

Entity-embeddings

- Case-studies from the industry -

→ users; products; jobs; applicants;
sellers;

Let us first derive simple strategies based on techniques we already know

→ for entity embeddings

⑤

- very doable -

first-cut
solution

Later :

→ study solutions by various companies

⑤

Embeddings

→ Text

↪ Word2Vec

→ BERT / GPT

{ dense semantic vectors }

Embeddings vs Sparse Vectors

→ Semantics
are captured
better

TF-IDF

Bow

one-hot encoding
(categorical)



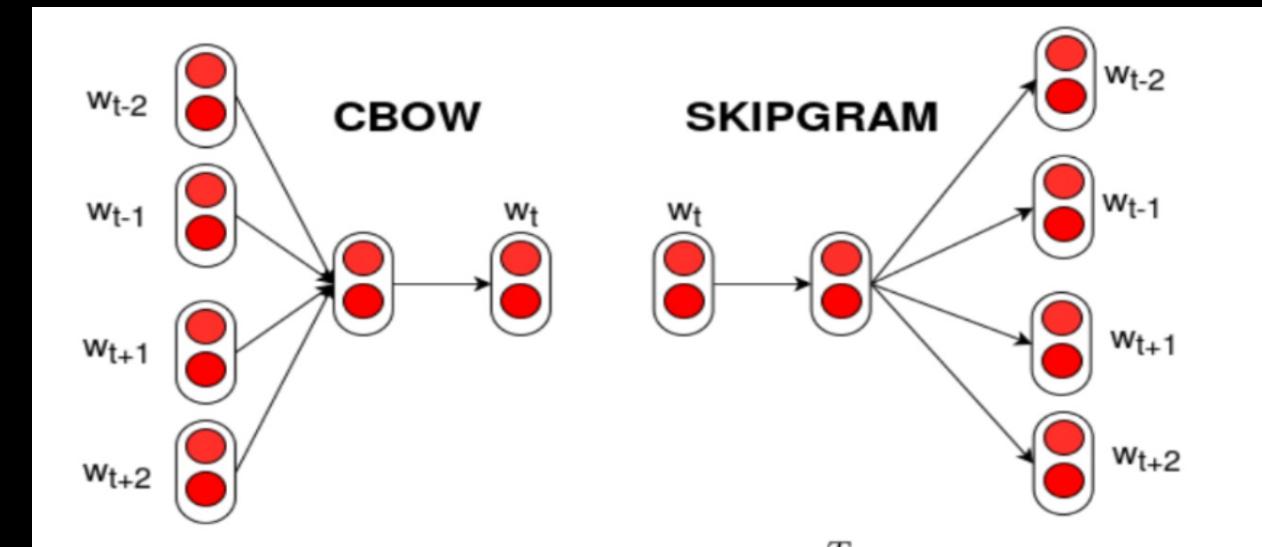
high
dimensional

word2vec

← focus word → context words

skip-gram

CBOW



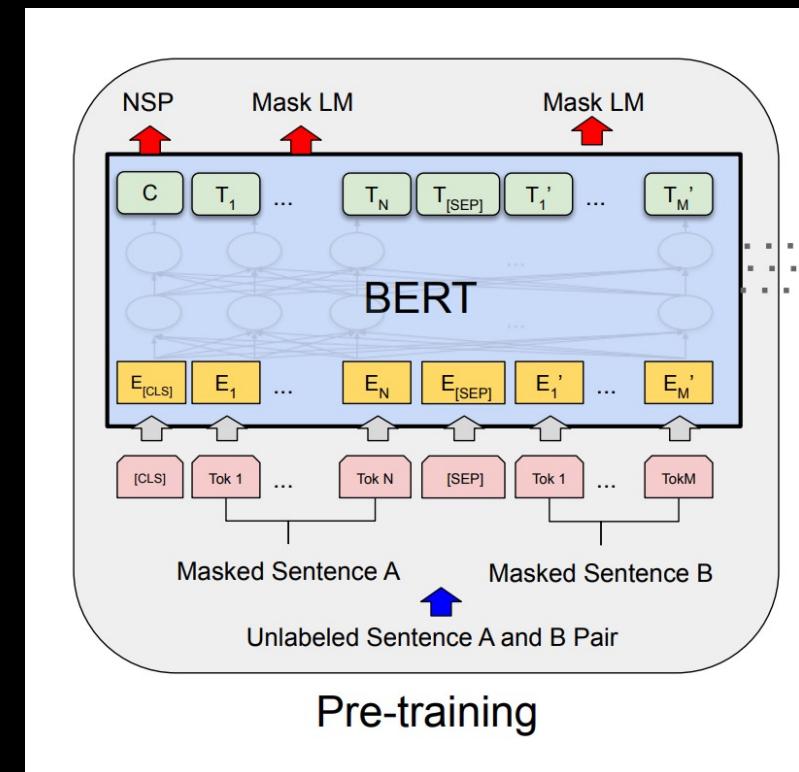
* Sequence of words

BERT (pretraining)

