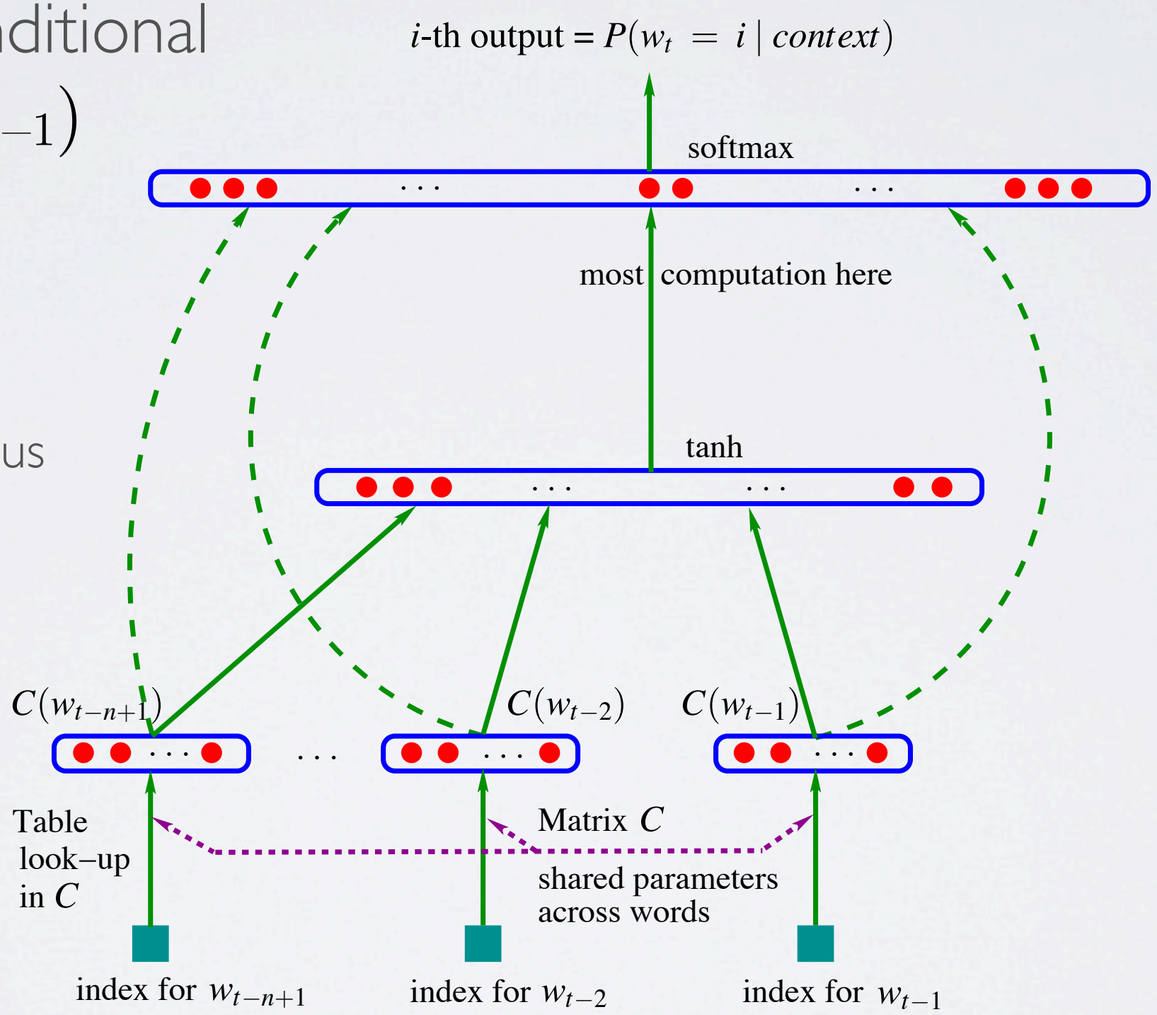
Neural networks

Natural language processing - hierarchical output layer

NEURAL NETWORK LANGUAGE MODEL

Topics: neural network language model

- Solution: model the conditional $p(w_t \mid w_{t-(n-1)}, \ldots, w_{t-1})$ with a neural network
  - learn word representations to allow transfer to n-grams not observed in training corpus

