CS 232: Artificial Intelligence

Spring 2024

Prof. Carolyn Anderson
Wellesley College

Reminders

- Reading for Tuesday: Illustrated Stable Diffusion blog post
- Reading for Friday: Chiang (2023)
- Tensorflow version compatibility issues: check email I sent for how to downgrade Tensorflow to 2.13

  pip install transformers

  pip install tensorflow==2.1
  3
- My help hours today: 3:30-4:30
- My help hours Monday: 4-5:15
- Lyra's Sunday help hours: 4-6

New Policy: Earn Bonus Late Days

You can earn bonus late days by attending a research talk. To be eligible:

- The talk must be on CS research or on research related to AI
- The talk must be live, not recorded (so you can ask questions)
- You must write a paragraph about the talk and what you learned and email it to me.

Upcoming Talks

BABSON COLLEGE

Renowned AI Ethics Pioneer is Coming to Babson!

6:00 PM | April 2 2024 | Winn Auditorium

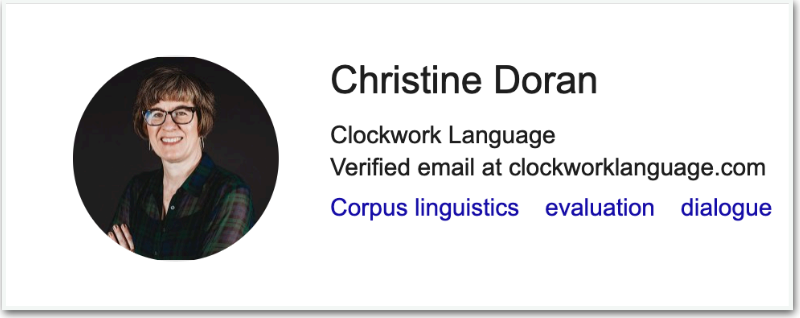
Dr. Rumman Chowdhury

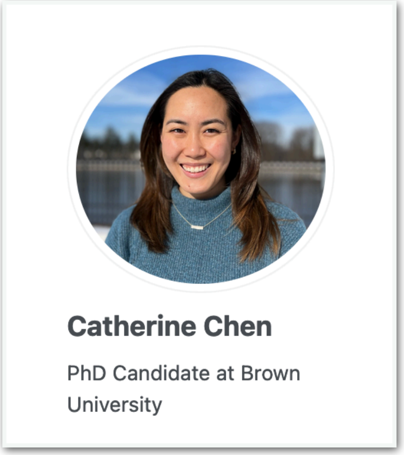
Named by Forbes as one of the five key people shaping AI - Dr. Chowdhury is the former Director of Machine Learning, Ethics, Transparency, and Accountability team at Twitter and now CEO and co-founder of Humane Intelligence.

Join us for an enlightening session that explores the intersection of AI, ethics, policy, and entrepreneurship.

