CS 232: Artificial Intelligence

Fall 2026

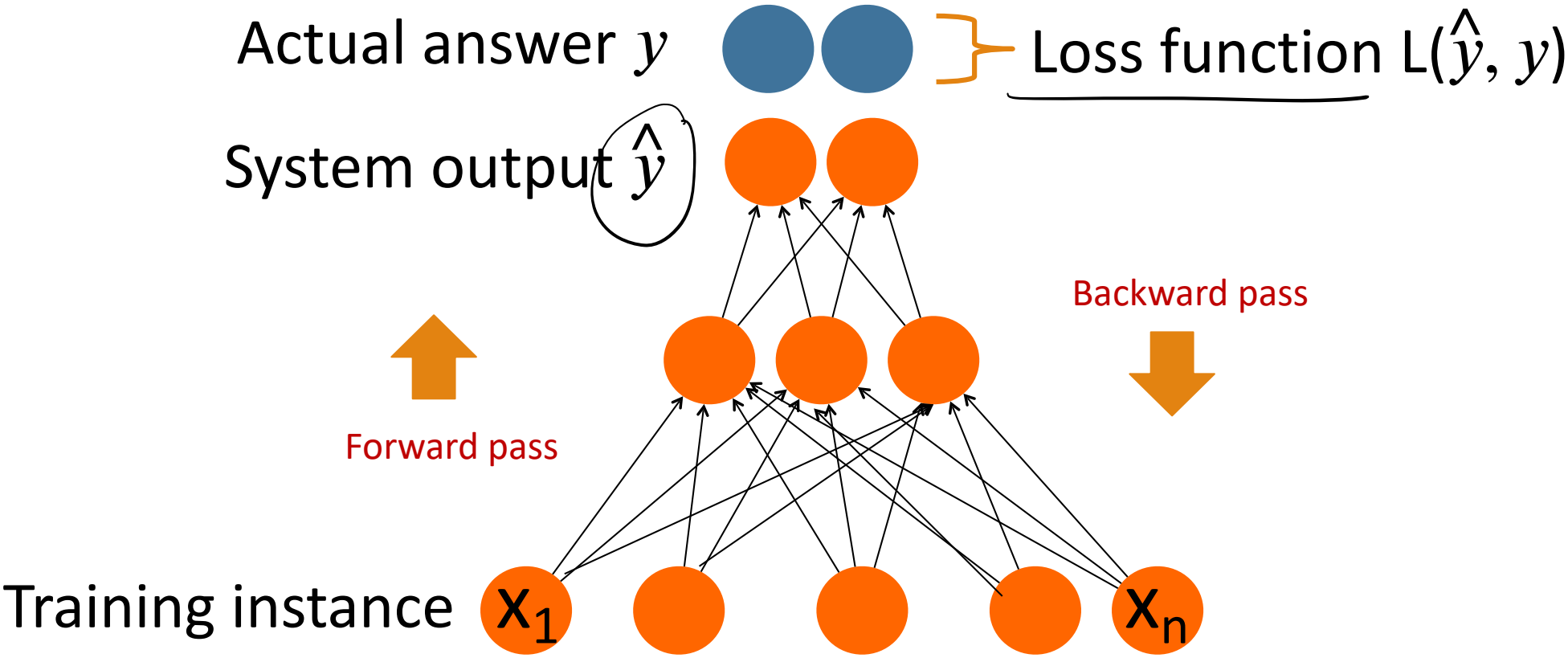
Prof. Carolyn Anderson
Wellesley College

Reminders

- Midterm 2 is next Tuesday!
  - You may bring an index card of notes
  - You may use a calculator
  - Midterm 2 review materials are up
- My help hours: Friday 3-4 in W423

Recap

Intuition: training a 2-layer Network



Our Semester

Search Unit

Uninformed Search

Informed Search

Adversarial Search

Reinforcement Learning

Classification Unit

Regression

Neural Networks

Image Classification

Text Classification

Generation Unit

Language Models

Chatbots

Prompt Engineering

Reinforcement Learning (Again!)