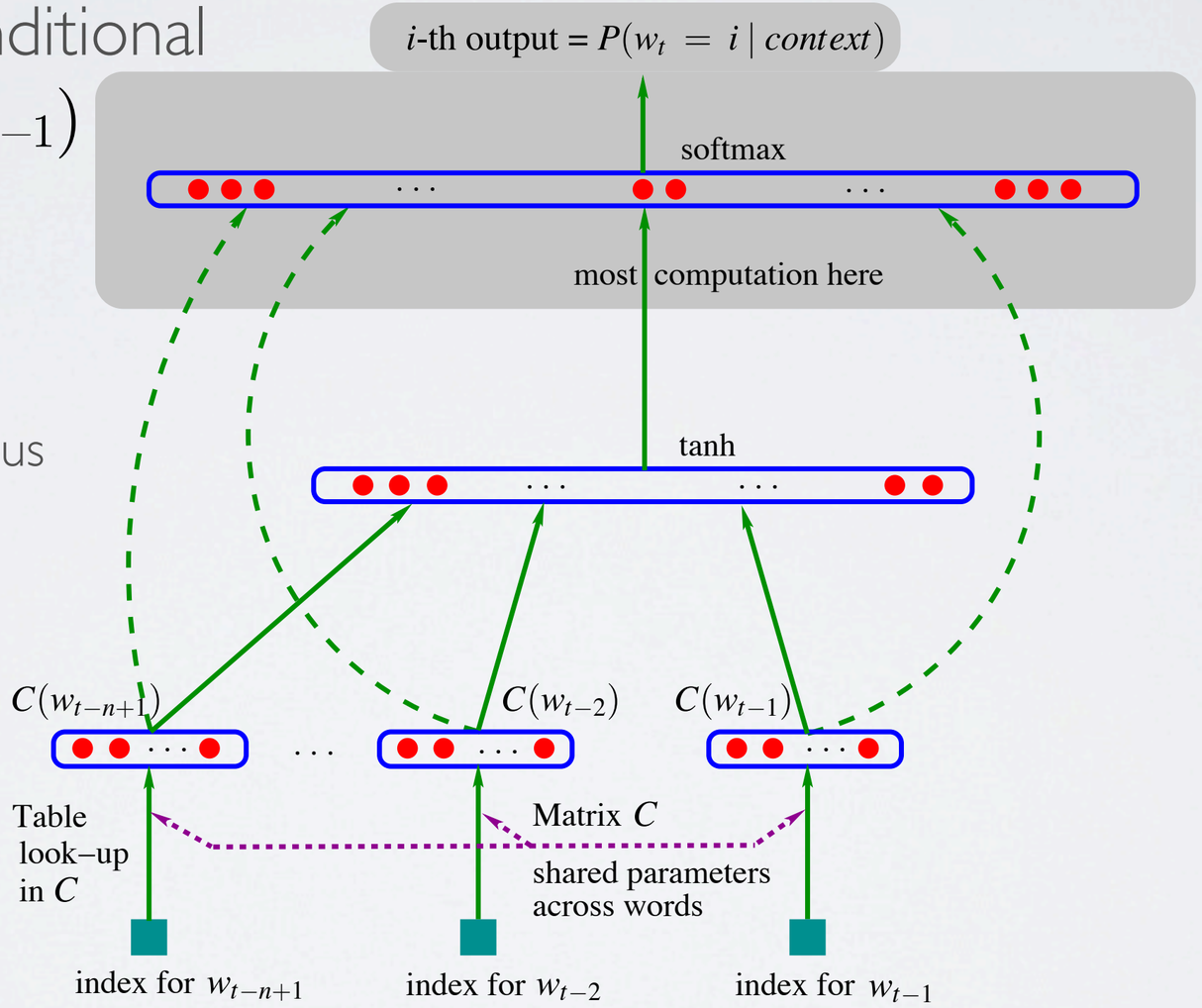


Bengio, Ducharme, Vincent and Jauvin, 2003

NEURAL NETWORK LANGUAGE MODEL

Topics: neural network language model

- Solution: model the conditional $p(w_t \mid w_{t-(n-1)}, \ldots, w_{t-1})$ with a neural network
  - learn word representations to allow transfer to n-grams not observed in training corpus



