

# Assignment 1

## Learning word representations

Nitish Srivastava  
[csc321a@cs.toronto.edu](mailto:csc321a@cs.toronto.edu)

# Word Representations

I-of-K Representation

Binary encoding

'and'

1	0	0	0	0	0	0	0
---	---	---	---	---	---	---	---

0	0	0
---	---	---

'but'

0	1	0	0	0	0	0	0
---	---	---	---	---	---	---	---

0	0	1
---	---	---

'she'

0	0	1	0	0	0	0	0
---	---	---	---	---	---	---	---

0	1	0
---	---	---

'he'

0	0	0	1	0	0	0	0
---	---	---	---	---	---	---	---

0	1	1
---	---	---

← vocabulary size →

←  $\log(\text{vocabulary size})$  →

'john'

0	0	0	0	0	0	0	1
---	---	---	---	---	---	---	---

1	1	1
---	---	---

# Word Representations

I-of-K Representation

Distributed Representation

‘and’

1	0	0	0	0	0	0	0
---	---	---	---	---	---	---	---

1.2	-2.3	0.1	0.2
-----	------	-----	-----

‘but’

0	1	0	0	0	0	0	0
---	---	---	---	---	---	---	---

1.3	-1.7	-0.2	-0.3
-----	------	------	------

‘she’

0	0	1	0	0	0	0	0
---	---	---	---	---	---	---	---

-1.9	2.6	-0.1	-0.3
------	-----	------	------

‘he’

0	0	0	1	0	0	0	0
---	---	---	---	---	---	---	---

-1.9	2.5	-0.2	0.4
------	-----	------	-----

← vocabulary size →

← embedding size →

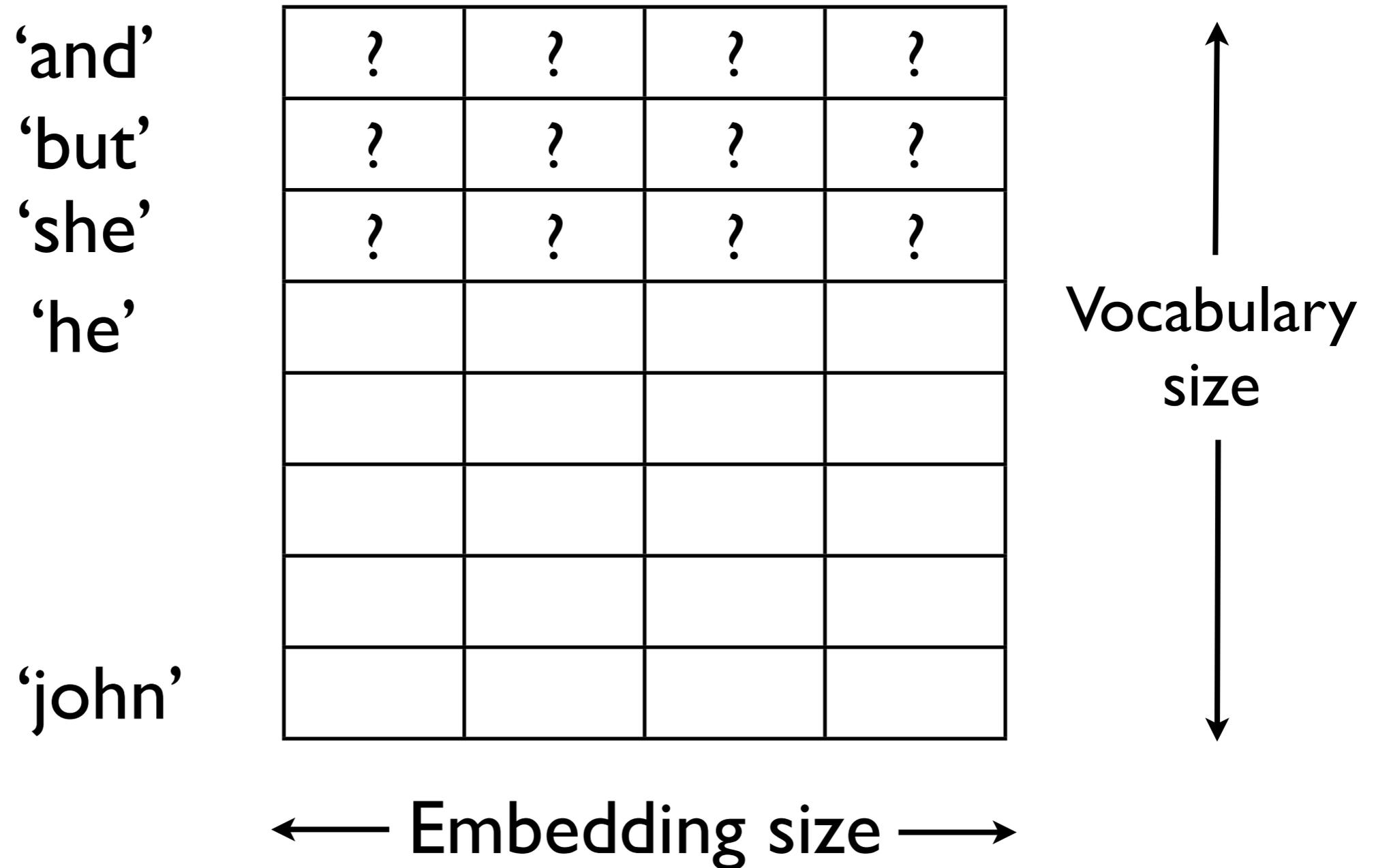
‘john’

0	0	0	0	0	0	0	1
---	---	---	---	---	---	---	---

-1.7	2.5	-0.2	0.4
------	-----	------	-----

# What we need

## Distributed Representation



# What we have

Text Data!