Butler Institute for Free Enterprise Through Entrepreneurship

Upcoming Talks



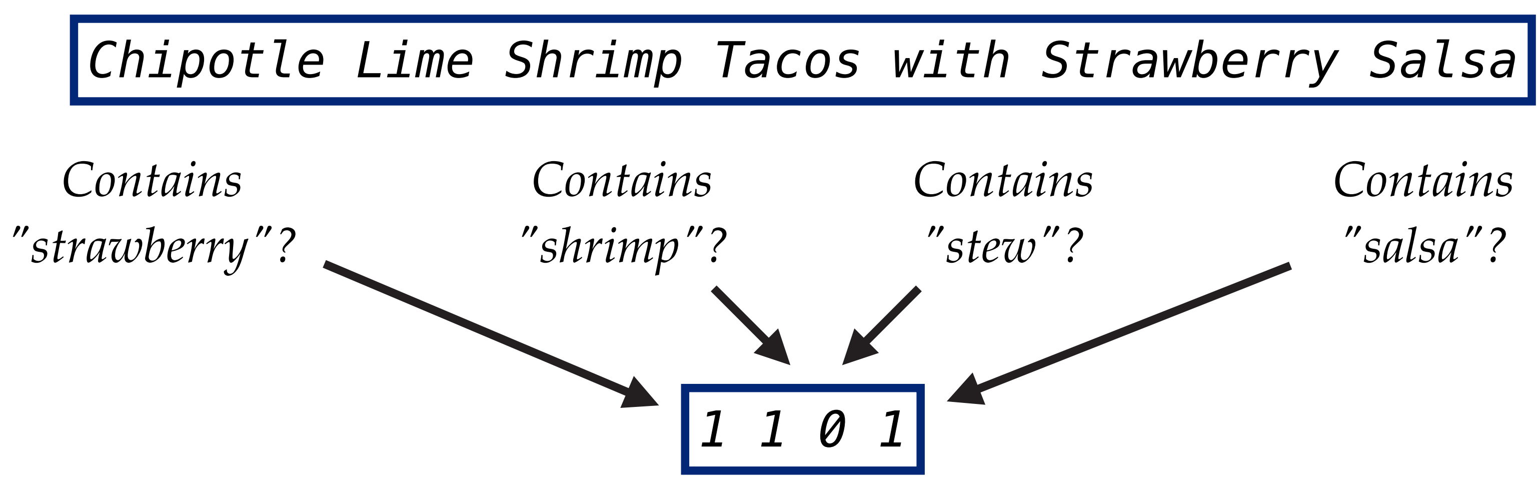


Representation Learning

How Do We Represent Text?

In the next homework assignment, you will try to improve our recipe classifier using neural networks instead of regression.

To feed text into a neural network, we need to turn it into numbers. In our regression classifier, we did this by **hand-crafting features**.



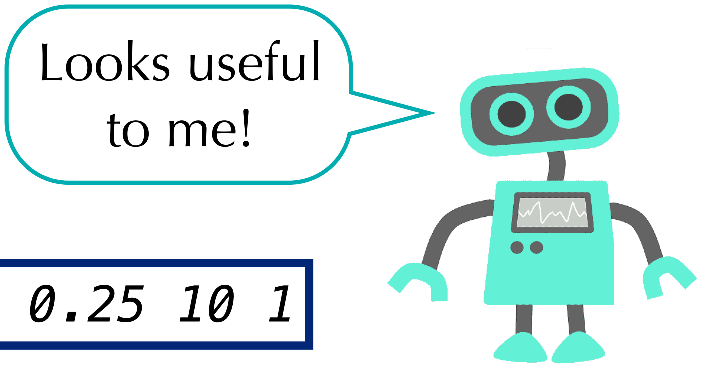
Representation Learning

From now on, we're going to use neural networks to **learn representations** for us.

`Chipotle Lime Shrimp Tacos with Strawberry Salsa`

↓

`2 10 -15 110 0 -31 475 19 ... -3 0.25 10 1`



Representation Learning: a machine learning technique for extracted features (informative aspects) from data.

Word Vectors

Idea: a word's meaning is based on its **distance** from other word meanings.

Each word = a vector (not just "good" or "w_{45}")

Similar words are "**nearby in semantic space**"

We build this space automatically by seeing which words are **nearby in text**

Word Embeddings

Which of these word pairs are most alike?

