Next Token Prediction

Nipun Batra

IIT Gandhinagar

August 2, 2025

NEXT TOKEN GENERATION

- Inspired by great lecture(s)

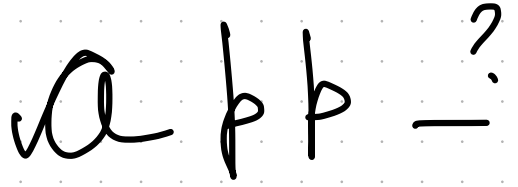
  from Andrej Karpathy.

  ↳ search for

  Neural Networks

  Zero to Hero

- We discussed relevance to chat GPT



What is the next character?

apple ?

What is the next character?

Pose as classification task

apple

c	P(c)
a	0.01
b	...
⋮	
l	0.4
⋮	
z	
_	

Specific Problem

- Generate Indian names
- Dataset : aabid
  aabida
  aadesh
  ⋮
  ⋮
  zeel

Specific Problem

- Generate Indian names
- Dataset : aabid
  aabida
  aadesh
  ⋮
  zeel

Assume

1) Only 26 lower case char
2) _ indicates end char
3) $4 < len < 10$

Generate Training Dataset

WORD #1

aabid

say we consider	'history / content' of 3 chars.

X	y
_ _ _	a
_ _ a	a
_ a a	b
a a b	i
a b i	d
b i d	_

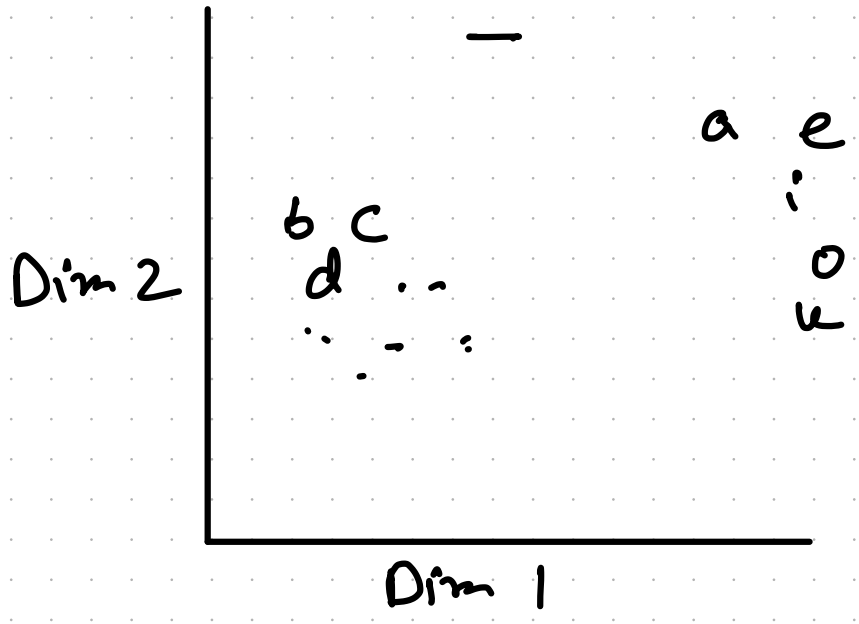
7 training examples from 1 name

Important Idea

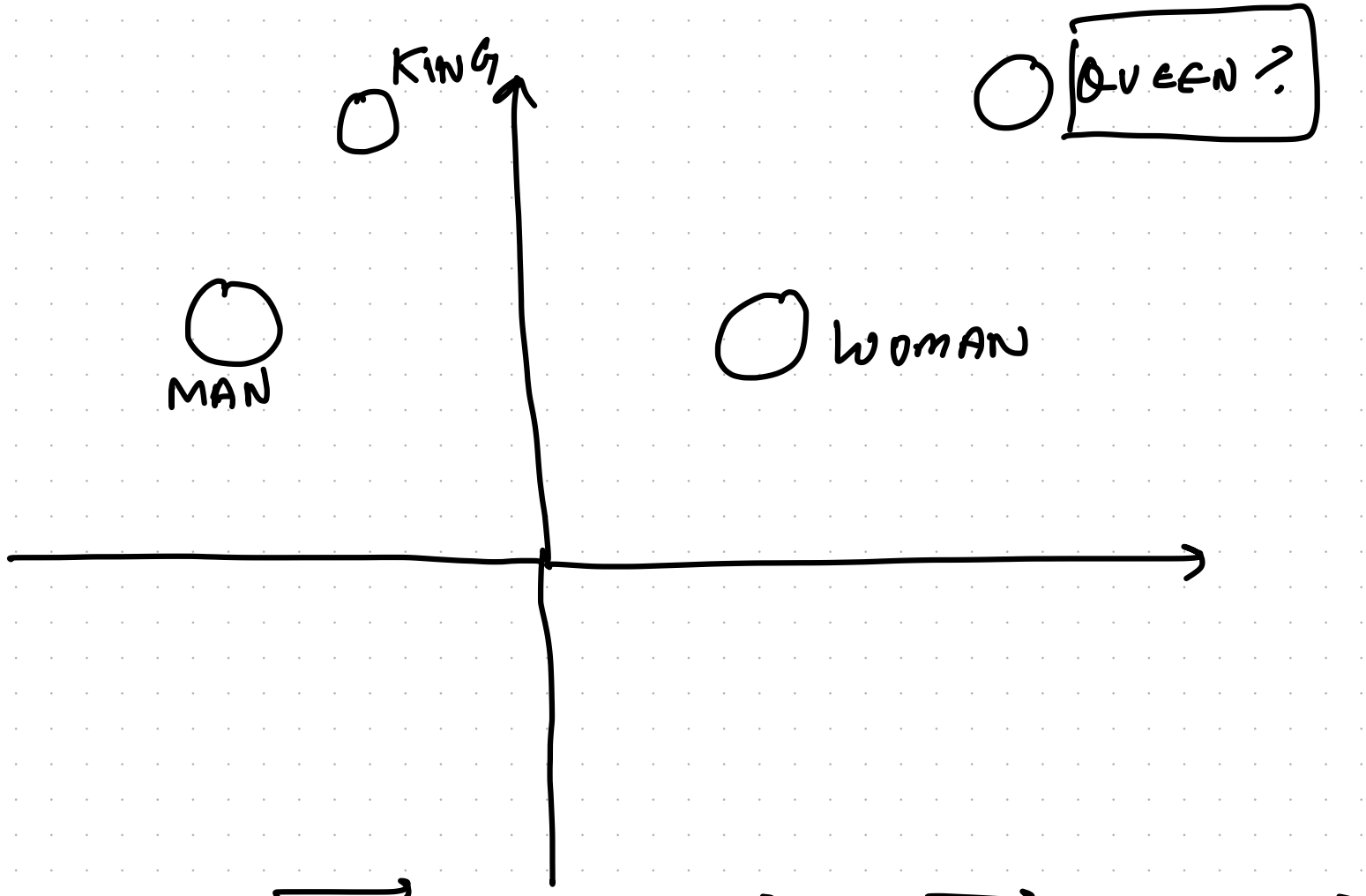
Representation

→ Learn a vector representation for each character

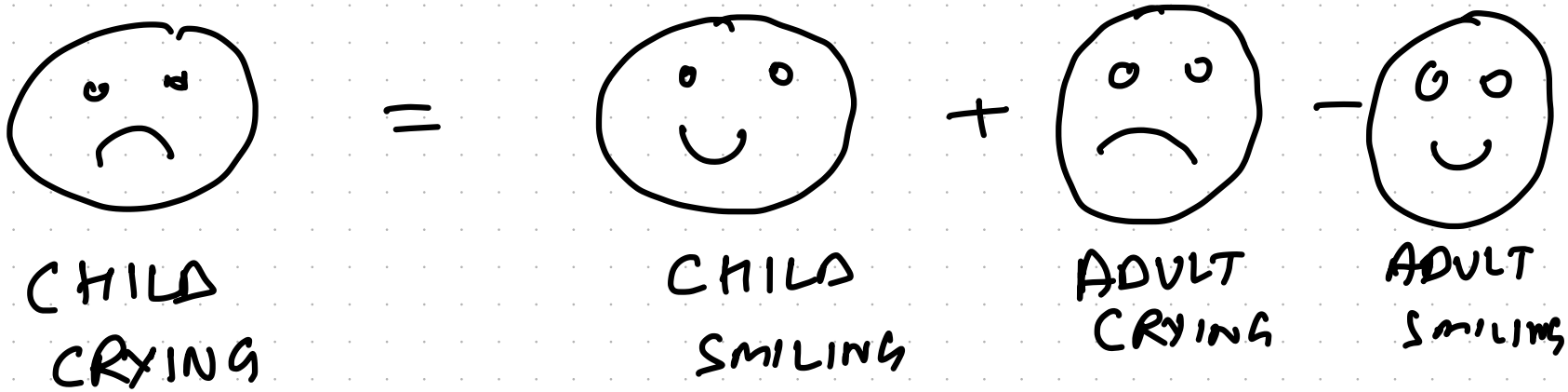
→ 'Similar' characters → closer in vector space