We're moving into generation!

Data Representation for Supervised Machine Learning

Data Representation for Machine Learning

Neural networks require numeric inputs. To train them on real-world data, we need a way of representing the data numerically.

Today we'll consider two challenging but common types of data: images and text.

Image Representations

Grayscale Images are Matrices

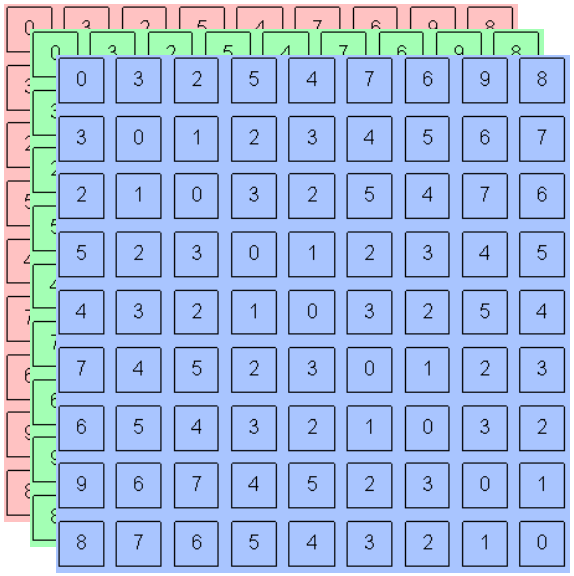


0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

height x width

Color Images are Tensors





channel x height x width

Channels are usually RGB: Red, Green, and Blue

Other color spaces: HSV, HSL, LUV, XYZ, Lab, CMYK

Alexnet

ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca

the paper that started the deep learning revolution!

image classification

Classify an image into 1000 possible classes:
e.g. Abyssinian cat, Bulldog, French Terrier, Cormorant, Chickadee,
red fox, banjo, barbell, hourglass, knot, maze, viaduct, etc.