celestial

sun --- moon 6 bright in the sky

round

sun --- lightbulb 5

sun --- mystical 2

moon --- lightbulb 3

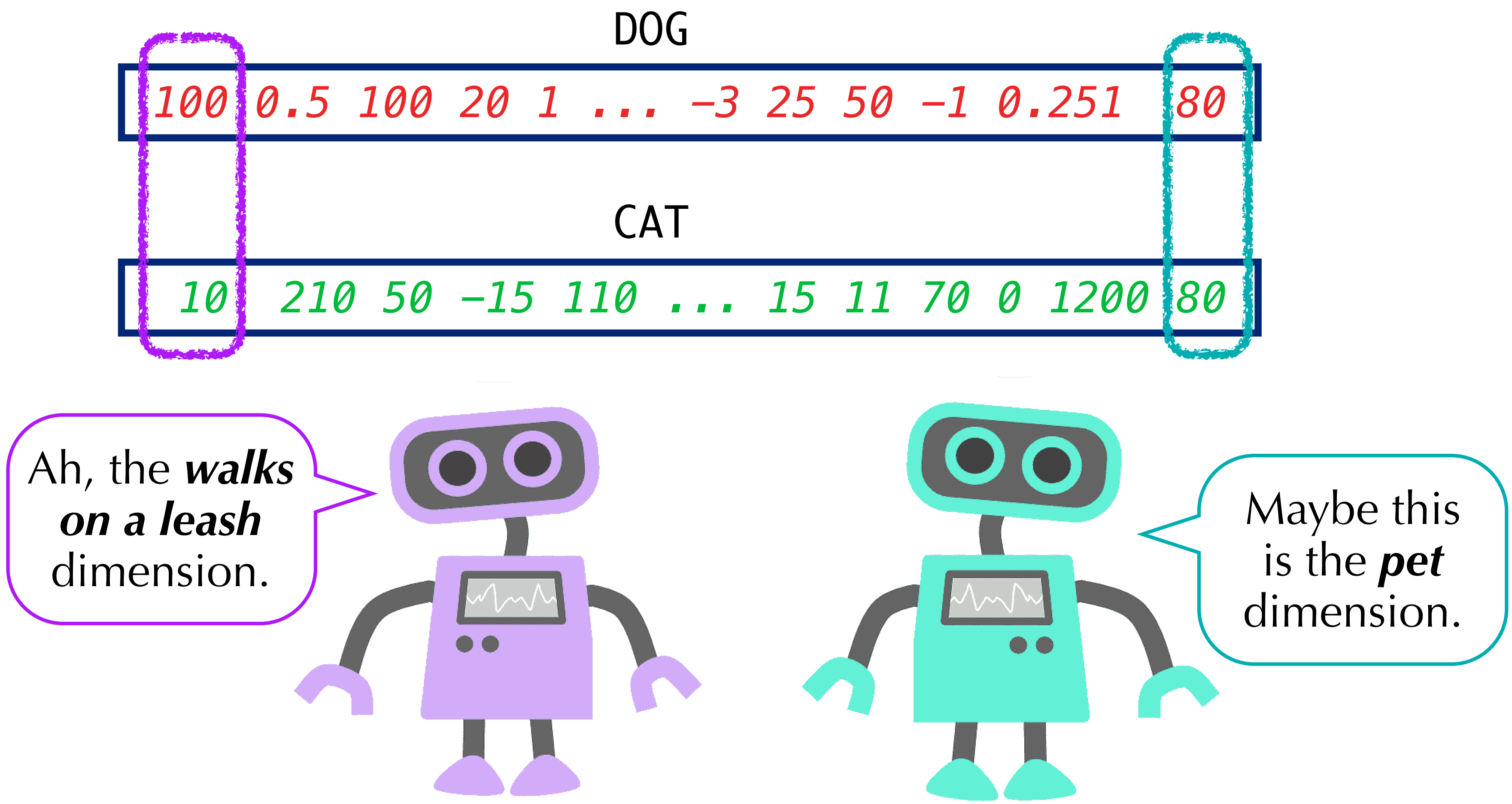
moon --- mystical 4 nights are more mystical