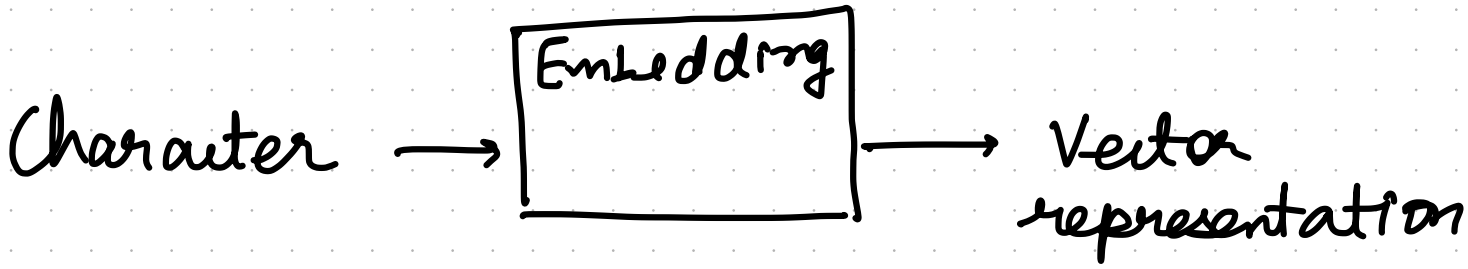
WORD2VEC



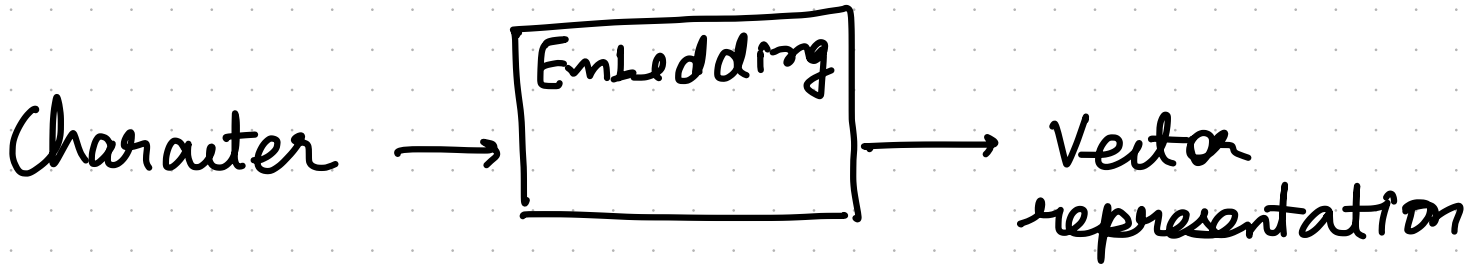
$$\overrightarrow{QUEEN} = \overrightarrow{KING} - \overrightarrow{MAN} + \overrightarrow{WOMAN}$$


