: ... ما دیروز داستان خواندیم ...

2. رابطه زیر را اثبات کنید.

$$P(w_1^n) = P(w_1)P(w_2|w_1)P(w_3|w_1^2) \dots P(w_n|w_1^{n-1}) = \prod_{k=1}^n P(w_k|w_1^{k-1})$$

3. روش beam search را می توان به صورت شبه کد زیر نوشت:

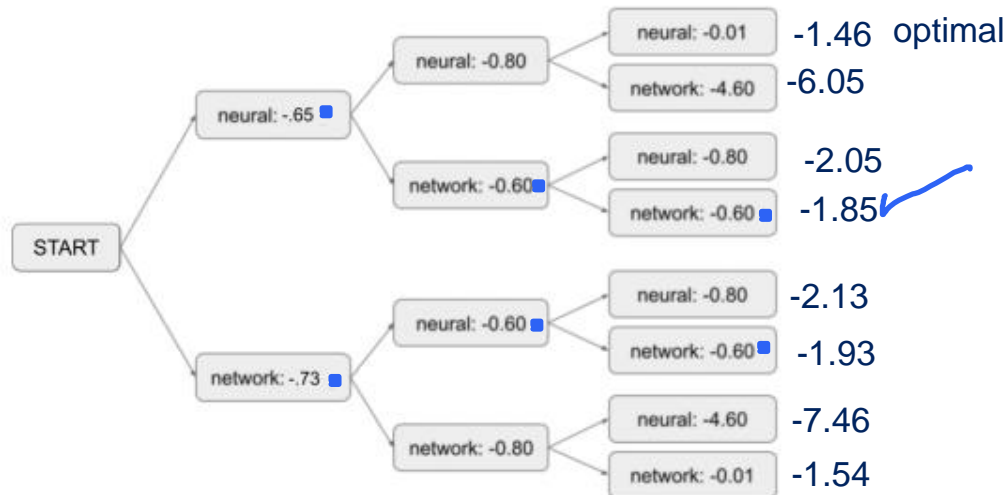
Algorithm 1 Beam Search

```

for each time step  $t$  do
  for each hypothesis  $y_{1:t-1,i}$  that we are tracking do
    find the top  $k$  tokens  $y_{t,i,1}, \dots, y_{t,i,k}$ 
  end for
  sort the resulting  $k^2$  length  $t$  sequences by their total log-probability
  store the top  $k$ 
  advance each hypothesis to time  $t + 1$ 
end for

```

$$O(T \cdot k \cdot M \cdot \log(M) + T \cdot k^2 \cdot \log(k^2)) = O(T \cdot k \cdot (M \cdot \log(M) + k \cdot \log(k))) = O(T \cdot k \cdot M \cdot \log(M))$$



Beam search را برای کدگشایی دنباله ای به طول 3 با $k = 2$ اجرا می کنیم. پیش بینی های یک کدگشا را در شکل بالا در نظر بگیرید، که در آن هر گره در درخت نشان دهنده پیش بینی احتمال ثبت توکن بعدی مشروط به توکن های قبلی در یک مرحله از کدگشا است (اعداد نشان داده شده، لگاریتم پیش بینی احتمال (log probability prediction) کدگشا از توکن فعلی با توجه به توکن های قبلی است). Vocabulary از دو کلمه تشکیل شده است:

"neural" و "network" $|M| = 2$

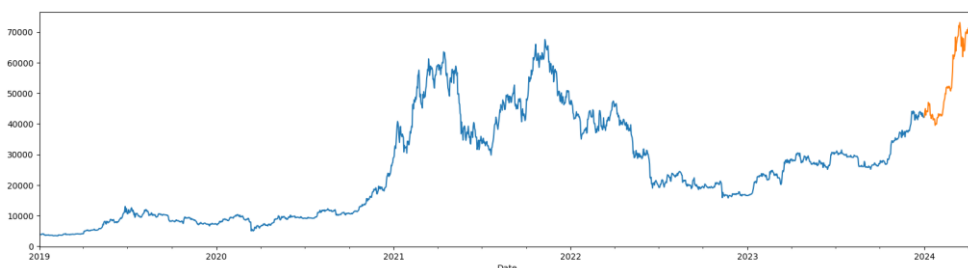
- در مرحله زمانی 1، beam search کدام دنباله را ذخیره می کند؟
- در مرحله زمانی 2، beam search کدام دنباله را ذخیره می کند؟
- در مرحله زمانی 3، beam search کدام دنباله را ذخیره می کند؟
- آیا beam search دنباله ی کلی با بیشترین احتمال (overall most-likely sequence) در این مثال را برمی گرداند؟ توضیح دهید.
- پیچیدگی زمان اجرا تولید یک دنباله با طول T با اندازه پرتو k با RNN چقدر است؟ بر حسب T و k و M پاسخ دهید (M اندازه vocabulary است).

4. ✓ به سوالات زیر پاسخ دهید.

- a. اگر در LSTM فقط بخواهیم گیت forget را داشته باشیم و گیت های input و output را حذف کنیم، چه اتفاقی می افتد و خروجی چه تغییری می کند؟
- b. اگر در یک LSTM مقدار گیت forget را به صفر تنظیم کنیم، چه اتفاقی می افتد و چطور این تغییر تاثیری بر روی توانایی شبکه در یادگیری و پیش بینی دارد؟
- c. توضیح دهید که چگونه افزایش تعداد لایه های LSTM در یک شبکه می تواند به کارایی و عملکرد شبکه کمک کند یا باعث افزایش پیچیدگی شود.