→ Self-attention & transformer based
(bi-directional)

→ sequence of words

→ Masked Language model
Next Sentence Prediction



[sequence]

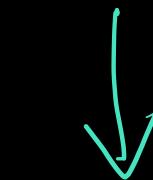
Key-idea: Sequence of words

1

→ very useful information

Self-supervision

Sequence *dalaá*
(large amounts) → Self-supervision

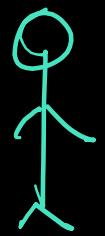


Word2Vec

BERT

(Q) Can you convert LinkedIn users
into a sequence?

[3 mins]



recursively

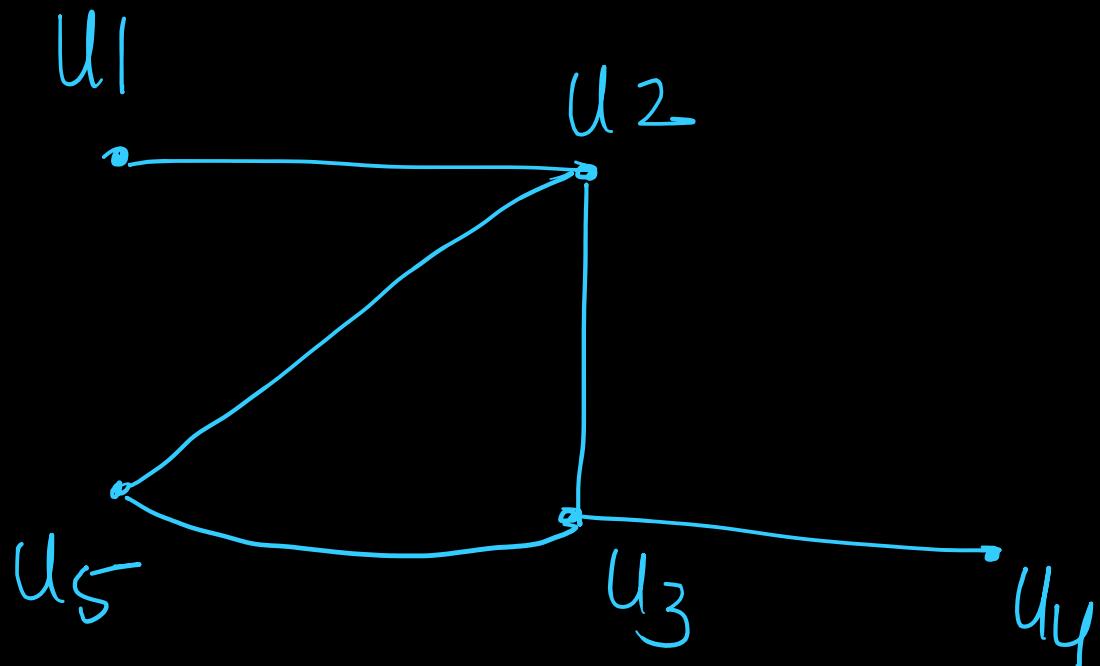
u_1

u_2

u_3

... .

Connections - graph:



random-walks:

$u_1 - u_2 - u_5 - u_3$

$u_1 - u_3 - u_4$

⋮

(Q) Can you convert job-postings
into a sequence?