CHILD CRYING
=
CHILD SMILING
+
ADULT CRYING
-
ADULT SMILING

Embedding matrix | table



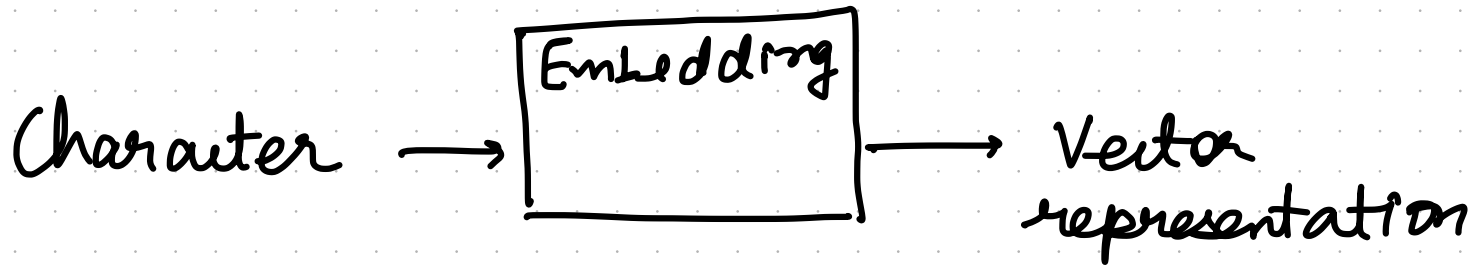
Embedding matrix | table



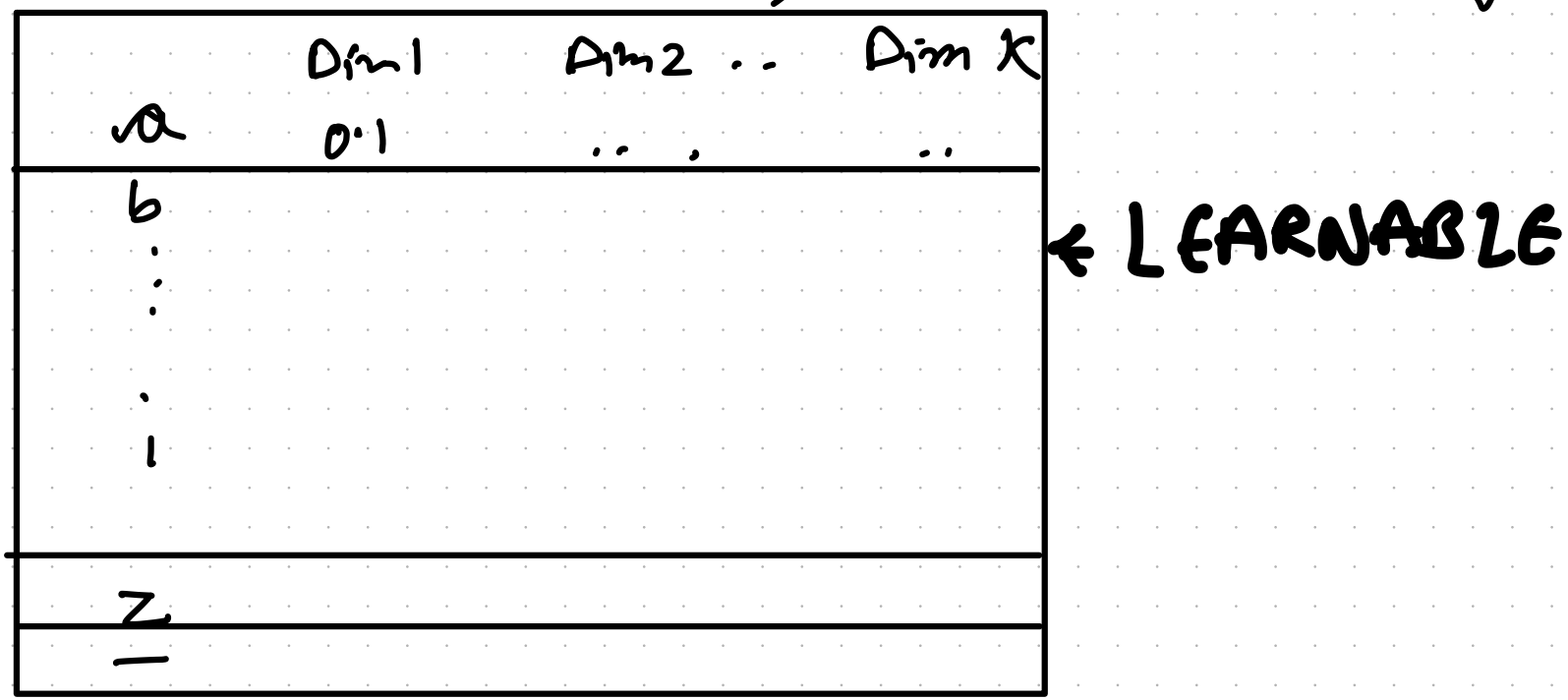
- Given 27 char (a, .. z, -), 'k' dim embedding