Bengio, Ducharme, Vincent and Jauvin, 2003

NEURAL NETWORK LANGUAGE MODEL

Topics: hierarchical output layer

- Example: [“ the ”,“ dog ”,“ and ”, “ the ”, “ cat ”]

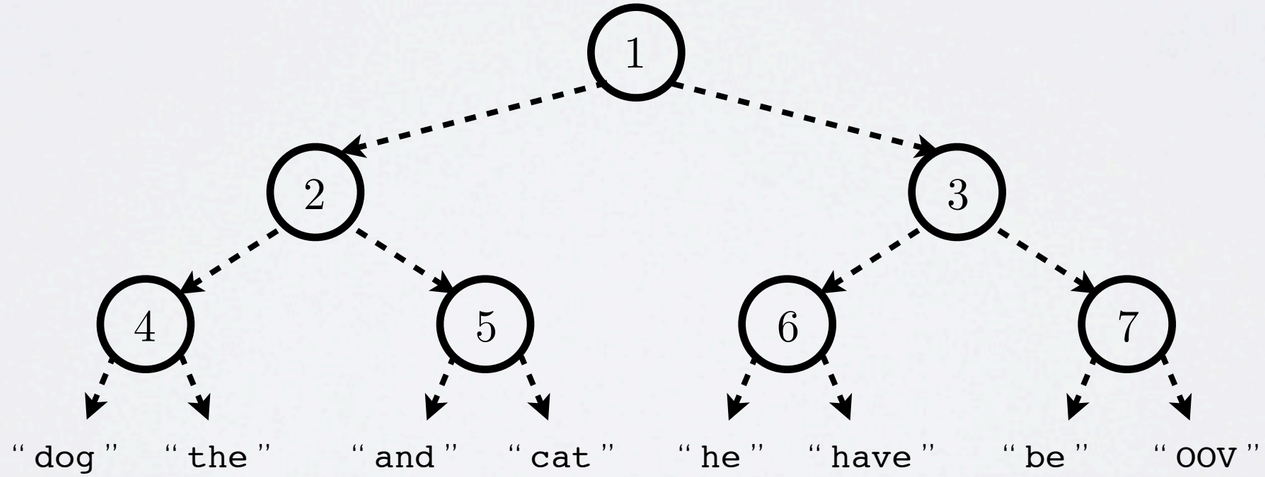
“ the ”

“ dog ”

“ and ”

“ the ”