twilight supernatural folklore

mystical --- lightbulb 1

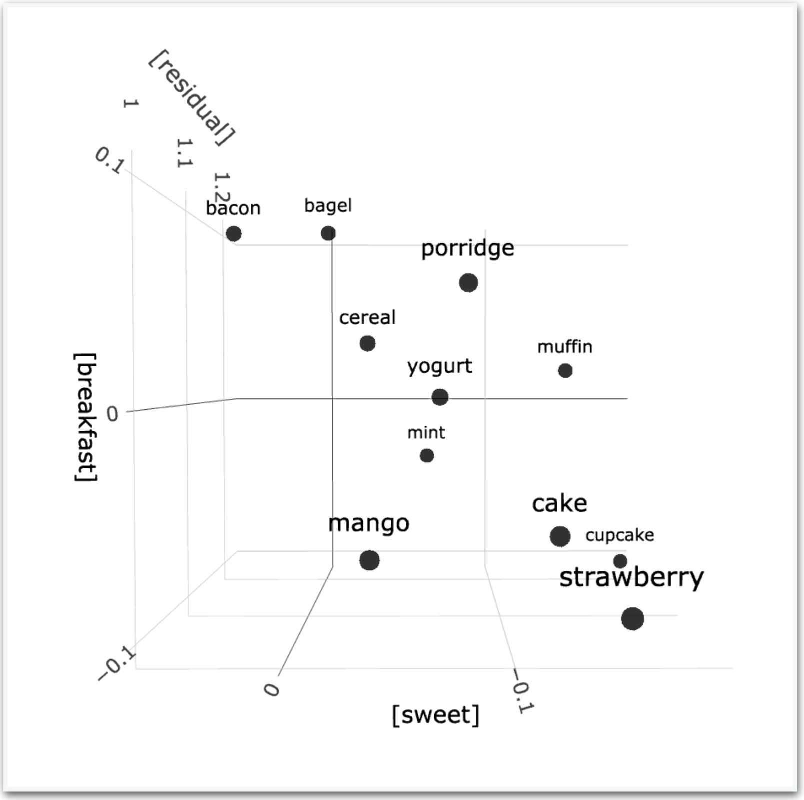
Word Embeddings

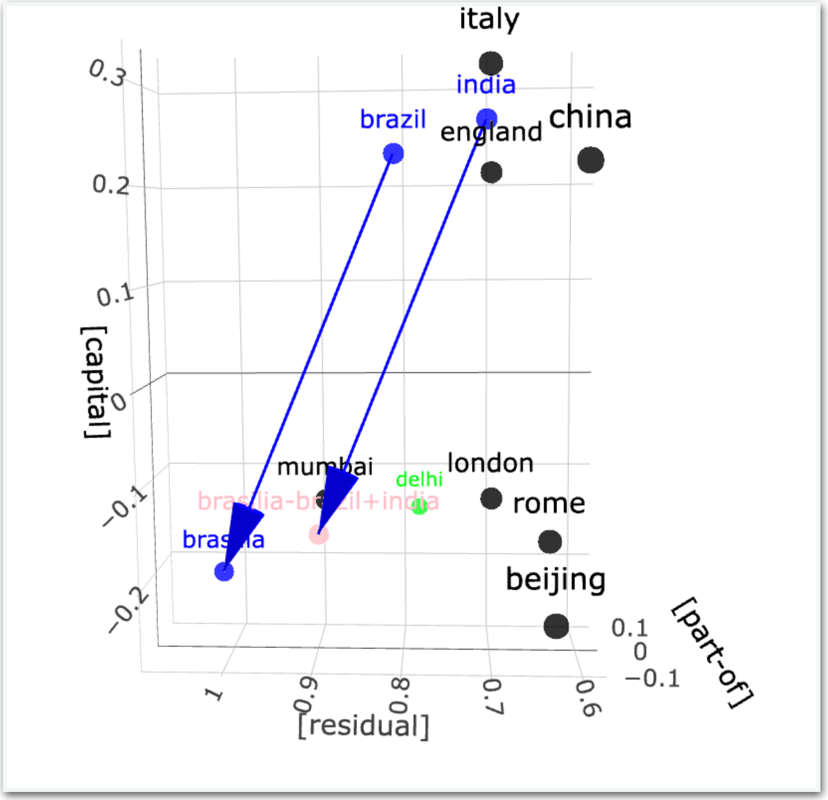
Imagine defining a large number of ways that words can be similar (*dimensions*). Maybe around 2000 ways?



Word Embeddings

If we have good word embeddings, their geometric relationships should be meaningful:





https://www.cs.cmu.edu/~dst/WordEmbeddingDemo

Neural Networks with Word Embedding Features

Neural Net Classification with embeddings as input features!

Single Unit View

hidden state output: 72

non-linear activation: ReLu

z-score: 72

take the dot product (multiply each pair & sum)

