



## مسئله ۱. Skip-gram

در روش Skip-gram به سوالات زیر پاسخ دهید:

- در این روش با ورودی دادن کلمه Center ، انتظار داریم کلمات Context آن پیش بینی شوند. اگر به جای این کار رویکردی مشابه Autoencoder داشته باشیم، یعنی آنکه سعی کنیم خود کلمه center را پیش بینی کنیم، چه مشکلی پیش می آید؟ به طور کلی تفاوت ساختار Word2Vec با ساختار Autoencoder چیست؟
- بعد از آموزش این مدل آیا ماتریسی را که فضای hidden را به فضای output نگاشت می دهد، می توانیم به عنوان یک embedding حساب کنیم؟ با توجه به جواب خود توضیح دهید به طور کلی چرا در این روش از دو ماتریس مختلف در مدل استفاده کردیم؟ آیا می توانستیم از یک ماتریس در هر دو بخش استفاده کنیم؟

## مسئله ۲. ترکیبی

- در بخش decoder مدل Transformer ، مکانیزم توجهی که روی ورودی decoder اعمال می شود، Masked Attention نام دارد که یعنی هر کلمه ورودی در decoder ، برای محاسبه attention تنها به کلمه هایی قبل از خود دسترسی دارد و کلمات بعدی به اصطلاح mask می شوند تا کلمه های قبل به آنها دسترسی نداشته باشند. علت این موضوع را شرح دهید.
- glove چگونه مزیت خود را نسبت به Word2Vec ایجاد می کند؟

(موفق باشید :)