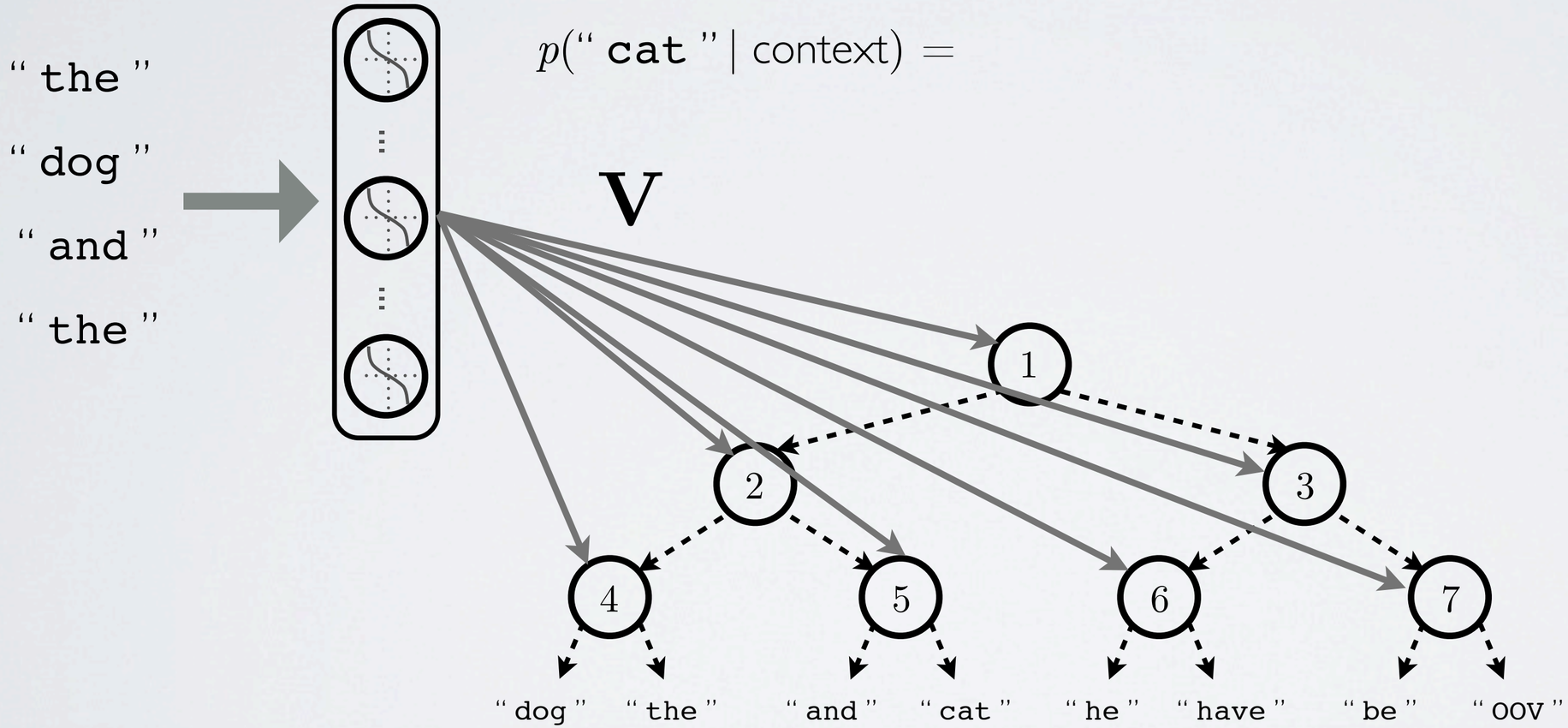
$p($“ cat ” | context$) =$



NEURAL NETWORK LANGUAGE MODEL

Topics: hierarchical output layer

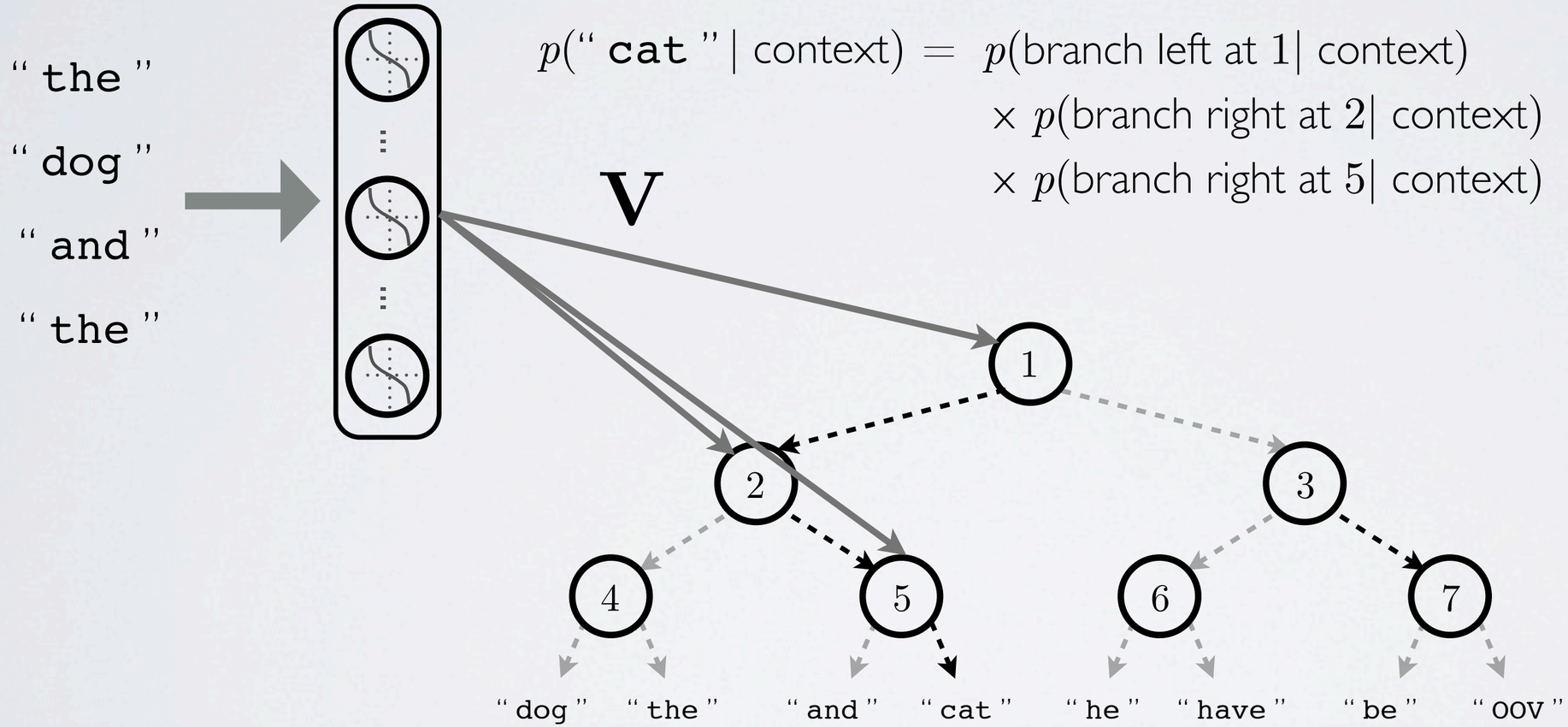
- Example: [“ the ”,“ dog ”,“ and ”, “ the ”, “ cat ”]