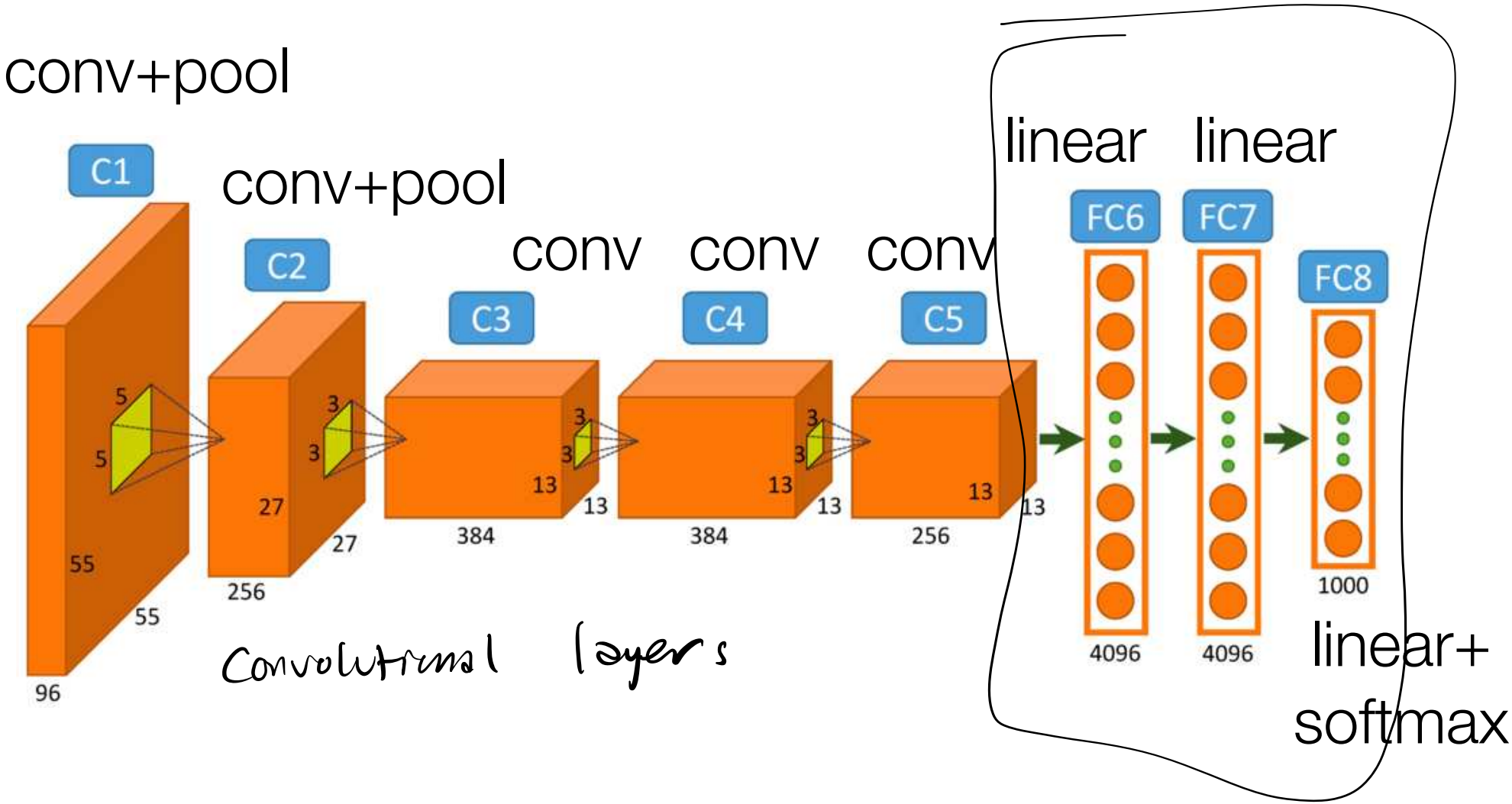


cat, tabby cat (0.71)
Egyptian cat (0.22)
red fox (0.11)
…..

