

Neural Language Models

Natalie Parde

UIC CS 421

Neural Language Models

- Popular application of neural networks
- Advantages over n -gram language models:
 - Can handle longer histories
 - Can generalize over contexts of similar words
- Disadvantage:
 - Slower to train
- Neural language models have higher predictive accuracy than n -gram language models trained on datasets of similar sizes

Neural Language Models

- Neural language models are used to boost performance for many modern NLP tasks
 - Machine translation
 - Dialogue systems
 - Language generation



Sample Generated by a Neural Language Model (GPT-2)

- Link to article: <https://openai.com/blog/better-language-models/>

System Prompt (Human-Written): *In a shocking finding, scientist discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.*

Sample Generated by a Neural Language Model (GPT-2)

Model Completion (Machine-Written, 10 Tries): The scientist named the population, after their distinctive horn, Ovid's Unicorn. These four-horned, silver-white unicorns were previously unknown to science.

Now, after almost two centuries, the mystery of what sparked this odd phenomenon is finally solved.

Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and several companions, were exploring the Andes Mountains when they found a small valley, with no other animals or humans. Pérez noticed that the valley had what appeared to be a natural fountain, surrounded by two peaks of rock and silver snow.

Pérez and the others then ventured further into the valley. "By the time we reached the top of one peak, the water looked blue, with some crystals on top," said Pérez.

Pérez and his friends were astonished to see the unicorn herd. These creatures could be seen from the air without having to move too much to see them – they were so close they could touch their horns.

While examining these bizarre creatures the scientists discovered that the creatures also spoke some fairly regular English. Pérez stated, "We can see, for example, that they have a common 'language,' something like a dialect or dialectic."

Dr. Pérez believes that the unicorns may have originated in Argentina, where the animals were believed to be descendants of a lost race of people who lived there before the arrival of humans in those parts of South America.

While their origins are still unclear, some believe that perhaps the creatures were created when a human and a unicorn met each other in a time before human civilization. According to Pérez, "In South America, such incidents seem to be quite common."

However, Pérez also pointed out that it is likely that the only way of knowing for sure if unicorns are indeed the descendants of a lost alien race is through DNA. "But they seem to be able to communicate in English quite well, which I believe is a sign of evolution, or at least a change in social organization," said the scientist.

Feedforward Neural Language Model

- Input: Representation of some number of previous words
 - w_{t-1}, w_{t-2} , etc.
- Output: Probability distribution over possible next words
- Goal: Approximate the probability of a word given the entire prior context $P(w_t | w_1^{t-1})$ based on the n previous words
 - $P(w_t | w_1^{t-1}) \approx P(w_t | w_{t-n+1}^{t-1})$

**Neural
language
models
represent
prior context
using
embeddings
of the
previous
words.**

- Allows them to generalize to unseen data better than n -gram models
- Embeddings can come from various sources
 - E.g., pretrained Word2Vec embeddings

