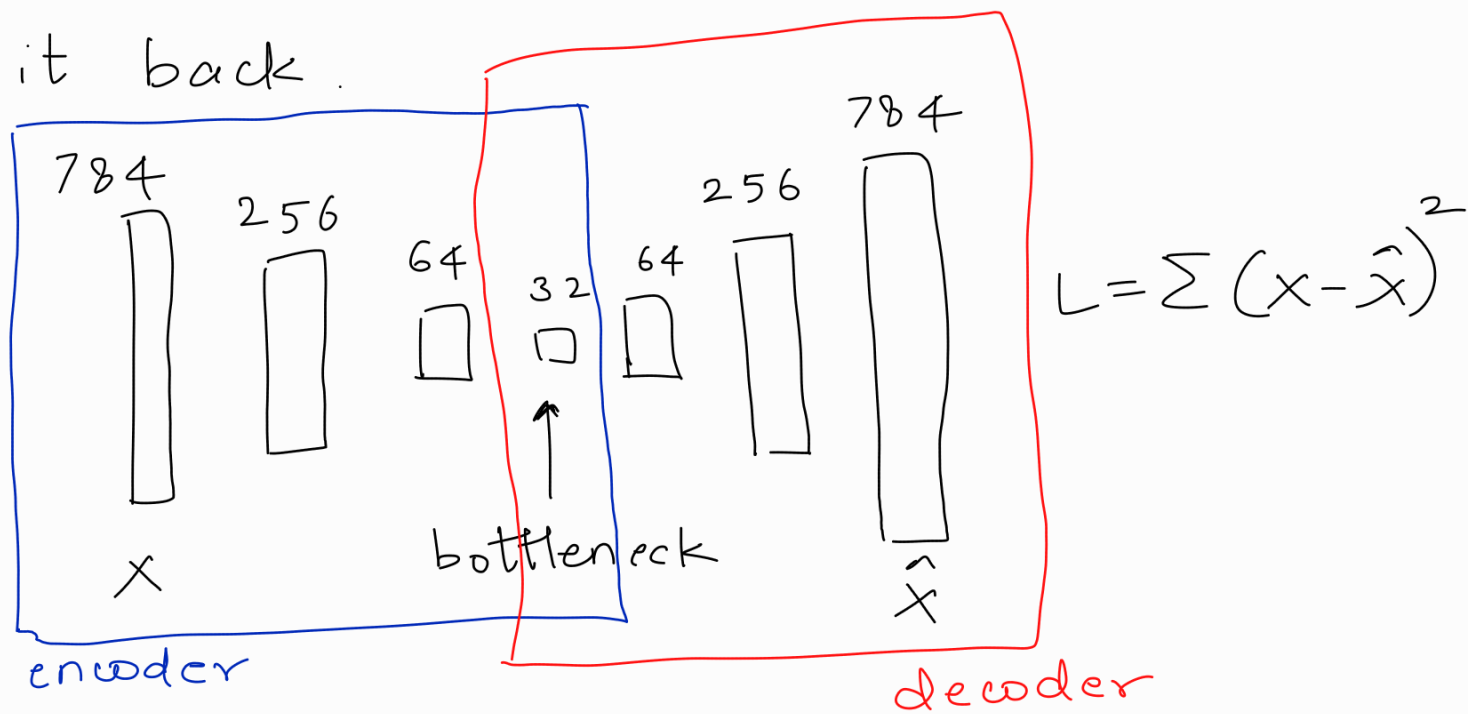


20/2/24

Send original data and try to predict it back.



Why do this?

If you do this successfully, it means that the 32 bottleneck has all the information required to represent the input. So, we have achieved compression. Reduced 784 to 32.

Works much better than PCA

Deeper auto-encoders work better

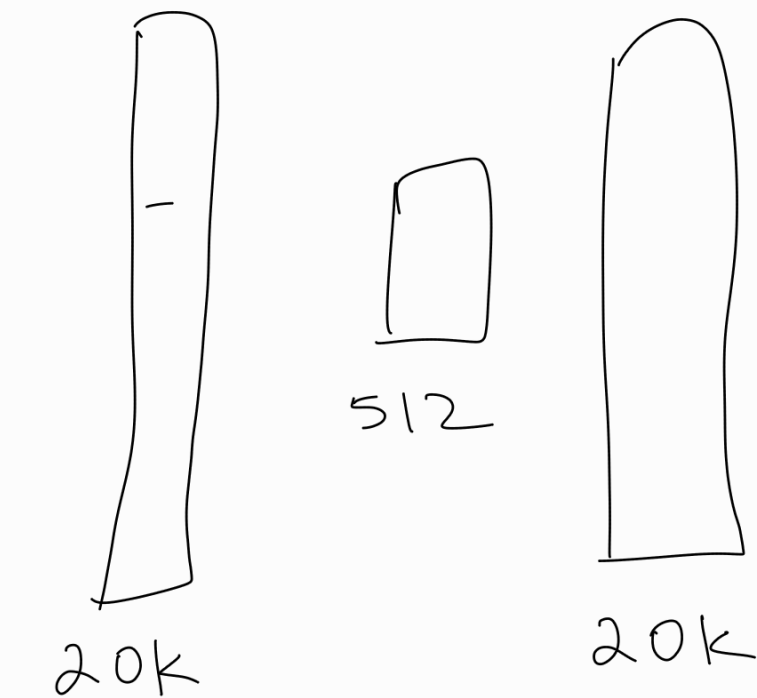
## Word 2 Vec

→ Earlier concept was one-hot embedding.

But we want embeddings of similar words to be similar.

→ One-hot does not capture context or similarity.

→ SKIP GRAM



$$L = -\log p$$

one-hot

→ predicts pre decided pairs

apple, juice

apple, drank