سوال عملی

5. در سال های اخیر توجه به رمزارزها بسیار گسترش یافته است. یکی از معروف ترین رمزارزهای موجود در بازار Bitcoin است. در این سوال قصد داریم قیمت Bitcoin را در آینده پیش بینی کنیم.
- a. ابتدا [کتابخانه yfinance](#) را نصب نمایید.
- b. حال میتوانید قیمت Bitcoin را دانلود نمایید. برای این کار از تابع download موجود در این [لینک](#) استفاده کنید. نماد شاخص مورد نظر برابر با BTC-USD است و تاریخ ذخیره سازی برای داده های آموزشی را از 2019-01-01 تا 2024-01-01 و تاریخ ذخیره سازی داده های آزمایشی را از 2024-01-01 تا آخرین روز موجود قرار دهید.
- c. برای آزمایش درستی مراحل فوق، نمودار این شاخص را بر حسب زمان رسم نمایید و به هر یک از داده های آموزشی و آزمایشی رنگ متفاوتی اختصاص دهید. نمودار حاصل شکلی مشابه با نمودار زیر خواهد داشت:



- d. برای آموزش بهتر مدل، مقادیر قیمت (محور y نمودار فوق) را نرمالیزه کنیم. برای این کار از تابع MinMaxScaler موجود در کتابخانه scikit-learn استفاده نمایید. لازم به توجه است که تنها داده های آموزشی به عنوان معیار scale کردن می باشند اما هر دو نوع داده های آموزشی و آزمایشی طبق این معیار scale می شوند.

e. در مرحله بعد داده‌های مورد نیاز برای آموزش و آزمایش مدل را تهیه می‌نماییم. برای این کار متغیری تعریف نمایید که نشان دهنده تعداد داده‌های گذشته برای پیش‌بینی داده مشخصی باشد. به عنوان مثال اگر این متغیر را برابر با ۶۰ قرار دهید، بدین معنی است که ۶۰ داده گذشته در پیش‌بینی آن داده تاثیرگذار هستند. طبق مثال فوق اندازه داده‌های این

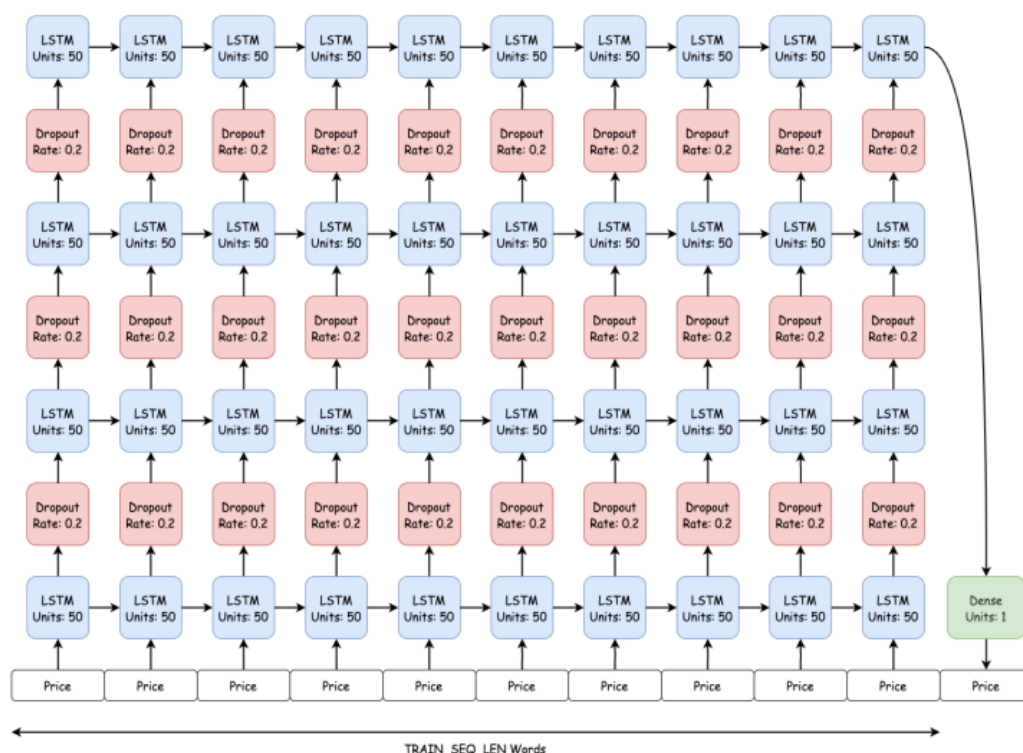
$$5 \times 365 - 60 = 1825 - 60 = 1765$$

مسئله به شرح زیر می‌شوند:

X_train.shape: (1766, 60, 1)

y_train.shape: (1766,)

f. حال مدل مشابه با شکل زیر بسازید:



g. مدل را در شرایط زیر آموزش دهید:

تابع ضرر: mean squared error

بهینه‌ساز: Adam

تعداد Epoch: 100

اندازه Batch: 32

h. پس از آموزش مدل، پیش‌بینی را بر روی داده‌های آزمایشی انجام دهید و نمودار را رسم

نمایید. در این نمودار که بر حسب زمان رسم می‌شود، هر دو مقدار واقعی و پیش‌بینی با

رنگ‌های متفاوت رسم شوند.

60

تشریحی i. به نظر شما افزایش یا کاهش متغیر تعریف شده در مرحله تهیه داده مورد نیاز برای آموزش

مدل یعنی تعداد داده‌های گذشته برای پیش‌بینی داده‌های مشخص چه مزایا یا معایبی دارد؟

شرح دهید.