

<p>تمرین سری 2</p> <p>درس NLP</p>	<p>نام مدرس: دکتر مرضیه داودآبادی فراهانی</p> <p>دستیاران آموزشی: امیرمحمد آزادی, بکتاش انصاری,</p> <p>سینا زمانی</p>
-----------------------------------	---

سوالات تئوری

۱. به سوالات زیر پاسخ دهید:

- تفاوت بین one-hot encoding و word embedding ها را بیان کنید.
- به طور خلاصه توضیح دهید که الگوریتم GloVe چگونه word embedding ها را تولید میکند.
- به طور خلاصه توضیح دهید که الگوریتم Word2Vec چگونه word embedding ها را تولید میکند.
- واژه‌های دارای چند معنی چگونه در word embedding ها کنترل می‌شوند و چه چالش‌هایی را تولید میکنند؟
- word embedding ها نیاز دارند که تمام کلمات در مجموعه آموزش حضور داشته باشند. چگونه با کلمات خارج از واژگان (out of vocabulary) برخورد می‌کنید؟ یک روش برای تولید word embedding برای کلماتی که در داده آموزش حاضر نبوده‌اند پیشنهاد دهید.

۲. ماتریس co-occurrence متن زیر را بنویسید. اندازه پنجره را ۲ در نظر بگیرید.

I love computer science and I love NLP even more.

سوالات عملی

به سوالات داخل نوتبوک درون خود آن پاسخ دهید.

۳. نوتبوک Q3 را تکمیل کنید.

۴. نوتبوک Q4 را تکمیل کنید.