Use γ 1

$$-C_1 - \gamma_1 -$$

$$-C_2 - \gamma_2 -$$

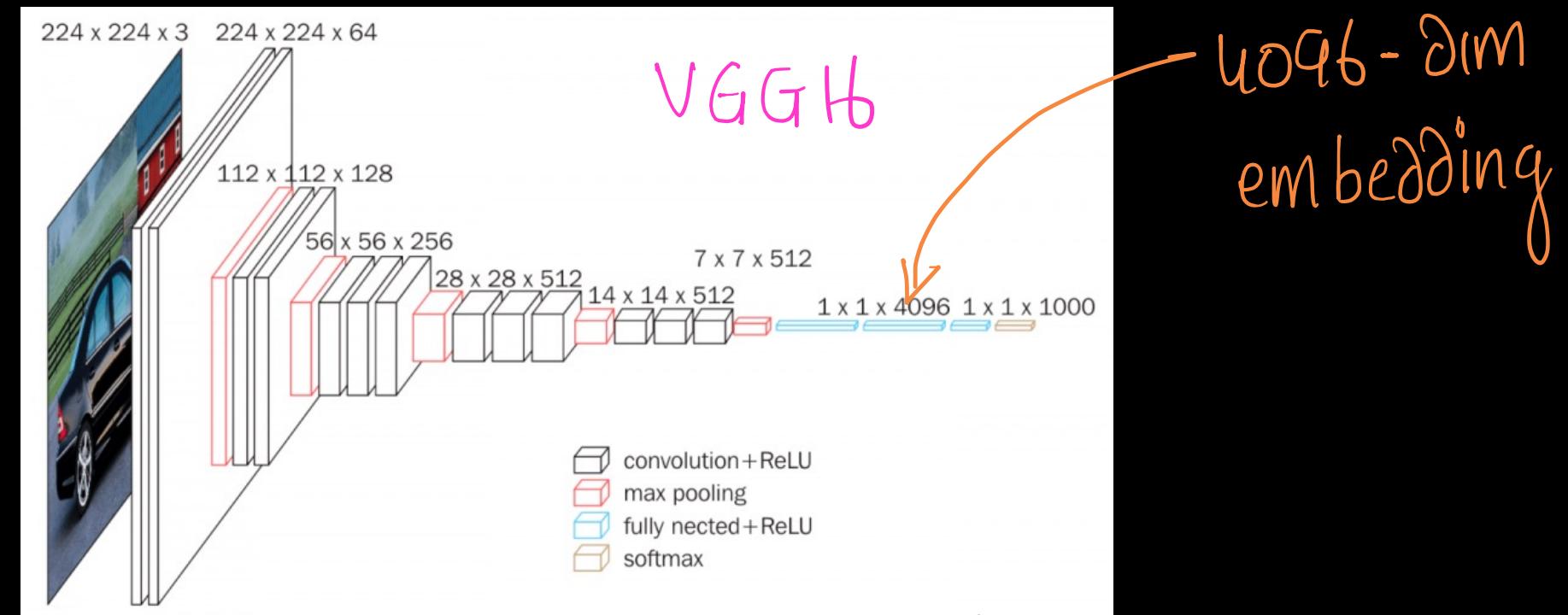
$$-C_3 - \gamma_3 -$$

(Q) Can you convert Amazon Products
into a sequence ?

(Q) Can you convert Amazon users
into a sequence?

Embeddings:

Images → Transfer-learning



visualize using t-SNE

<https://cs.stanford.edu/people/karpathy/cnnembed/>

→ Semantic similarity

based proximity

Images → ResNet 50 (popular)

↓
Pretrained for ImageNet classification

[Supervision]

Key-idea :

②

Deep Learning

+

Pretrained models for classification

+

large + general purpose dataset
(ImageNet)

can result in good embeddings.

(Q) How can we leverage this strategy [Classification] to embed/encode Amazon's users? [2 min]

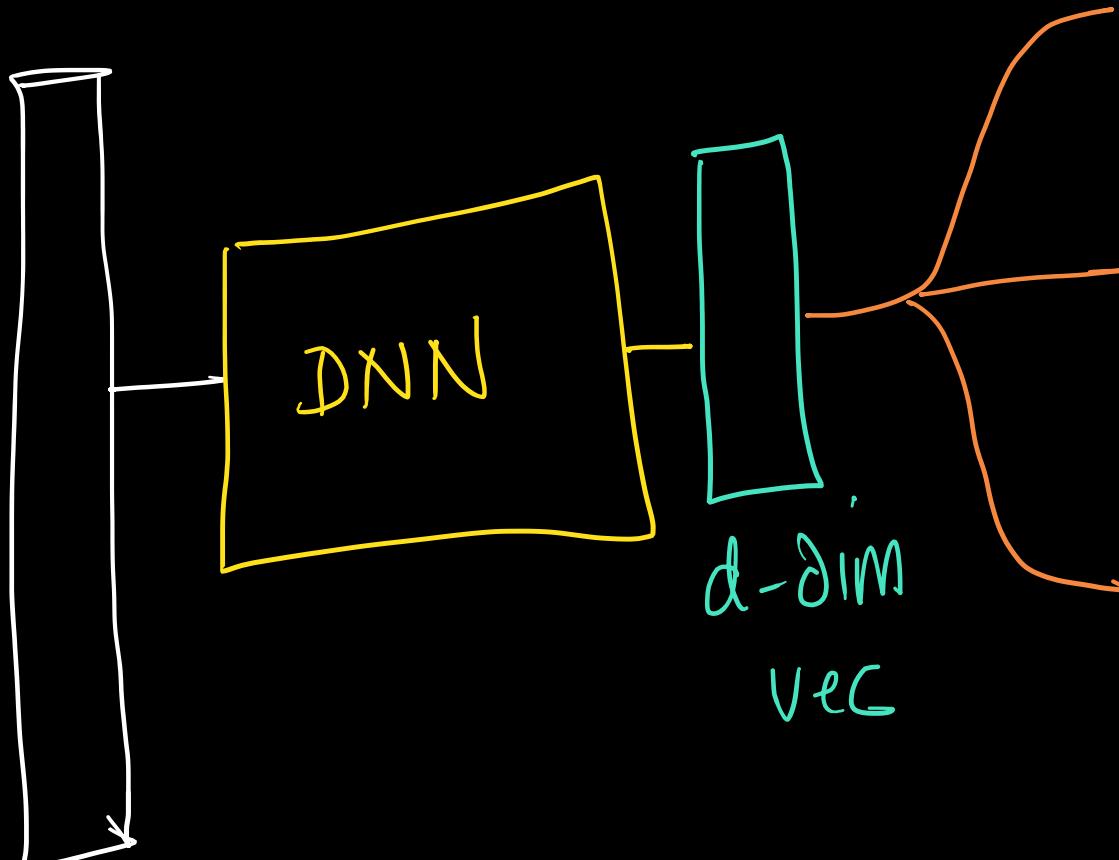
users' features:

↳ categorical

→ numeric

→ purchase - history . . .

u_i



\hat{y}_i^1 : \$ spent in next month

\hat{y}_i^2 : isPrime

\hat{y}_i^3 : #visits next month

multi-task learning

(Q) Where did we obtain d-dim
dense embeddings for users/customers
in our course?

[2 mins]

Matrix factorization based RecSys:

$$R_{n \times m} = U_{n \times d} \cdot M_{d \times m}^T$$

U_i : i^{th} row = d-dim dense vector
of U for U

ratings
 ↓
 $R_{n \times m} = U_{n \times d} \cdot M^T_{d \times m}$
 ↓ ↓
 U_i M_j
 ↗ Co-embedding
 due to interaction

RecSys = regression + entity interaction
(MF)

e.g.: LinkedIn

user - job-posting
entity interaction
+ classification

Key-idea:

2 or more entities interacting

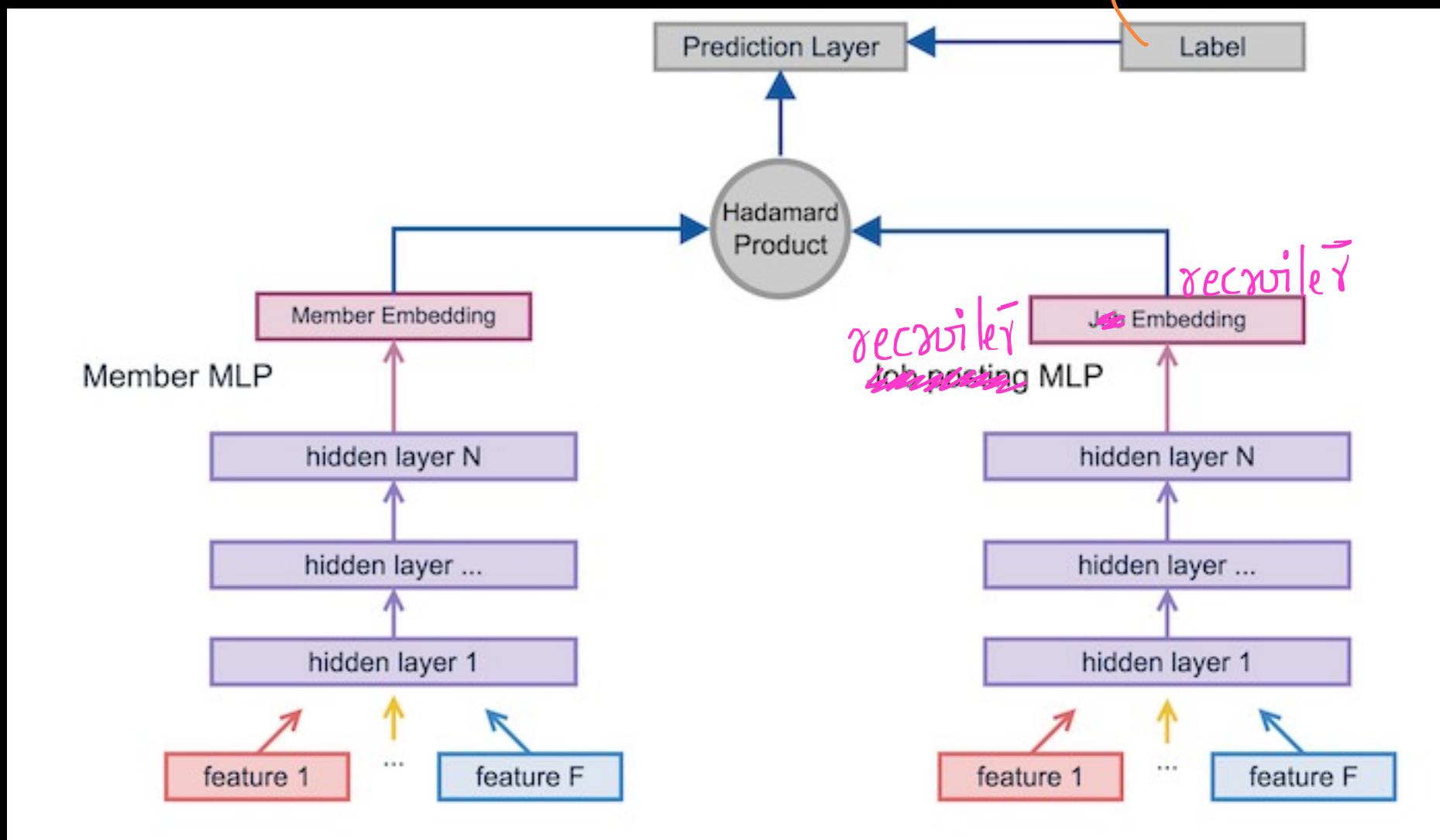
3
+
Co-embedding

+
Recsys (Supervised)

+
Embeddings for all entities

(Q) How can we jointly learn embeddings
of technicians & candidates on LinkedIn?

is interested



(Q) Can you think of 3-entity interactions on LinkedIn?

user/member ; role ; company

Label: isApply?

Sent2Vec = weighted Word2Vec

Seller = weighted Product2Vec
on Amazon

all products sold
by the seller

Key-idea:

④

weighted sum

seq/
Set of entities

+

weighted - sum

↳ another
entity

e.g:

embedding for a job - posting



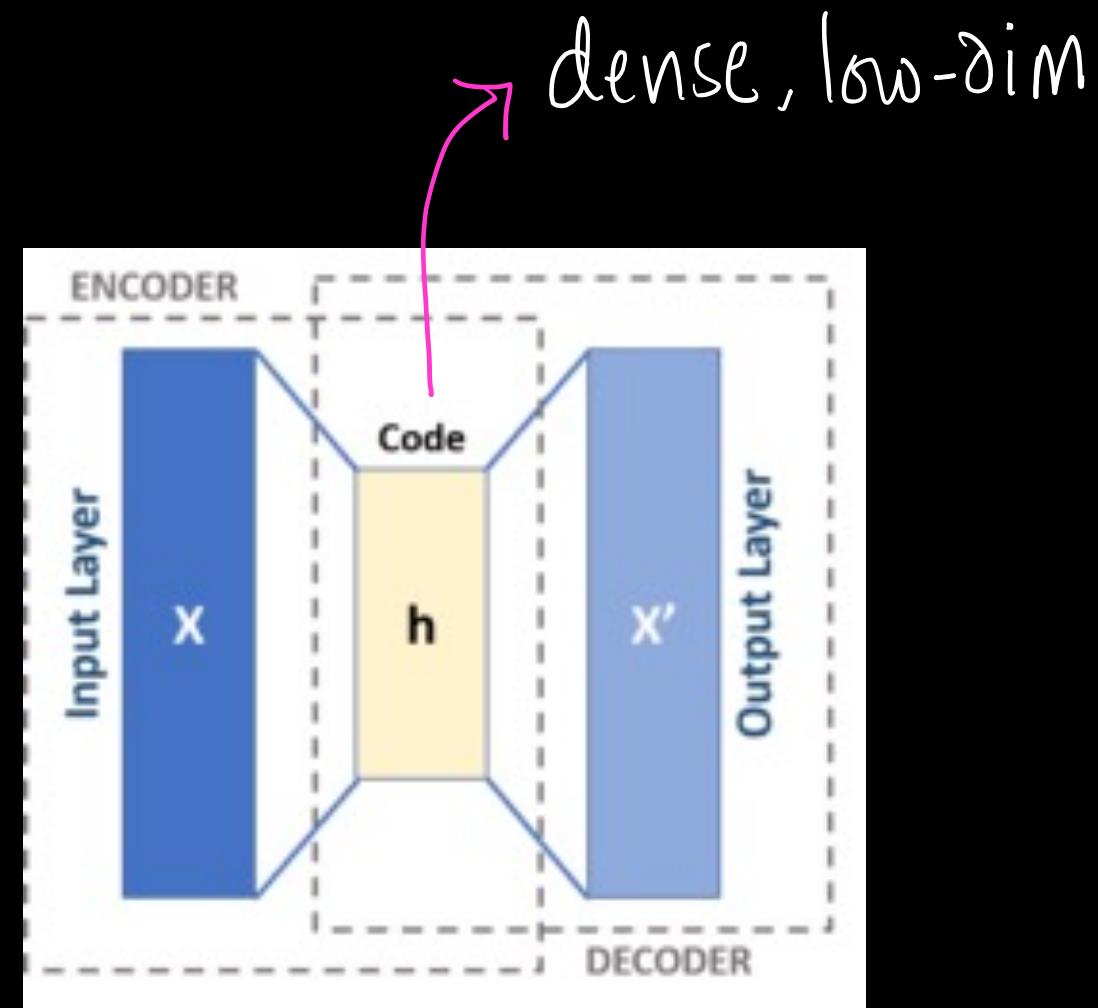
$$\sum_i w_i \text{ cand}_i$$



all those who
applied -

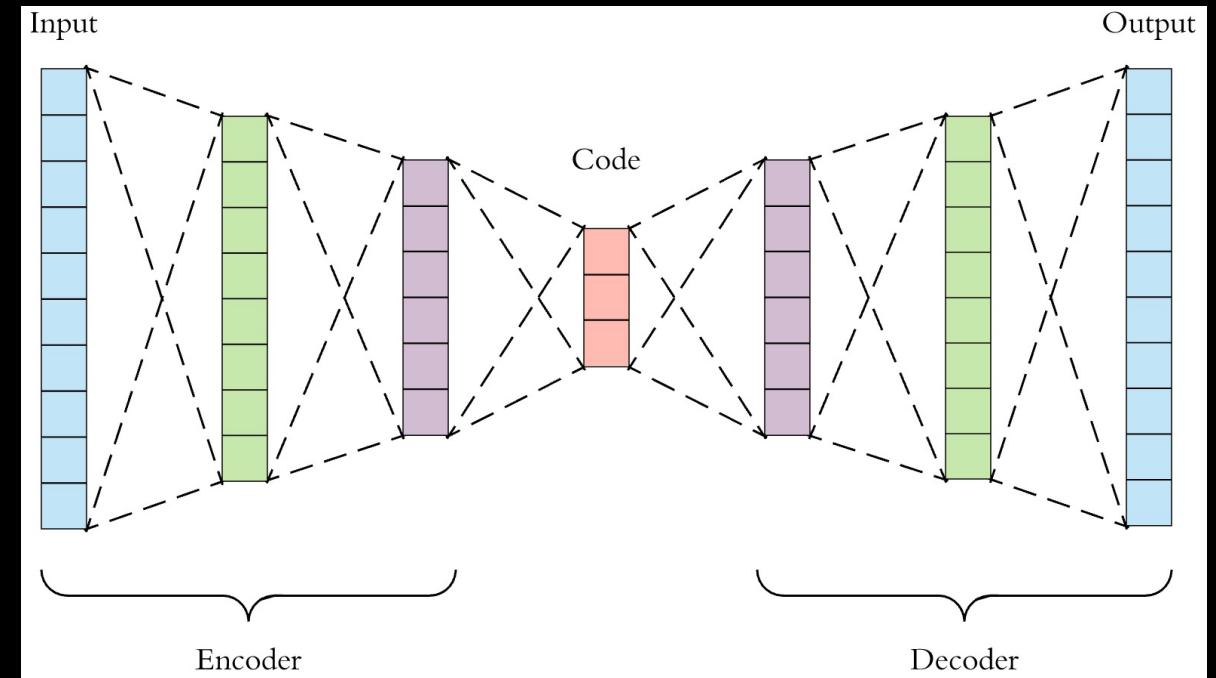
SVD, Autoencoder

↳ based on reconstruction
error.