Lots of it!

Newspaper articles, blogs, wikipedia,  
just about any text on any website is useful data!

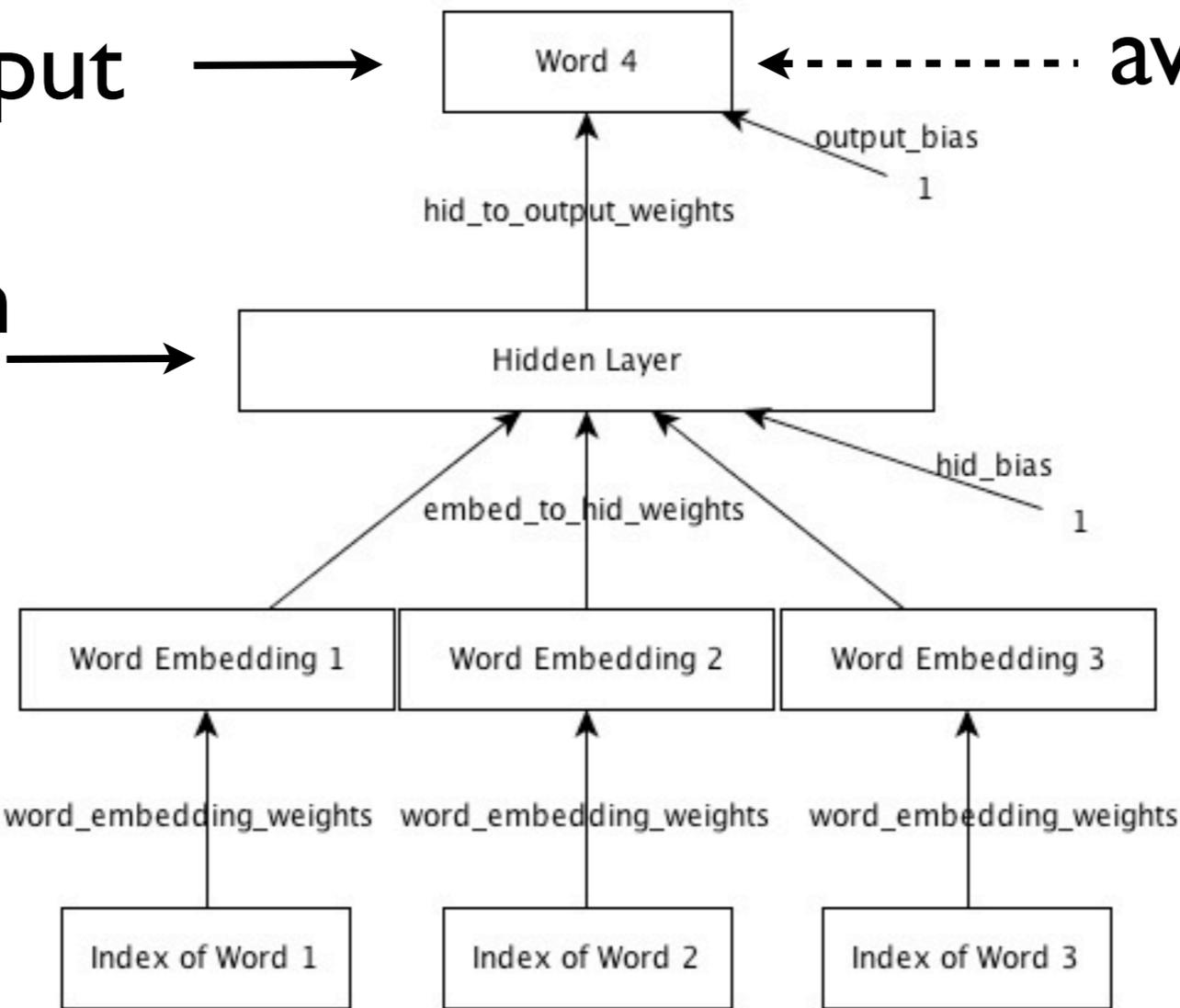
**Design a task for a neural net that  
requires it to learn good word  
representations.**

# Next word prediction

Neural nets are ?

Softmax output → Word 4 ← ..... awesome

Logistic hidden units →



What we want to learn

Neural nets are

# Dataset

- Training set contains 372,500 4-grams.
- Validation and test sets contain 46,500 4-grams.
- Training set - data that the network is trained on.
- Validation set - data used during training to choose hyper-parameters and do early stopping.
- Test set - data used to evaluate the model.

# What you have to do

## Part I [4 points]

- Try different settings of embedding size and hidden layer size.
- Report the training, validation, test errors and time it took to train the network (number of epochs)[3 points].
- Choose the best network. Describe what metric you used to choose it. [1 point]

# What you have to do

## Part 2 [6 points]

- Experiment with the best model using the tools provided and write down your observations.
- Things to try : look at distances between words, nearest neighbours of words, word predictions.
- When do you think the model would put two words close together in embedding space ? Think about what the task is trying to achieve and how that affects the word representation that is being learned.
- Read through the assignment for more suggestions. You are welcome to build upon the code or the tools.



**Due next Tuesday - 5th Feb 2013.**