	Dim 1	Dim 2 ..	Dim K
a	0.1	.. .	..
b			
:			
.			
i			
z			
-			

Embedding matrix | table



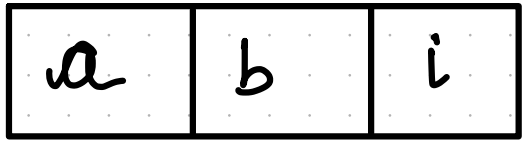
- Given 27 char (a, .. z, -), 'k' dim embedding



OVERALL ARCHITECTURE

- For illustrat², 2dim embedding

X



OVERALL ARCHITECTURE

1) LOOKUP EMBEDDING

X

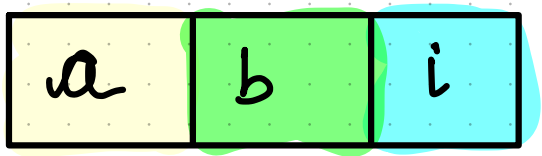
a	b	i

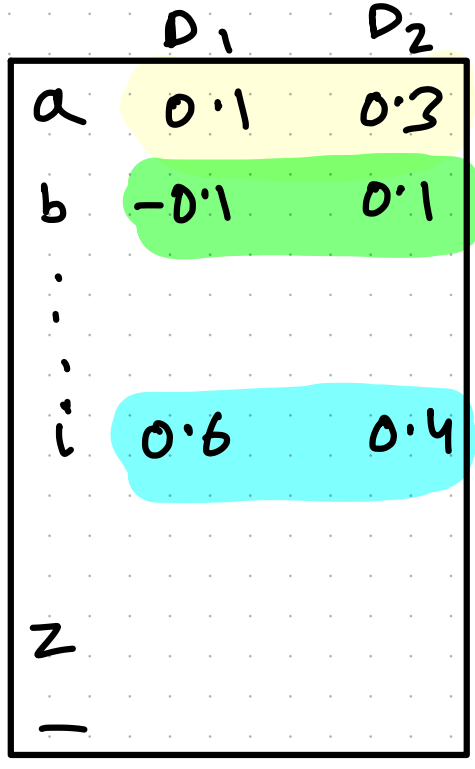
	D_1	D_2
a	0·1	0·3
b	-0·1	0·1
⋮		
i	0·6	0·4
z		
—		