weights

0.25 0 ... −1 0.5 1   2 0.1 ... 1 0.5 0   1 0 ... −3 0.5 0

embeddings

2 19 ... −3 0.25 10

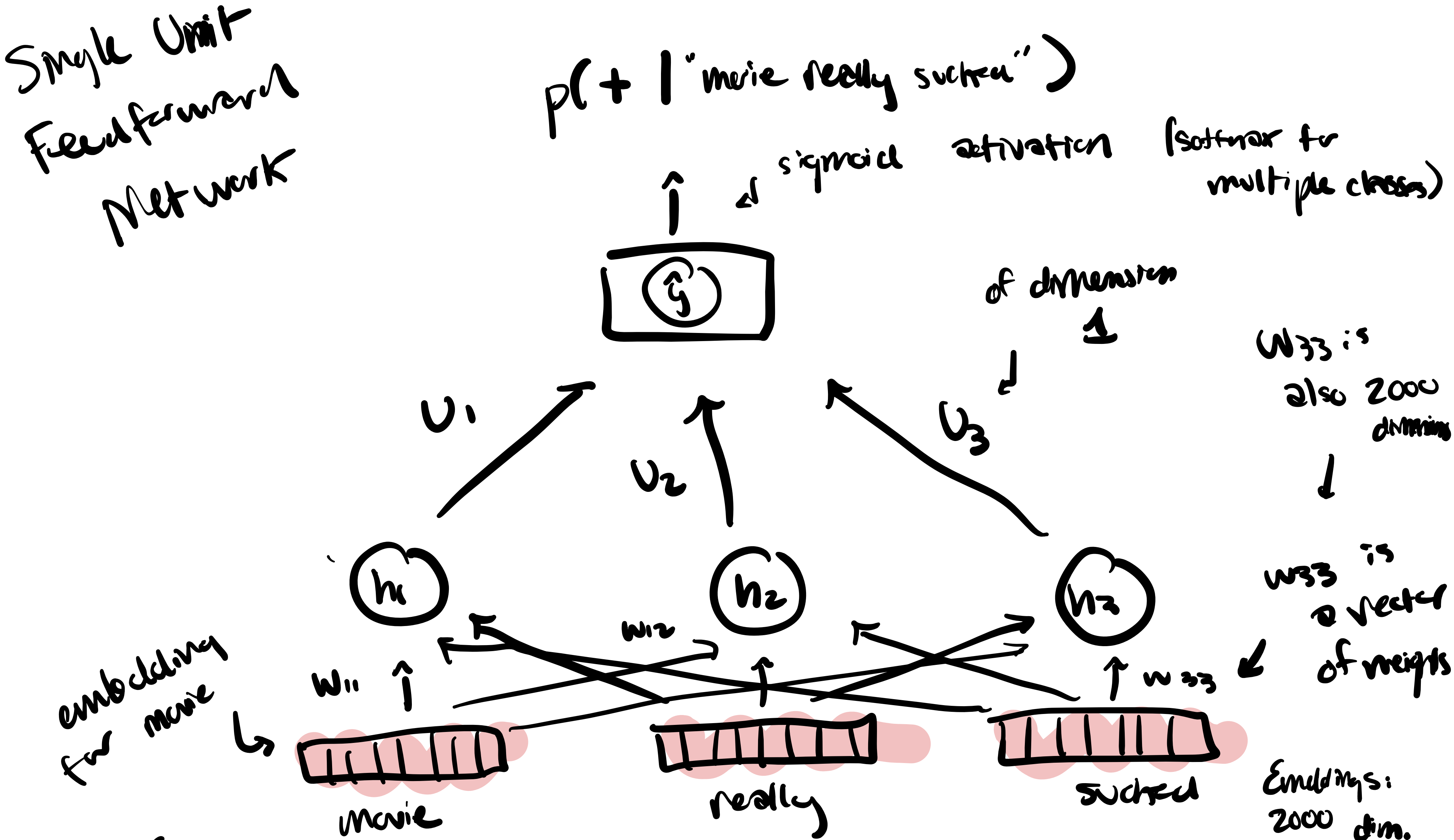
−15 110 ... 105 1 5

−31 475 ... 0.5 72 1

tokenization: 21 movie | 781 really | 5100 sucked

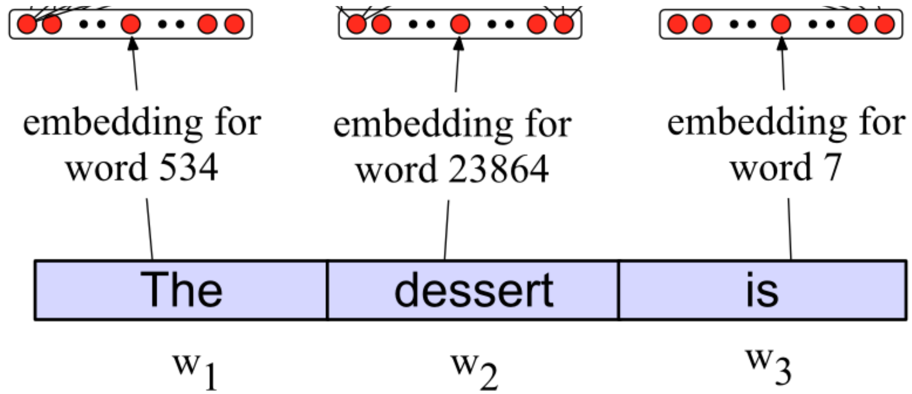
this movie really sucked

Neural Net Classification with embeddings as input features!



$$W = \{w_{11}, w_{12}, w_{13}, w_{21}, w_{22}, w_{23}, w_{31}, w_{32}, w_{33}\}$$

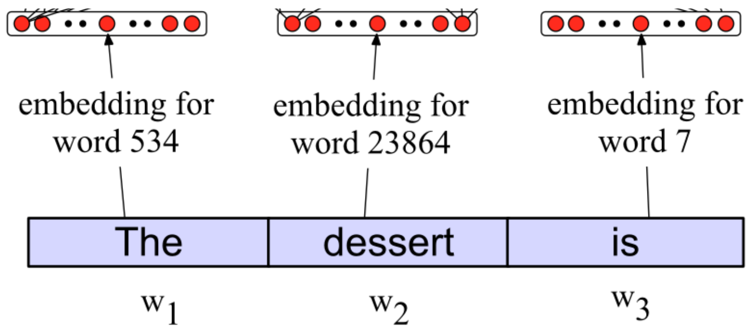
Issue: texts come in different sizes



1. Make the input the length of the longest review
   If a review is long, "pad" it with zero embeddings
   Truncate if review is too long.
2. Create a "sentence embedding" to represent all words