NEURAL NETWORK LANGUAGE MODEL

Topics: hierarchical output layer

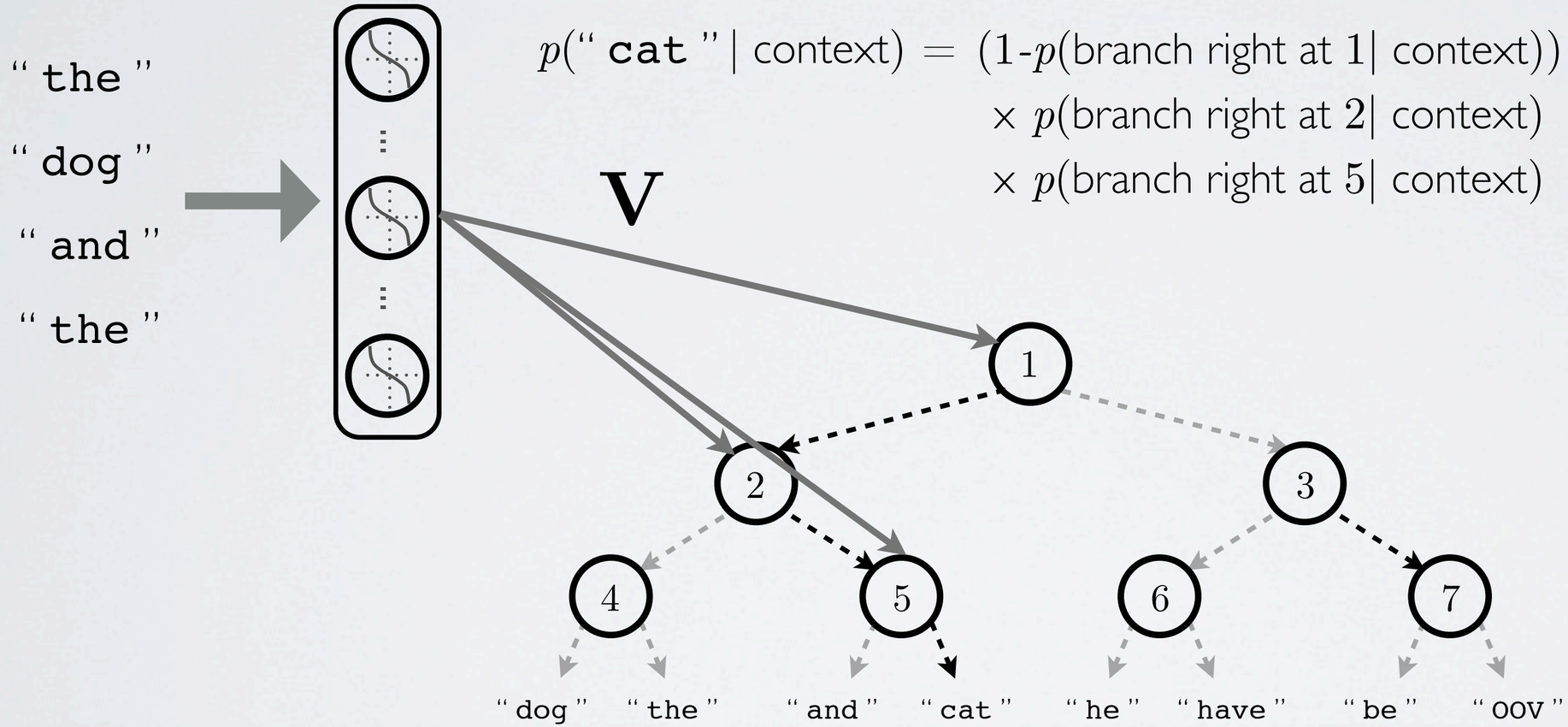
- Example: [“ the ”,“ dog ”,“ and ”, “ the ”, “ cat ”]