OVERALL ARCHITECTURE

1) LOOKUP EMBEDDING

X

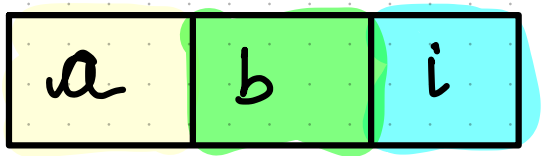


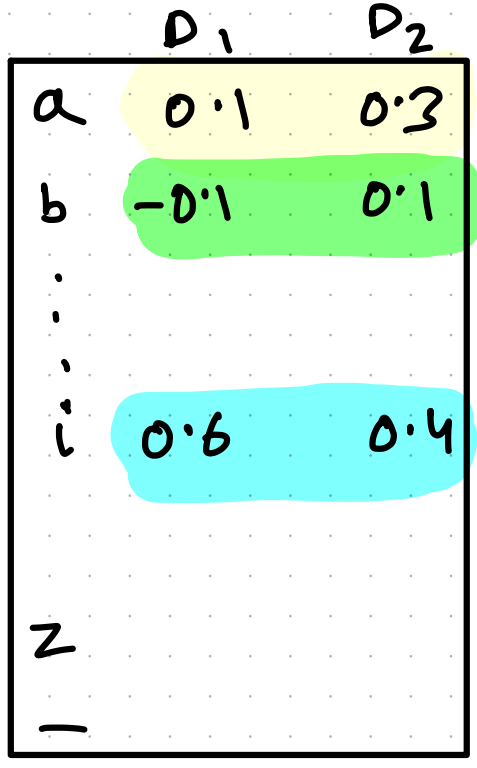


OVERALL ARCHITECTURE

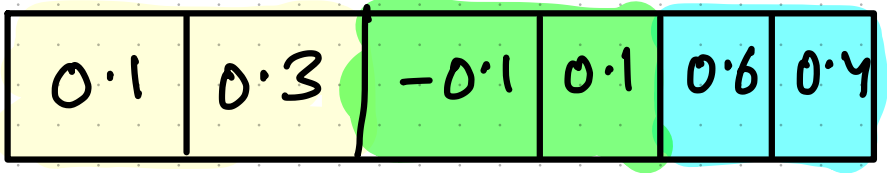
2) CONCATENATE EMBEDDINGS

X



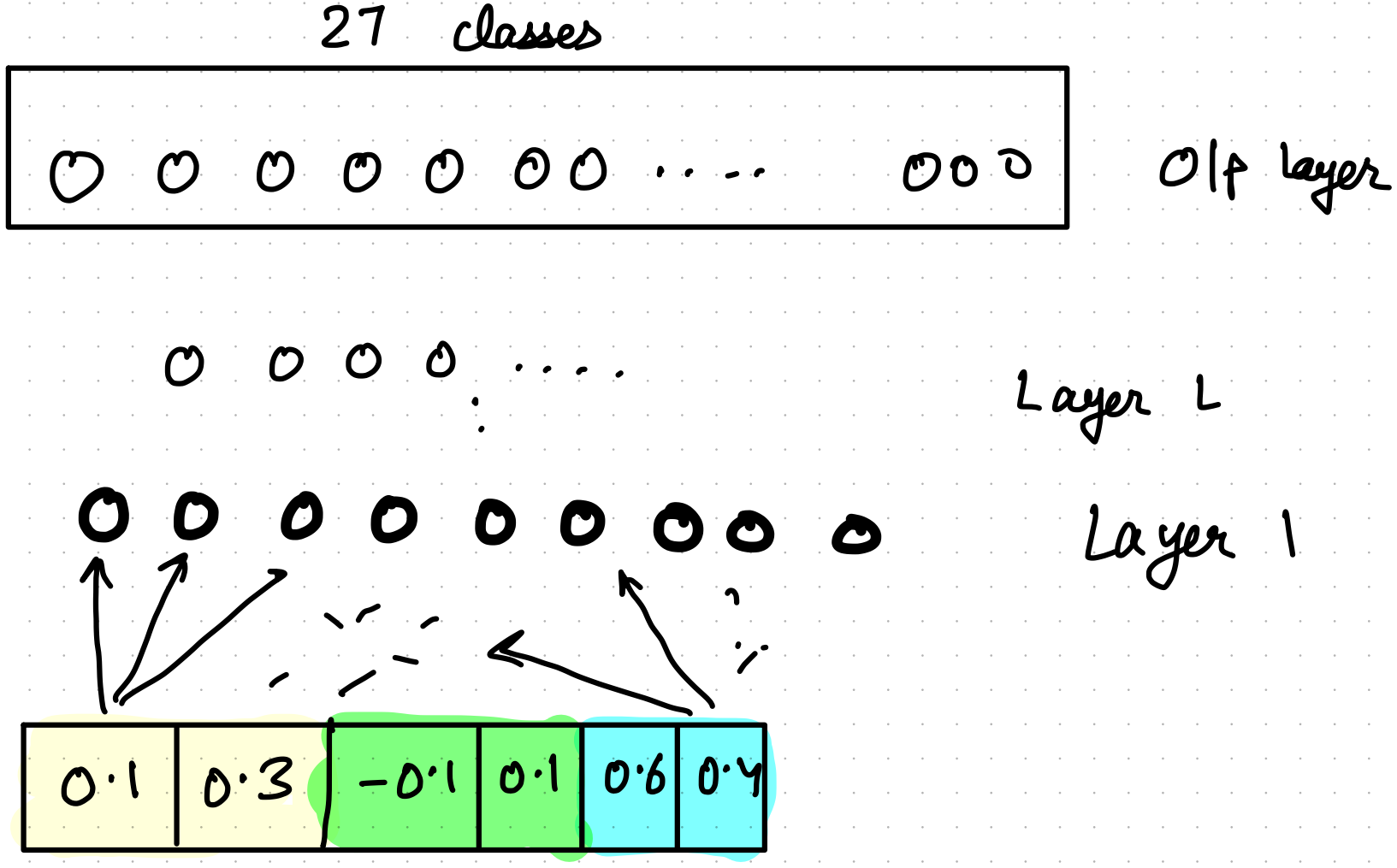


Feature Vector



OVERALL ARCHITECTURE