Issue: texts come in different sizes

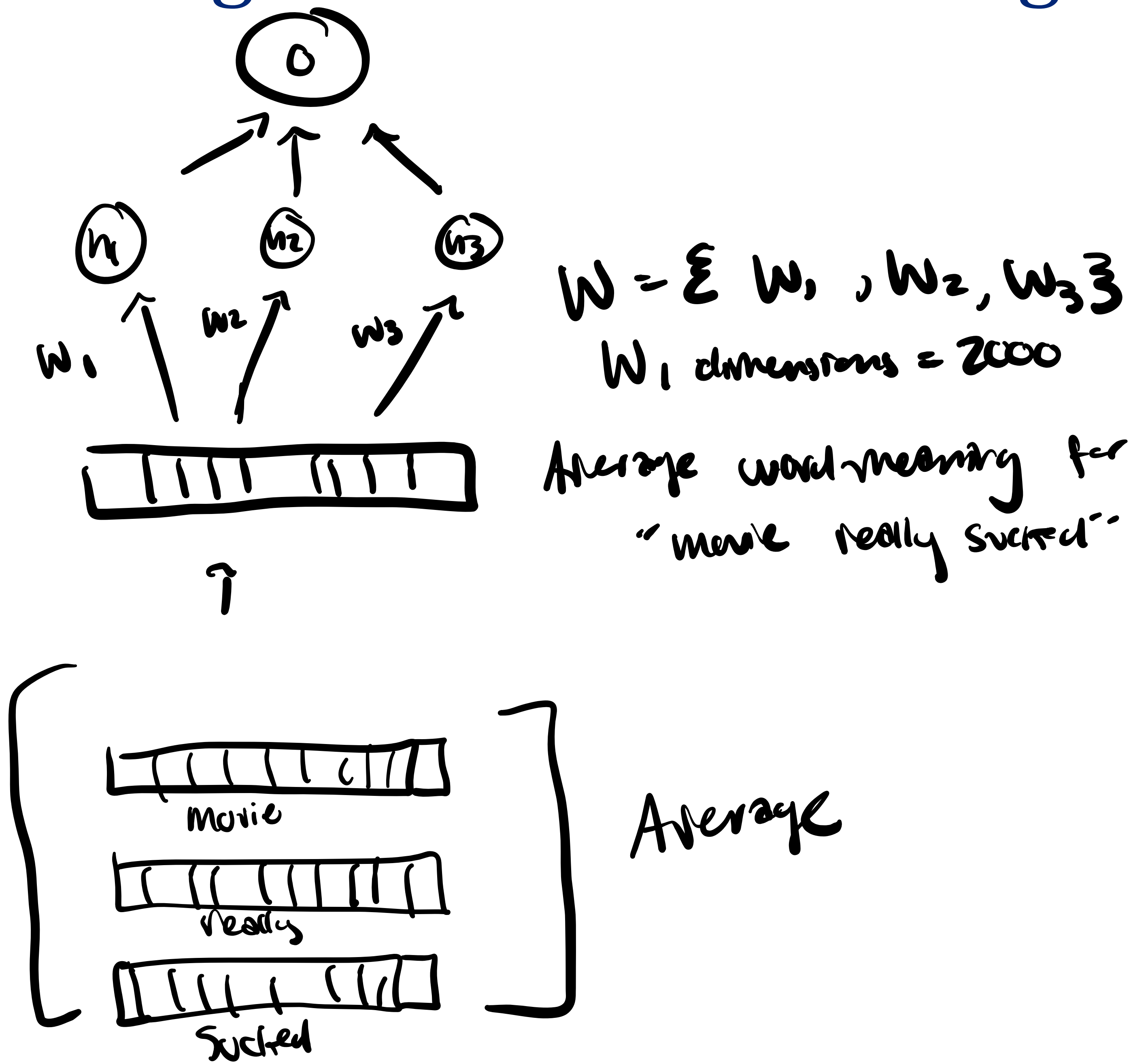
This assumes a fixed size length (3)!