Code ↴

<https://blog.keras.io/building-autoencoders-in-keras.html>



key-idea

⑤

{}

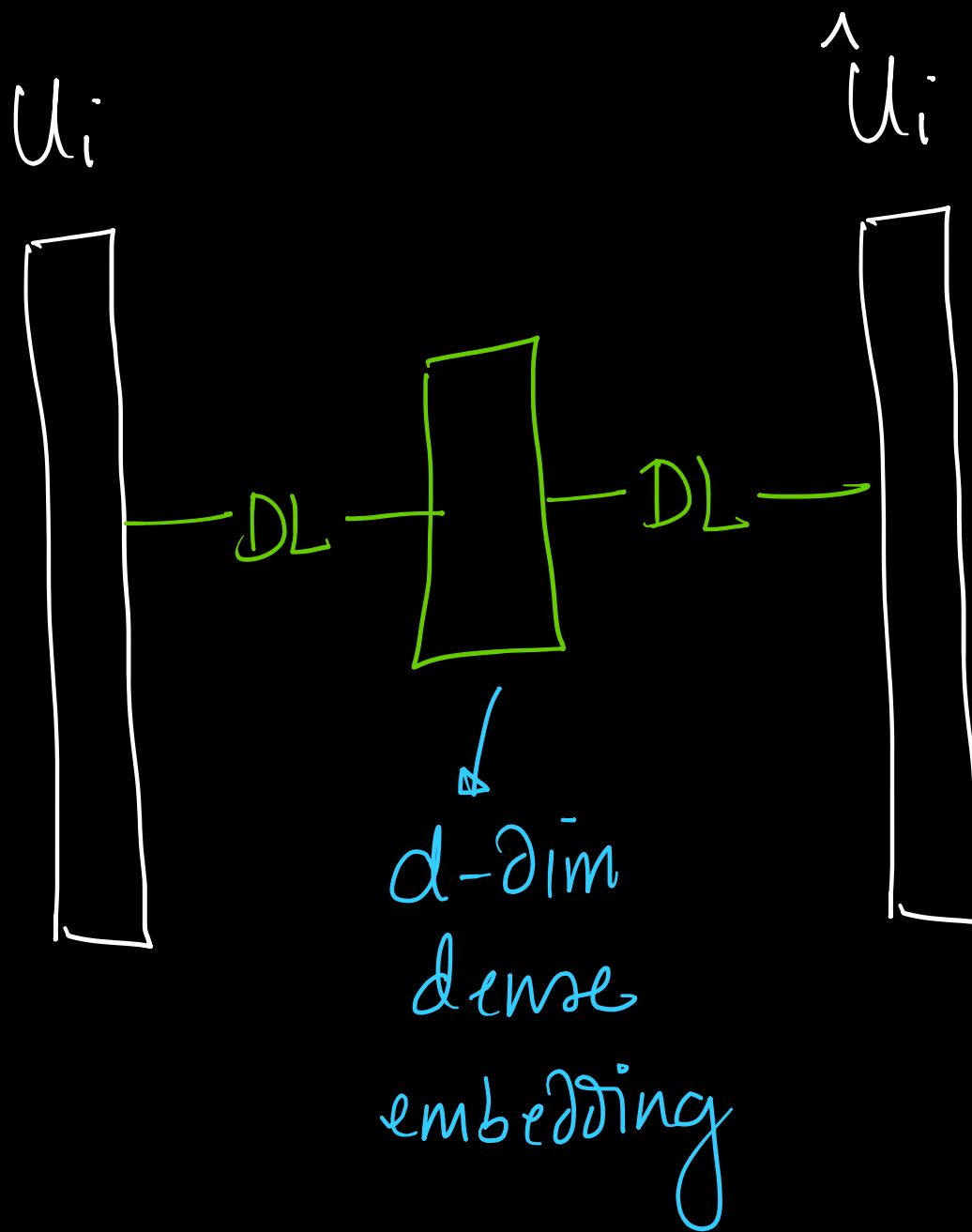
Auto encoders

: any vector representation

+

Auto encoder

(Q) How to use autoencoders to
embed users on Amazon?



Not an exhaustive list of ideas /