NEURAL NETWORK LANGUAGE MODEL

Topics: hierarchical output layer

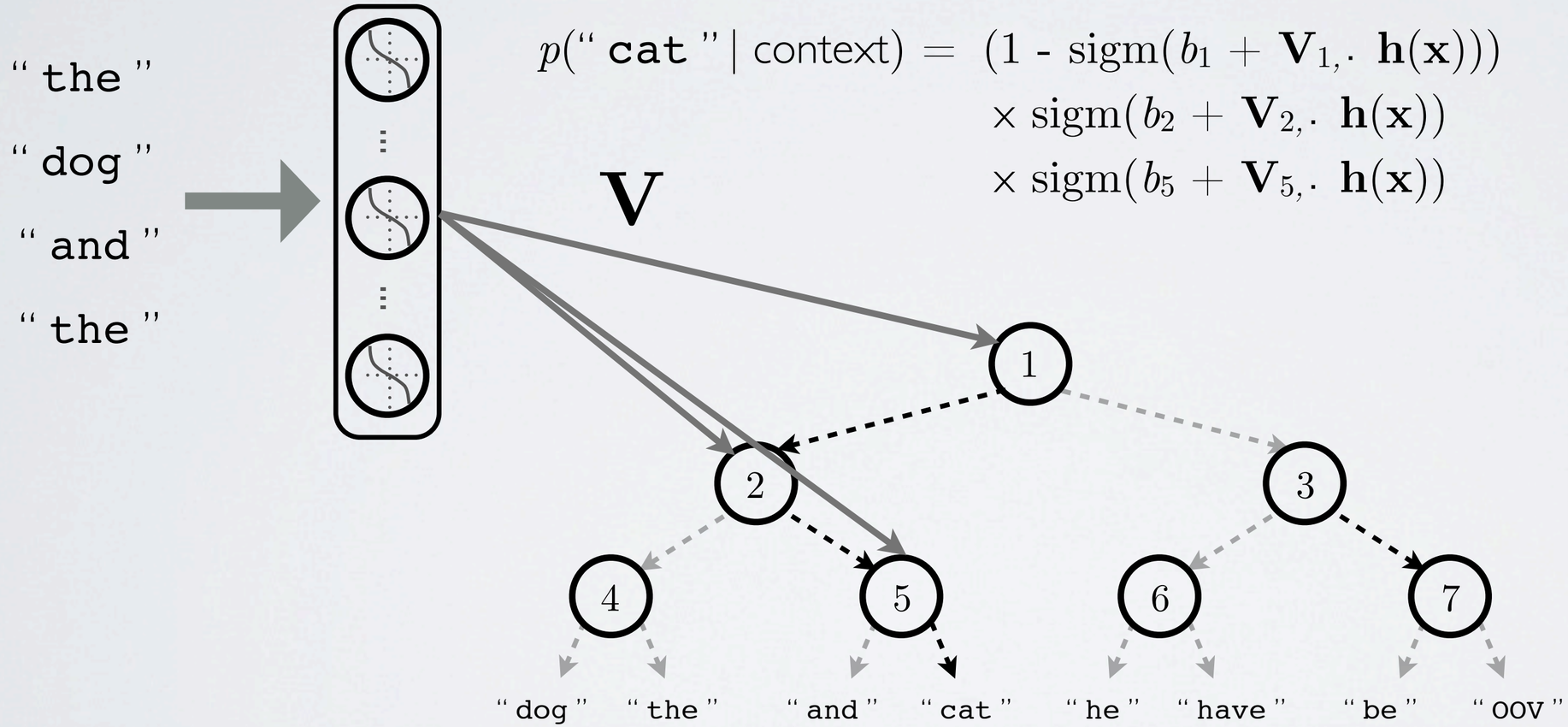
- Example: [“ the ”,“ dog ”,“ and ”, “ the ”, “ cat ”]