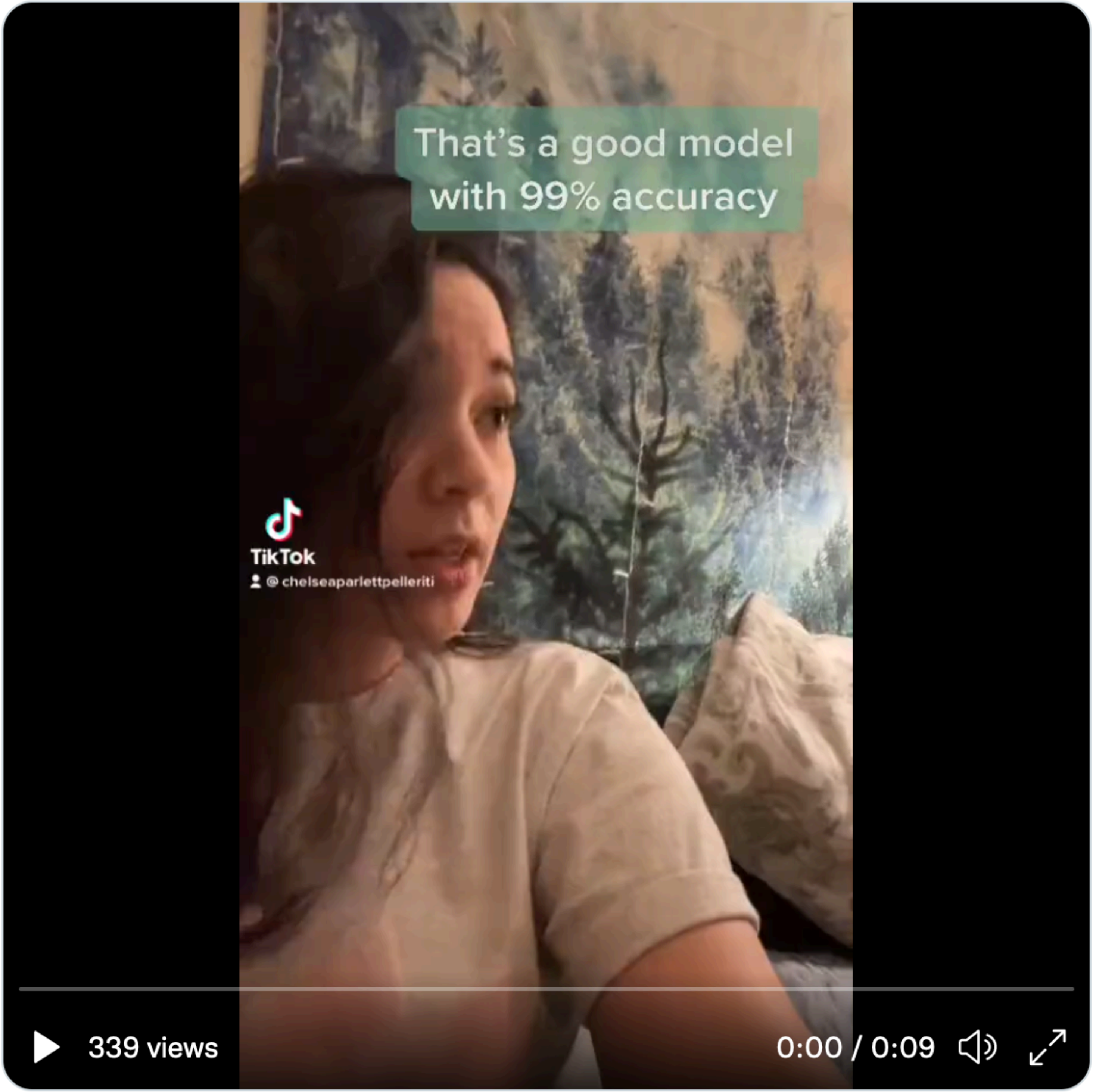
Some simple solutions (more sophisticated solutions later)

1. Make the input the length of the longest review
   - If shorter then pad with zero embeddings
   - Truncate if you get longer reviews at test time
2. Create a single "sentence embedding" (the same dimensionality as a word) to represent all the words
   - Take the mean of all the word embeddings
   - Take the element-wise max of all the word embeddings
     - For each dimension, pick the max value from all words

Solution 2: Average the word embeddings



Revisiting Our Classifier



https://www.tiktok.com/@chelseaparlettpelleriti/video/7072586373064248622?is_from_webapp=1&sender_device=pc&web_id=7159271848050869802