strategies

- good enough -

All 5 ideas → based on what-
we already know.

Case-studies from the industry

① LinkedIn Pensieve

↳ embedding feature platform

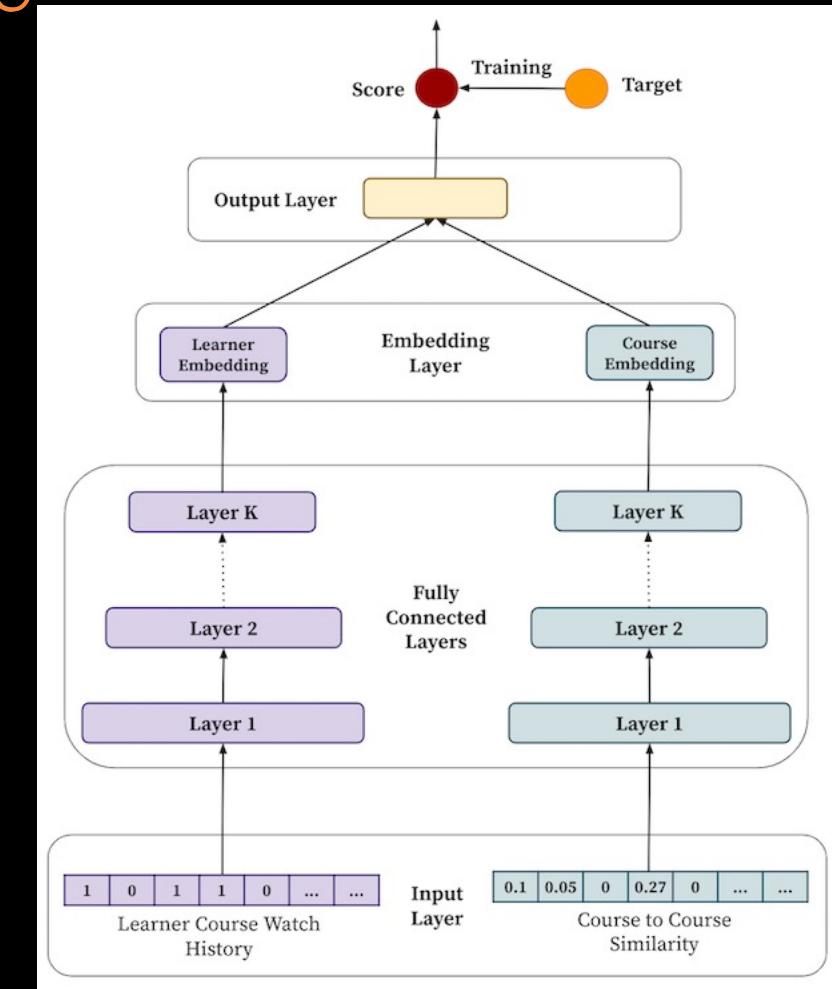
<https://engineering.linkedin.com/blog/2020/pensieve>