train on the ImageNet challenge dataset, ~1.2 million images

Slides adapted from Mohit Iyyer

Alexnet

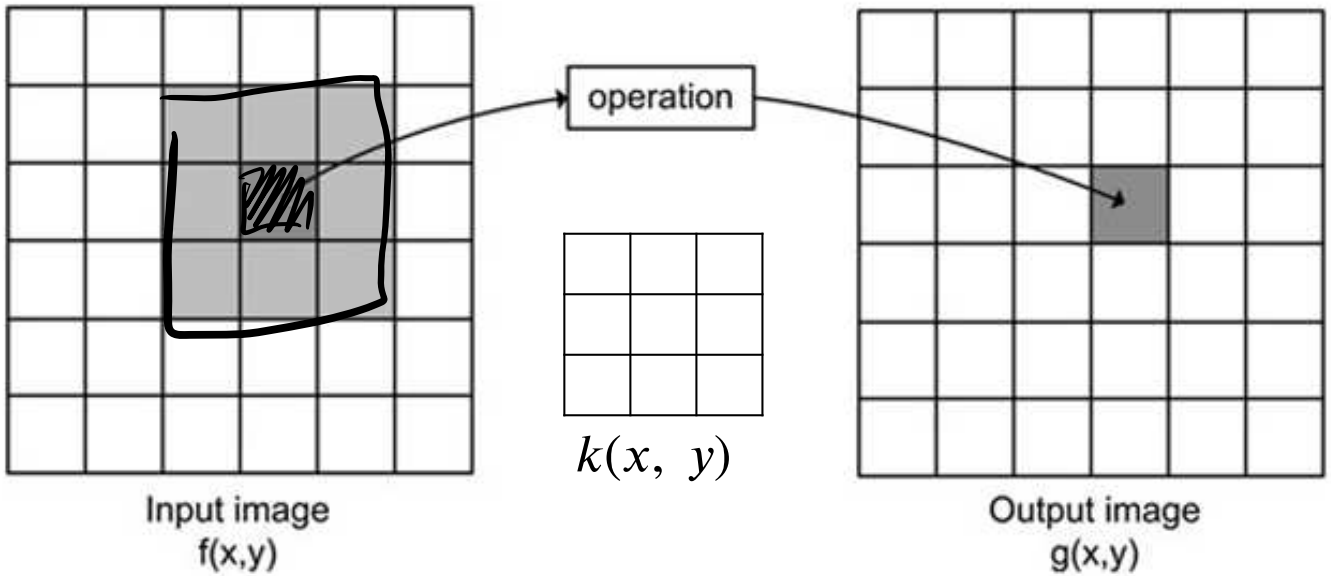


https://www.saagie.com/fr/blog/object-detection-part1

Slides adapted from Mohit Iyyer

Convolution Operator

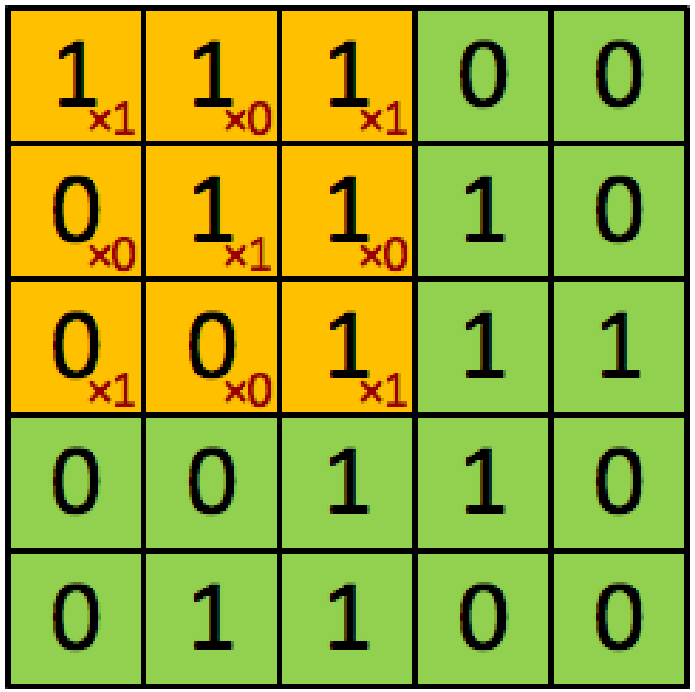
3 x 3 filter



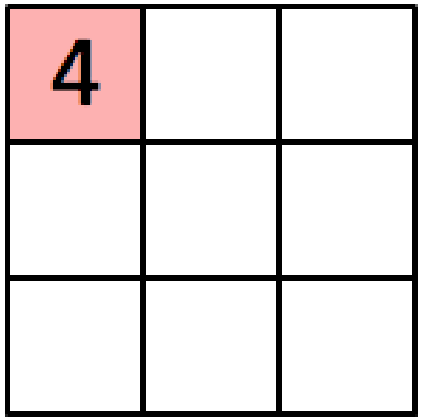
$$g(x,\ y) = \sum_{v} \sum_{u} k(u,v) f(x\ - u,\ y - v)$$

Image Credit: http://what-when-how.com/introduction-to-video-and-image-processing/neighborhood-processing-introduction-to-video-and-image-processing-part-1/