NEURAL NETWORK LANGUAGE MODEL

Topics: hierarchical output layer

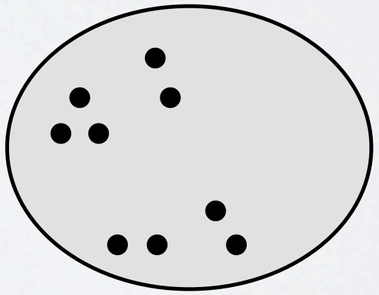
- Example: [“ the ”,“ dog ”,“ and ”, “ the ”, “ cat ”]



NEURAL NETWORK LANGUAGE MODEL

Topics: hierarchical output layer

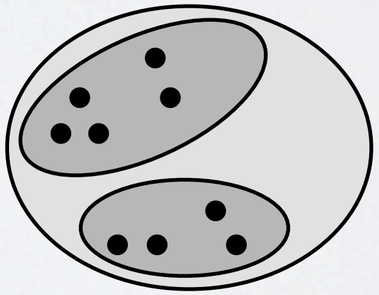
- How to define the word hierarchy
  - can use a randomly generated tree
    - this is likely to be suboptimal
  - can use existing linguistic resources, such as WordNet
    - Hierarchical Probabilistic Neural Network Language Model
      Morin and Bengio, 2005
    - they report a speedup of 258x, with a slight decrease in performance
  - can learn the hierarchy using a recursive partitioning strategy
    - A Scalable Hierarchical Distributed Language Model
      Mnih and Hinton, 2008
    - similar speedup factors are reported, without a performance decrease



NEURAL NETWORK LANGUAGE MODEL

Topics: hierarchical output layer

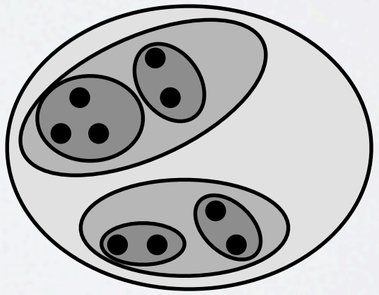
- How to define the word hierarchy
  - can use a randomly generated tree
    - this is likely to be suboptimal
  - can use existing linguistic resources, such as WordNet
    - Hierarchical Probabilistic Neural Network Language Model
      Morin and Bengio, 2005
    - they report a speedup of 258x, with a slight decrease in performance
  - can learn the hierarchy using a recursive partitioning strategy
    - A Scalable Hierarchical Distributed Language Model
      Mnih and Hinton, 2008
    - similar speedup factors are reported, without a performance decrease



NEURAL NETWORK LANGUAGE MODEL

Topics: hierarchical output layer

- How to define the word hierarchy
  - can use a randomly generated tree
    - this is likely to be suboptimal
  - can use existing linguistic resources, such as WordNet
    - Hierarchical Probabilistic Neural Network Language Model
      Morin and Bengio, 2005
    - they report a speedup of 258x, with a slight decrease in performance
  - can learn the hierarchy using a recursive partitioning strategy
    - A Scalable Hierarchical Distributed Language Model
      Mnih and Hinton, 2008
    - similar speedup factors are reported, without a performance decrease



NEURAL NETWORK LANGUAGE MODEL

Topics: hierarchical output layer

- How to define the word hierarchy
  - can use a randomly generated tree
    - this is likely to be suboptimal
  - can use existing linguistic resources, such as WordNet
    - Hierarchical Probabilistic Neural Network Language Model
      Morin and Bengio, 2005
    - they report a speedup of 258x, with a slight decrease in performance
  - can learn the hierarchy using a recursive partitioning strategy
    - A Scalable Hierarchical Distributed Language Model
      Mnih and Hinton, 2008
    - similar speedup factors are reported, without a performance decrease