→
(dive - deep)
Oct 2020

② LinkedIn embeddings using Neural CF

<https://engineering.linkedin.com/blog/2020/course-recommendations-ai-part-two>

(Jul 2020)



③

Amazon Object2Vec

<https://aws.amazon.com/blogs/machine-learning/introduction-to-amazon-sagemaker-object2vec/>

(dive - Deep)

SageMaker - integration :

<https://docs.aws.amazon.com/sagemaker/latest/dg/object2vec.html>

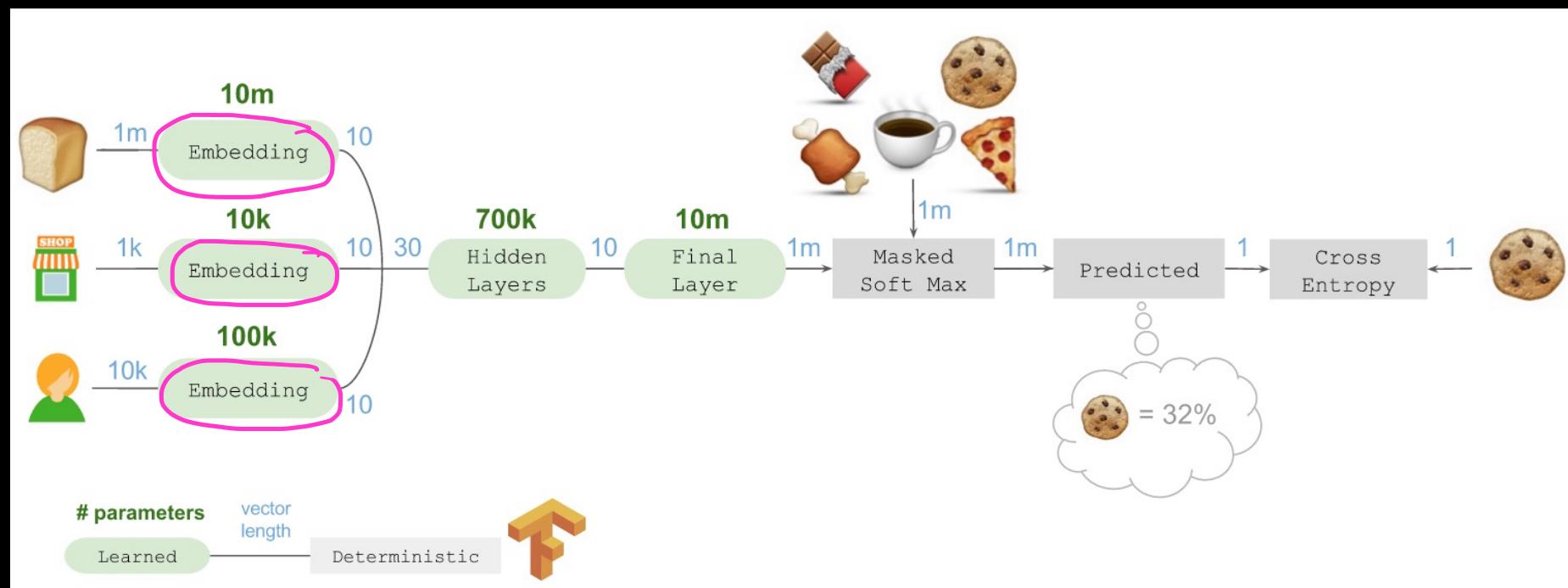
④

Instacart

<https://tech.instacart.com/deep-learning-with-emojis-not-math-660ba1ad6cdc>

+ code }

Product
store
User



⑤ 1mg (Tata)'s MetaProduct2Vec +

<https://medium.com/1mgofficial/moving-beyond-meta-better-product-embeddings-for-better-recommendations-fa6dd1578777>

[Oct-2020]

Q & A