Chat_Bot

Training Data

"Hi
"How are you?" → greeting

"Bye"
"See you later" → goodbye

X ? **y**

all words

["Hi", "How", "are", "you", "bye", "see", "later"]

Training Data

bag of words

all words

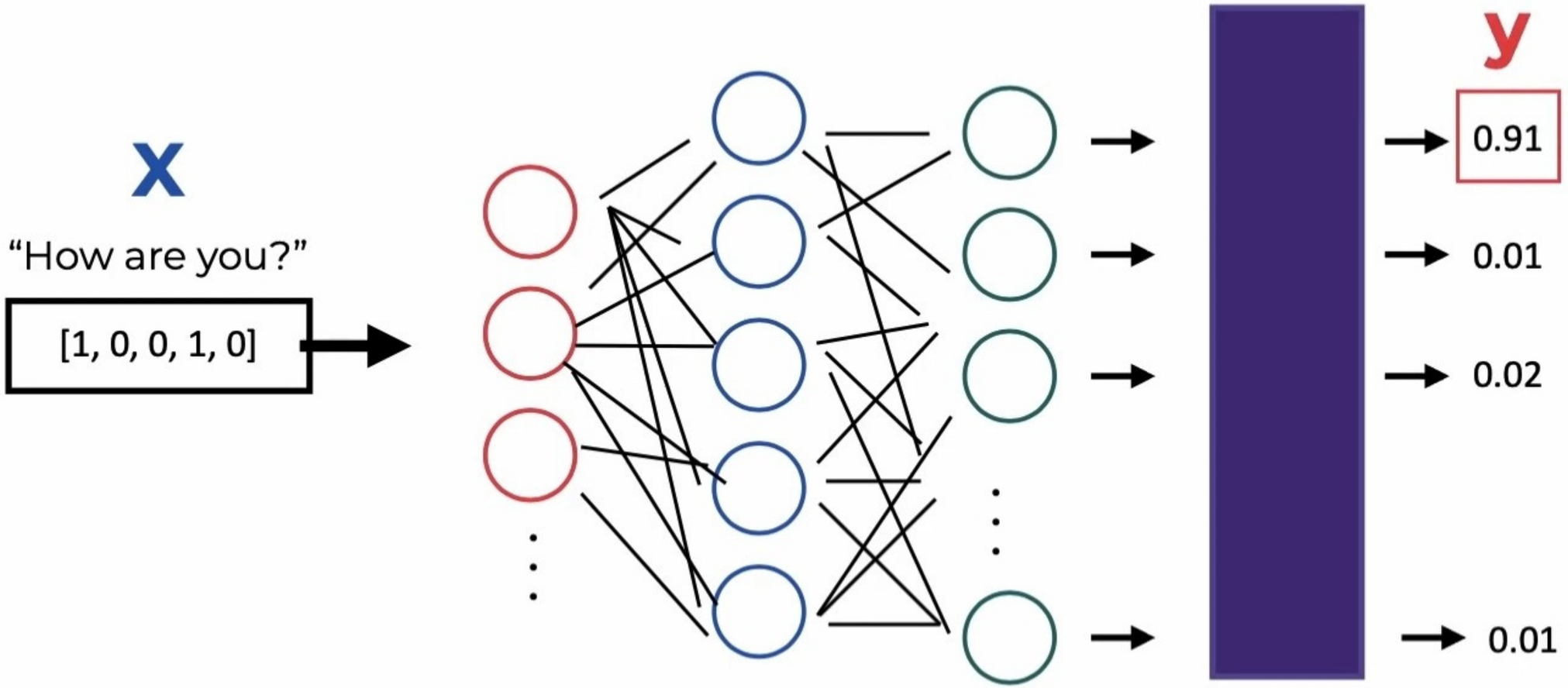
["Hi", "How", "are", "you", "bye", "see", "later"]

"Hi	→	[1, 0, 0, 0, 0, 0 , 0]	0 (greeting)
"How are you?"	→	[0, 1, 1, 1, 0, 0 , 0]	
"Bye"	→	[0, 0, 0, 0, 1, 0 , 0]	1 (goodbye)
"See you later"	→	[0, 0, 0, 1, 0, 1 , 1]	

X

y

Feed Forward Neural Net



NLP Basics

Tokenization: splitting a string into meaningful units
(e.g. words, punctuation characters, numbers)

"what would you do with 1000000$?"

→ [“what”, “would”, “you”, “do”, “with”, “1000000”, “$”, “?”]

“aren’t you happy with so much money?”

→ [“are”, “n’t”, “you”, “happy”, “with”, “so”, “much”, “money”, “?”]

NLP Basics

Stemming: Generate the root form of the words.
Crude heuristic that chops of the ends off of words

"organize", "organizes", "organizing"

→ ["organ", "organ", "organ"]

"universe", "university"

→ ["univers", "univers"]

Our NLP Preprocessing Pipeline

"Is anyone there?"

↓ tokenize

["Is", "anyone", "there", "?"]

↓ lower + stem

["is", "anyon", "there", "?"]

↓ exclude punctuation characters

["is", "anyon", "there"]

↓ bag of words

X [0, 0, 0, 1, 0, 1, 0, 1]

Thank You