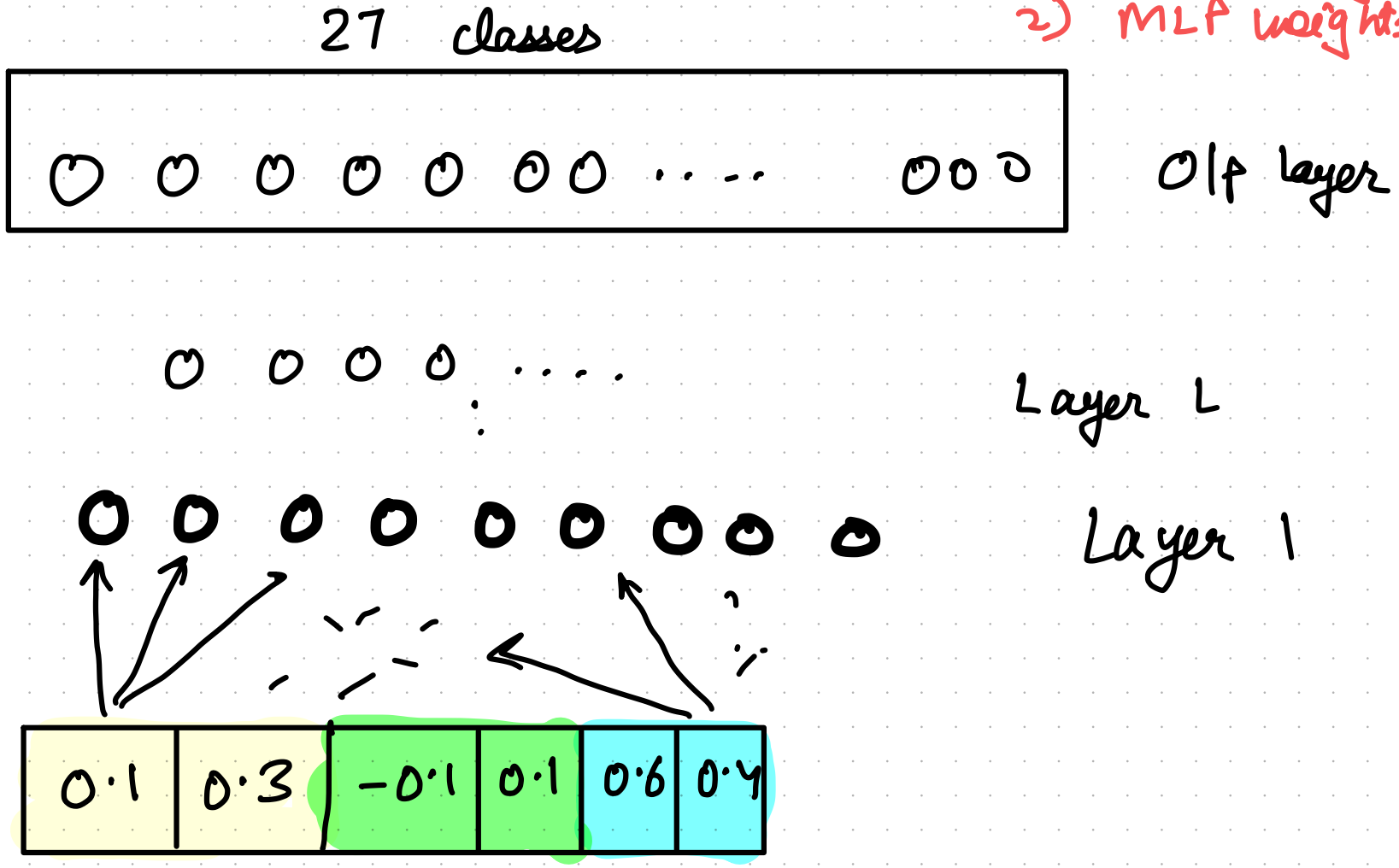
3) MLP



OVERALL ARCHITECTURE

4) USE CROSS ENTROPY LOSS TO LEARN

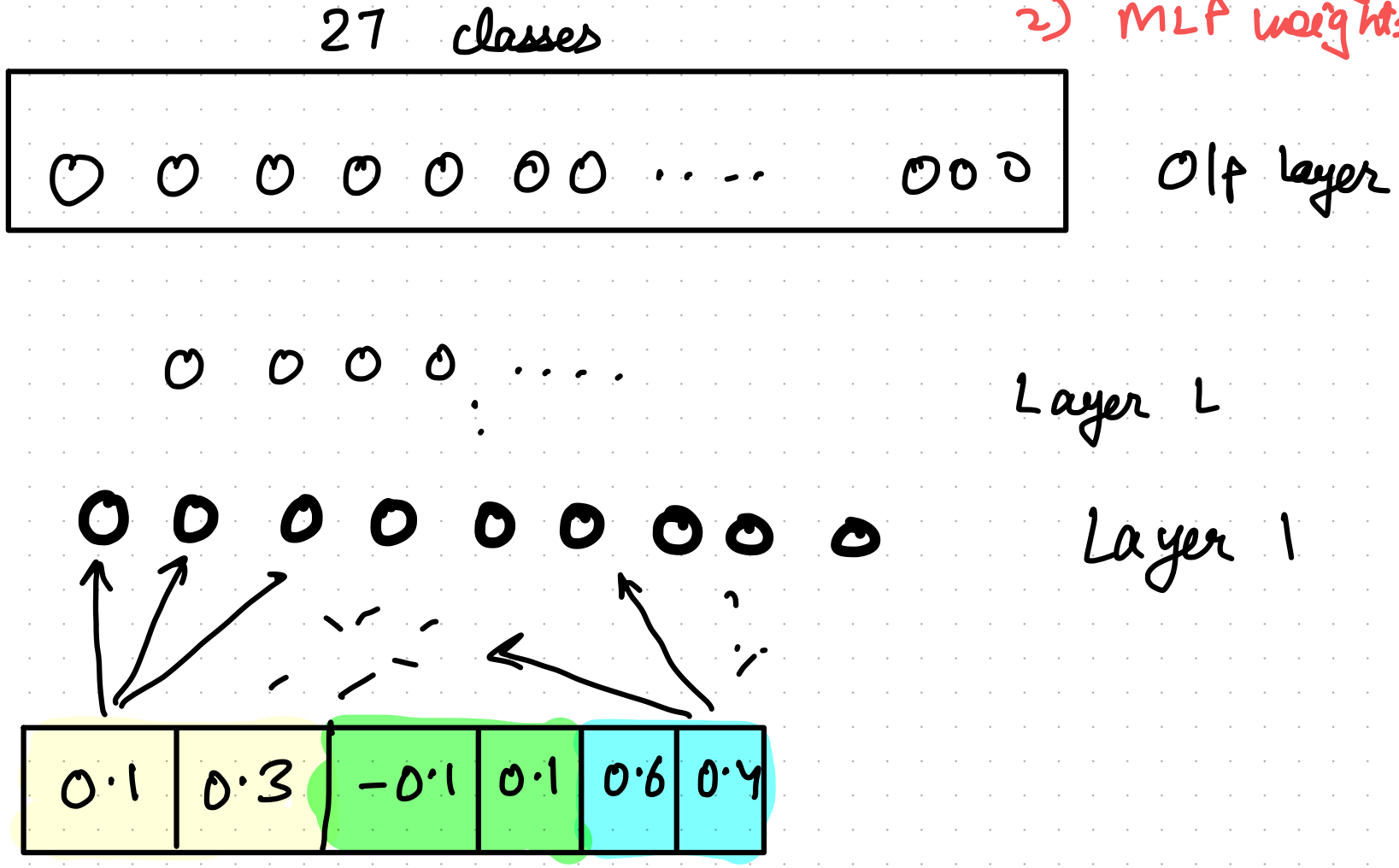
1) Embeddings
2) MLP weights



OVERALL ARCHITECTURE

4) USE CROSS ENTROPY LOSS TO LEARN

1) Embeddings
2) MLP weights



OVERALL ARCHITECTURE

5) GENERATION | SAMPLING

Test i/p

a b i

Test o/p

c	P(c)
a	0.01
b	0.01
c	0.01
d	0.6
e	:
:	:
:	:
:	:
:	:
z	:
_	

OVERALL ARCHITECTURE

5) GENERATION | SAMPLING

Test i/p

a b i

Sample from Prob. distribution

ab i a 1%.

ab i b 1%.

ab i d 60%.

⋮

Test o/p

c	P(c)
a	0.01
b	0.01
c	0.01
d	0.6
e	⋮
⋮	⋮
z	⋮

OVERALL ARCHITECTURE

5) GENERATION | SAMPLING