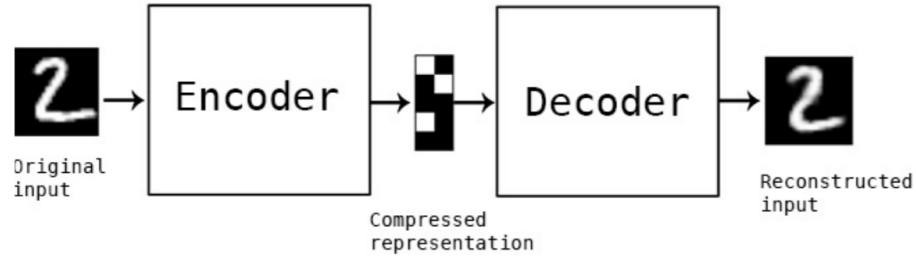


اسلاید مربوط به هر سوال زیر آن ذکر شده است.

سوال ۱:

در یک شبکه autoencoder فضای نهان ما فقط یک نود دارد و هیچ تابع activation ای نیز برای نود موجود در لایه نهان وجود ندارد. همچنین فرض کنید که ماتریس وزن های شبکه در بخش encoder و decoder با یکدیگر Transpose هستند. به عبارت دیگر  $\hat{x} = x W^T W$  باشد.

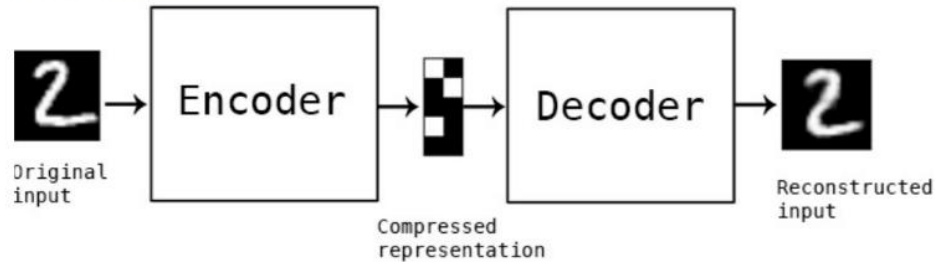
1. Autoencoders can be thought of as a non linear PCA.



$$\min_{h,g} \sum \|\hat{x} - x\|^2$$

$$\min_{h,g} \sum \|g(f(x)) - x\|^2$$

1. PCA can be described as



$$\min_W \sum \|\hat{x} - x\|^2$$

$$W^T W = I$$

$$\min_W \sum \|W^T W x - x\|^2$$

## سوال ۲:

مدل بردار سراسری یا به اختصار glove به منظور ترکیب مدل تجزیه ماتریس مبتنی بر شمارش (count-based matrix factorization) و مدل skip-gram مبتنی بر محتوا (context-based skip gram) ارائه گردید. همانطور که میدانیم تعداد به کارگیری کلمات و میزان رخداد همزمان آنها میتواند در تعیین معنای کلمات کمک کند. میدانیم که word2vec تنها از محتوی محلی (local context windows) برای تعیین بردار کلمات استفاده میکند در حالی که مدل glove با استفاده از تعداد رخدادهای همزمان کلمات در کل متن بردار کلمات را تعیین میکند.

1. GloVe becomes a global model for unsupervised learning of word representations that outperforms other models on word analogy, word similarity, and named entity recognition tasks.
2. Advantages
  - ▶ Fast training
  - ▶ Scalable to huge corpora
  - ▶ Good performance even with small corpus, and small vectors
  - ▶ Early stopping. We can stop training when improvements become small.
3. Drawbacks
  - ▶ Uses a lot of memory: the fastest way to construct a term-co-occurrence matrix is to keep it in RAM as a hash map and perform co-occurrence increments in a global manner
  - ▶ Sometimes quite sensitive to initial learning rate

## سوال ۳:

LSTMها مدل های ترتیبی هستند که باید واحدها (کلمات) به طور ترتیبی پردازش شوند. این فرآیند مانع از پردازش همزمان چند واحد و کاهش سرعت میشود. transformerها با استفاده از مکانیزم های self-attention و بدون استفاده از ساختارهای recurrent ساخته میشود. در این مکانیزم جمله ها به جای کلمه به کلمه، به صورت کلی پردازش میشوند. همچنین بر خلاف lstm به stateهای پیش از خود وابسته نیست و خطر فراموشی در ورودی های طولانی وجود ندارد.

در attention معمولی سیستم یاد میگیرد که به کدام قسمت از ورودی جدید توجه کند در حالی که self-attention تعدادی ورودی را میگیرد و نحوه ارتباط این ورودی ها را با هم بررسی میکند.

1. Consider the following example

#### Example (Self-Attention)

- ▶ Consider the following sentence  
**The animal didn't cross the street because it was too tired.**
- ▶ What does **it** in this sentence refer to?
- ▶ Is it referring to **the street** or to **the animal**?

2. Self-attention (intra-attention) is an attention mechanism relating different positions of a single sequence in order to compute a representation of the same sequence (Cheng, Dong, and Lapata 2016).
3. It is very useful in
  - ▶ Machine reading (the automatic, unsupervised understanding of text)
  - ▶ Abstractive summarization
  - ▶ Image description generation

1. The self-attention mechanism enables us to learn the correlation between the current words and the previous part of the sentence.

The FBI is chasing a criminal on the run .  
 The FBI is chasing a criminal on the run .  
 The FBI is chasing a criminal on the run .  
 The FBI is chasing a criminal on the run .  
 The FBI is chasing a criminal on the run .  
 The FBI is chasing a criminal on the run .  
 The FBI is chasing a criminal on the run .  
 The FBI is chasing a criminal on the run .  
 The FBI is chasing a criminal on the run .

2. The current word is in red and the size of the blue shade indicates the activation level.