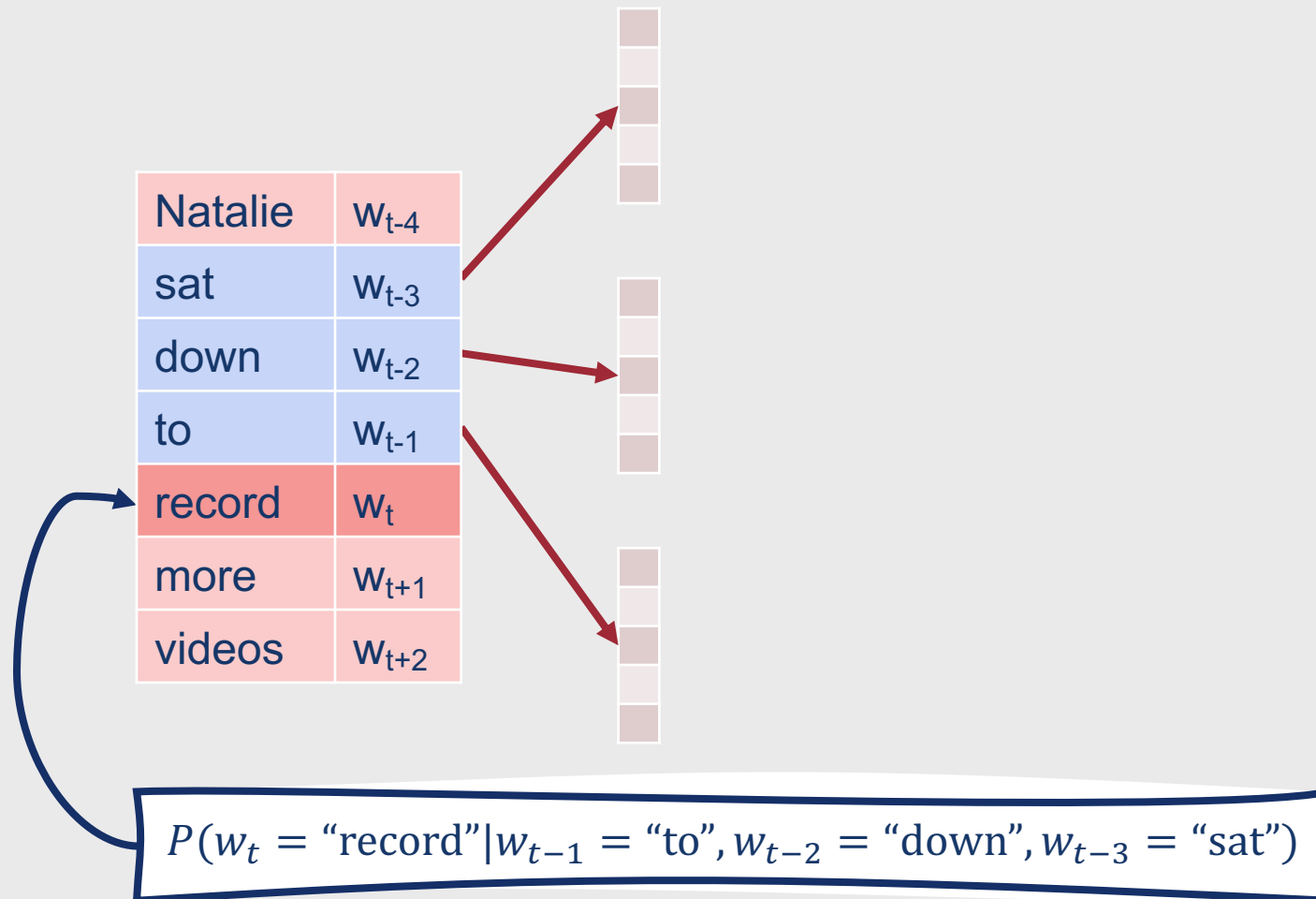


Neural Language Model

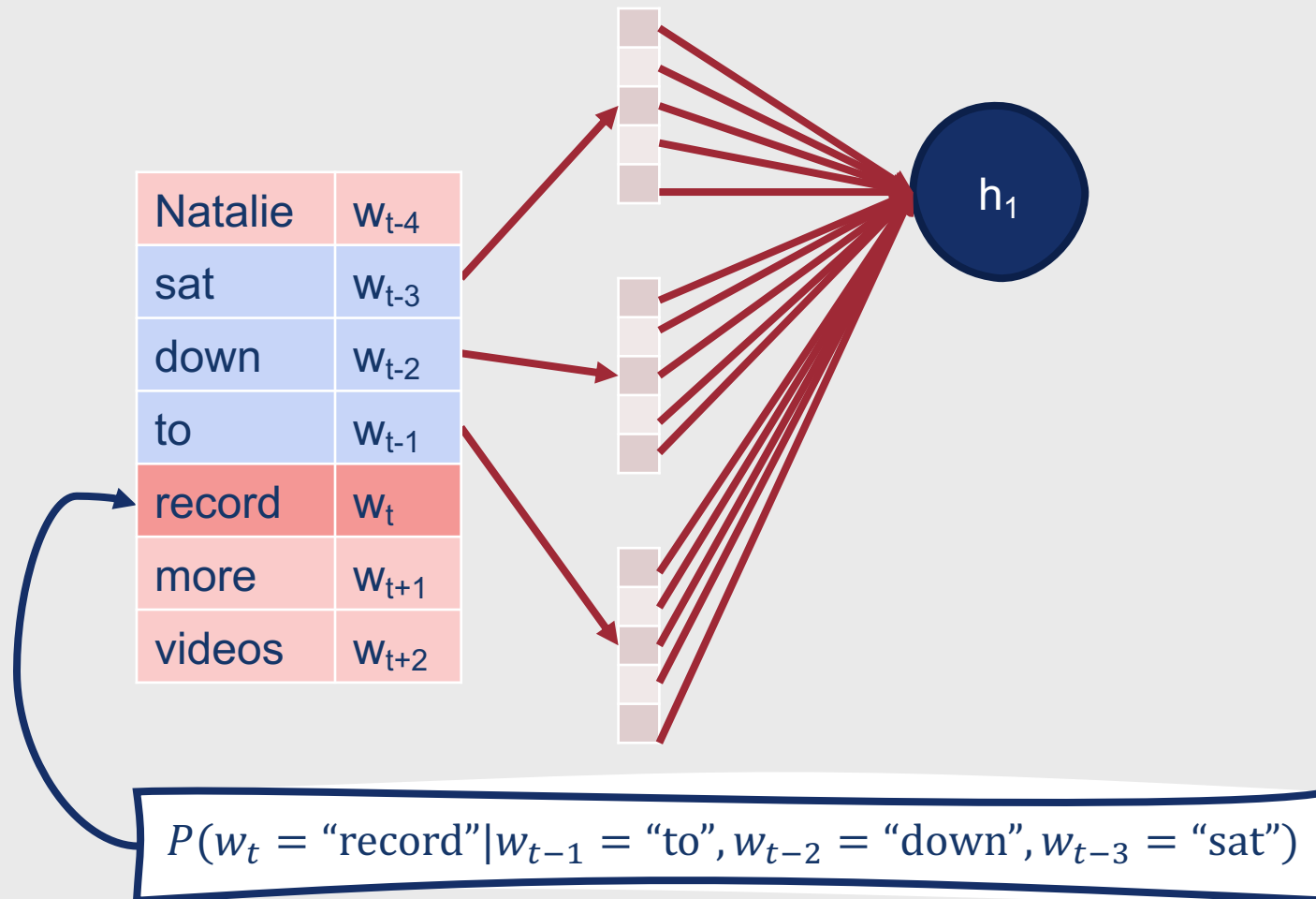
Natalie	w_{t-4}
sat	w_{t-3}
down	w_{t-2}
to	w_{t-1}
record	w_t
more	w_{t+1}
videos	w_{t+2}


$$P(w_t = \text{"record"} | w_{t-1} = \text{"to"}, w_{t-2} = \text{"down"}, w_{t-3} = \text{"sat"})$$