AI Tasks

Search

Uninformed Search

Informed Search

Adversarial Games

Navigation

Learning Under Uncertainty

Classification

Regression

Sentiment Analysis

Neural Networks

Image Classification

Text Classification

Generation

Language Models

Image Generation

Chatbots

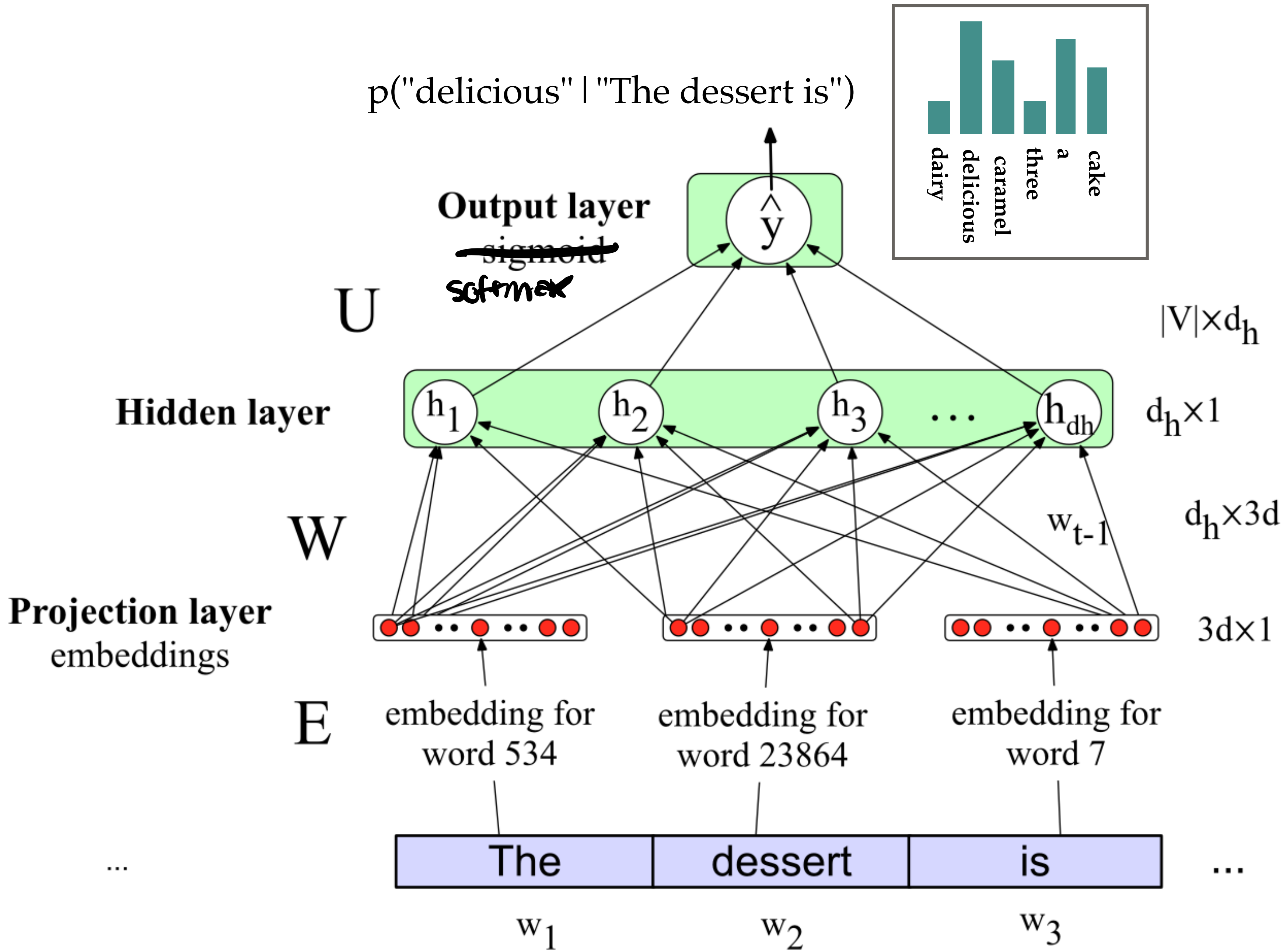
Finetuning

Prompt Engineering

We're moving into generation!

Language Modeling
(Text Generation)

Neural Net Classification with embeddings as input features!



Language Generation

So far we have used language models to predict the next word in a sequence and estimate the probability of a sentence.

How do we **generate** sentences?

Language Generation

We sample words according to their estimated probabilities:

$$P(\text{english} \mid \text{want}) = .0011$$

$$P(\text{chinese} \mid \text{want}) = .0065$$

$$P(\text{to} \mid \text{want}) = .66$$

$$P(\text{eat} \mid \text{to}) = .28$$

$$P(\text{food} \mid \text{to}) = 0$$

$$P(\text{want} \mid \text{spend}) = 0$$

$$P(\text{i} \mid \langle s \rangle) = .25$$

Language Generation

- Start the sentence
- Sample a next word according to its probability
- Keep going!

represent beginning of sentence

1st guess

2nd guess

```
<s> I
    I want
      want to
           to eat
              eat Chinese
                  Chinese food
                          food  </s>
I want to eat Chinese food
```