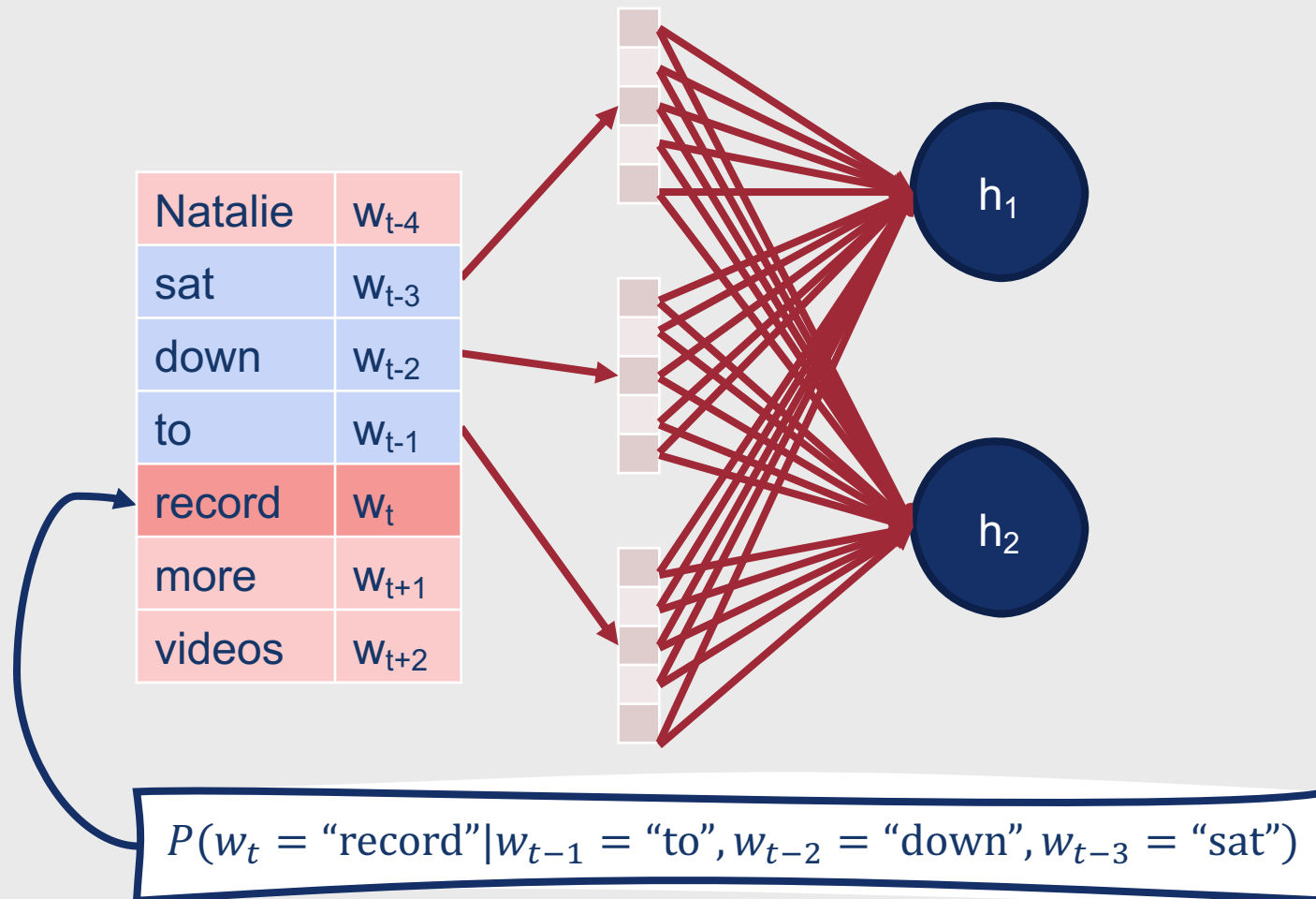
Neural Language Model



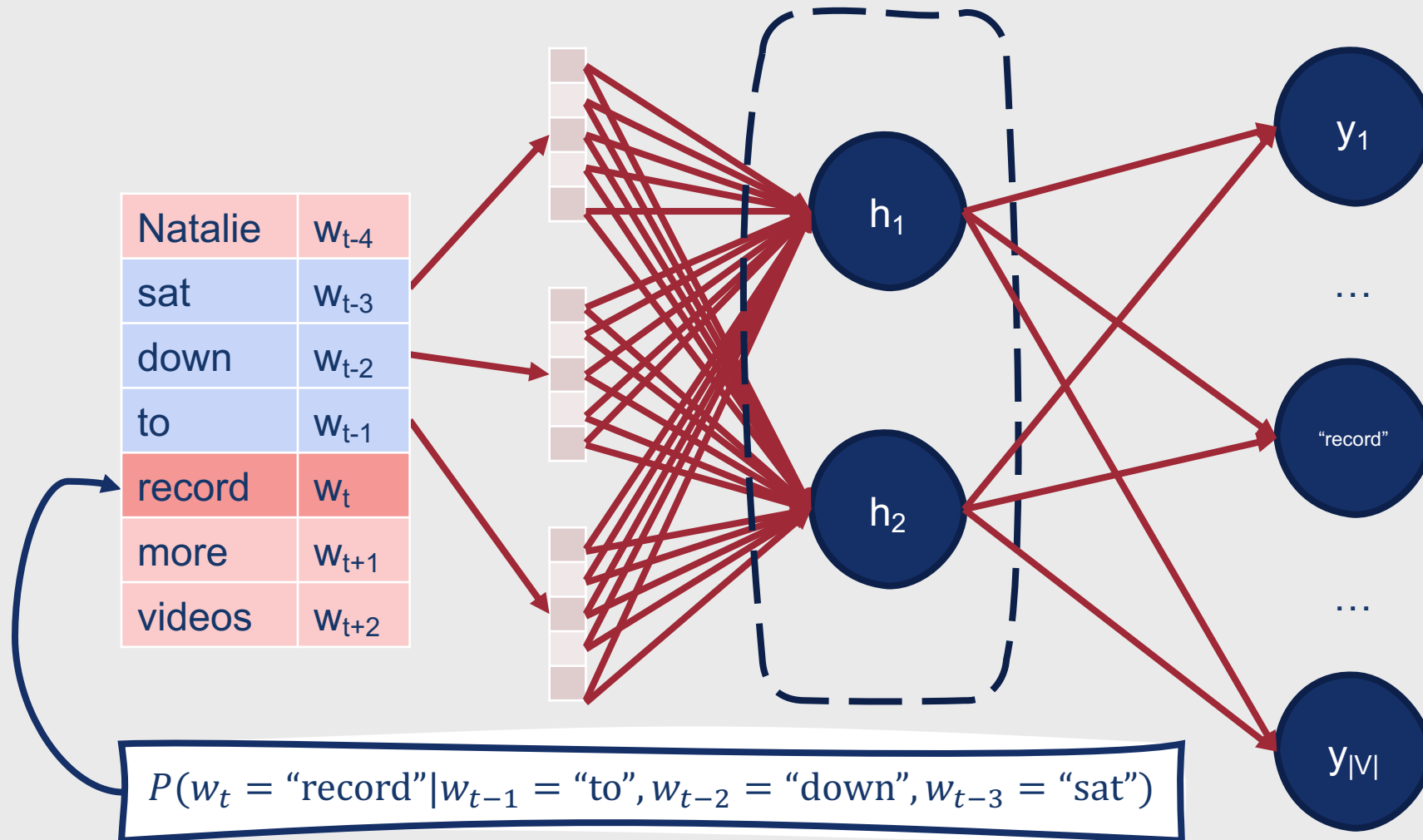
Neural Language Model



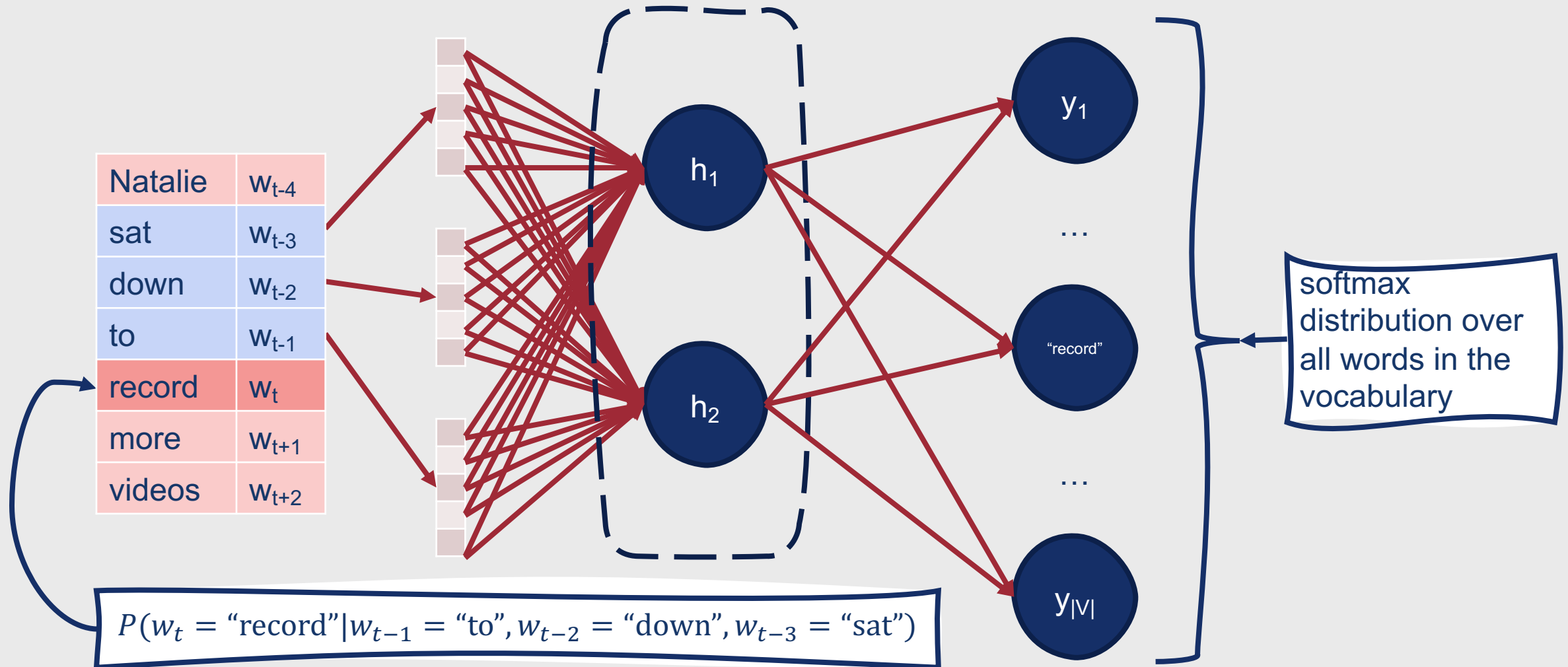
Neural Language Model



Neural Language Model



Neural Language Model



- When we use another algorithm to learn the embeddings for our input words, this is called **pretraining**
- However, sometimes it's preferable to learn embeddings while training the network, rather than using **pretrained embeddings**
 - E.g., if the desired application places strong constraints on what makes a good representation

**What if we
don't already
have dense
word
embeddings?**

Learning New Embeddings

- Start with a **one-hot vector** for each word in the vocabulary
 - Element for a given word is set to 1
 - All other elements are set to 0
- Randomly initialize the hidden (weight/embedding) layer
- Maintain a separate vector of weights for that layer, for each vocabulary word

Formal Definition: Learning New Embeddings

- Letting E be an embedding matrix, with one row for each word in the vocabulary:
 - $\mathbf{e} = (E_{x_1}, E_{x_2}, \dots, E_{x_n})$
 - $\mathbf{h} = \sigma(W\mathbf{e} + \mathbf{b})$
 - $\mathbf{z} = U\mathbf{h}$
 - $y = \text{softmax}(\mathbf{z})$
- Optimizing this network using the same techniques discussed for other neural networks will result in both
 - A model that predicts words
 - A new set of word embeddings that can be used for other tasks

Neural Language Model