Convolution operation



Image



Convolved Feature

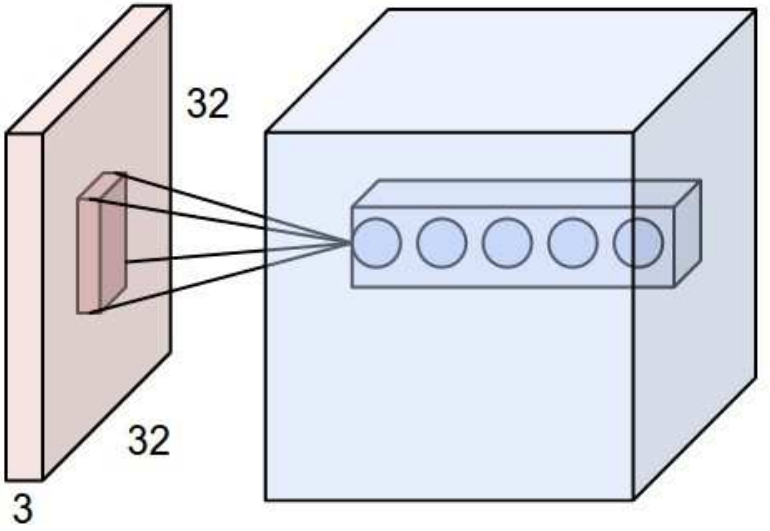
http://deeplearning.stanford.edu/tutorial/supervised/FeatureExtractionUsingConvolution

demo:

http://setosa.io/ev/image-kernels/

pooling layers also used to reduce dimensionality

Convolutional Layers: slide a set of small filters over the image



Pooling Layers: reduce dimensionality of representation

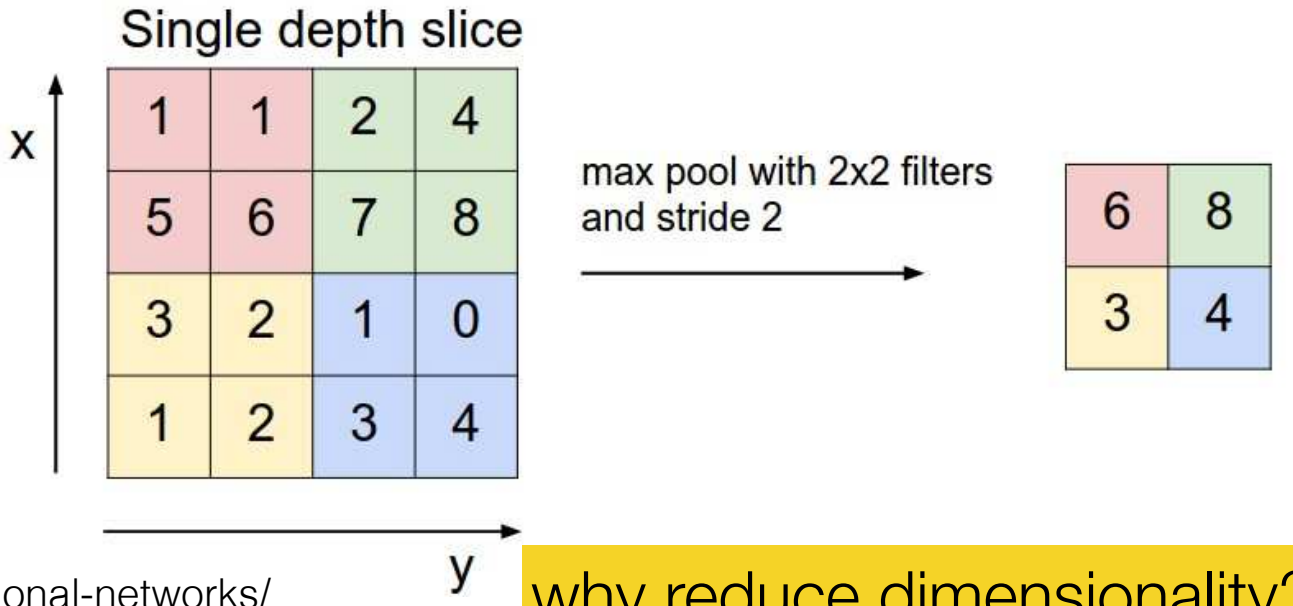
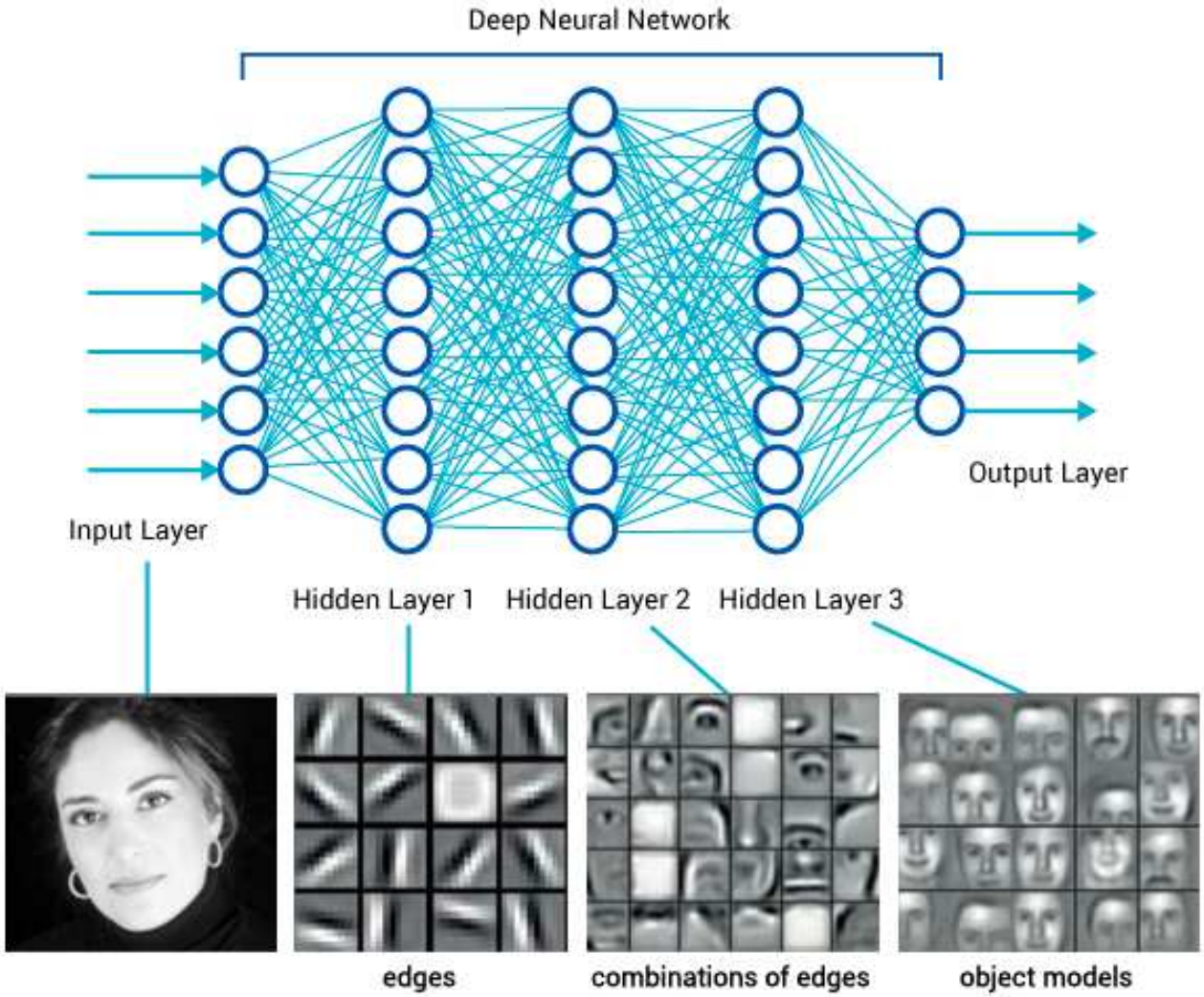


image: https://cs231n.github.io/convolutional-networks/

why reduce dimensionality?

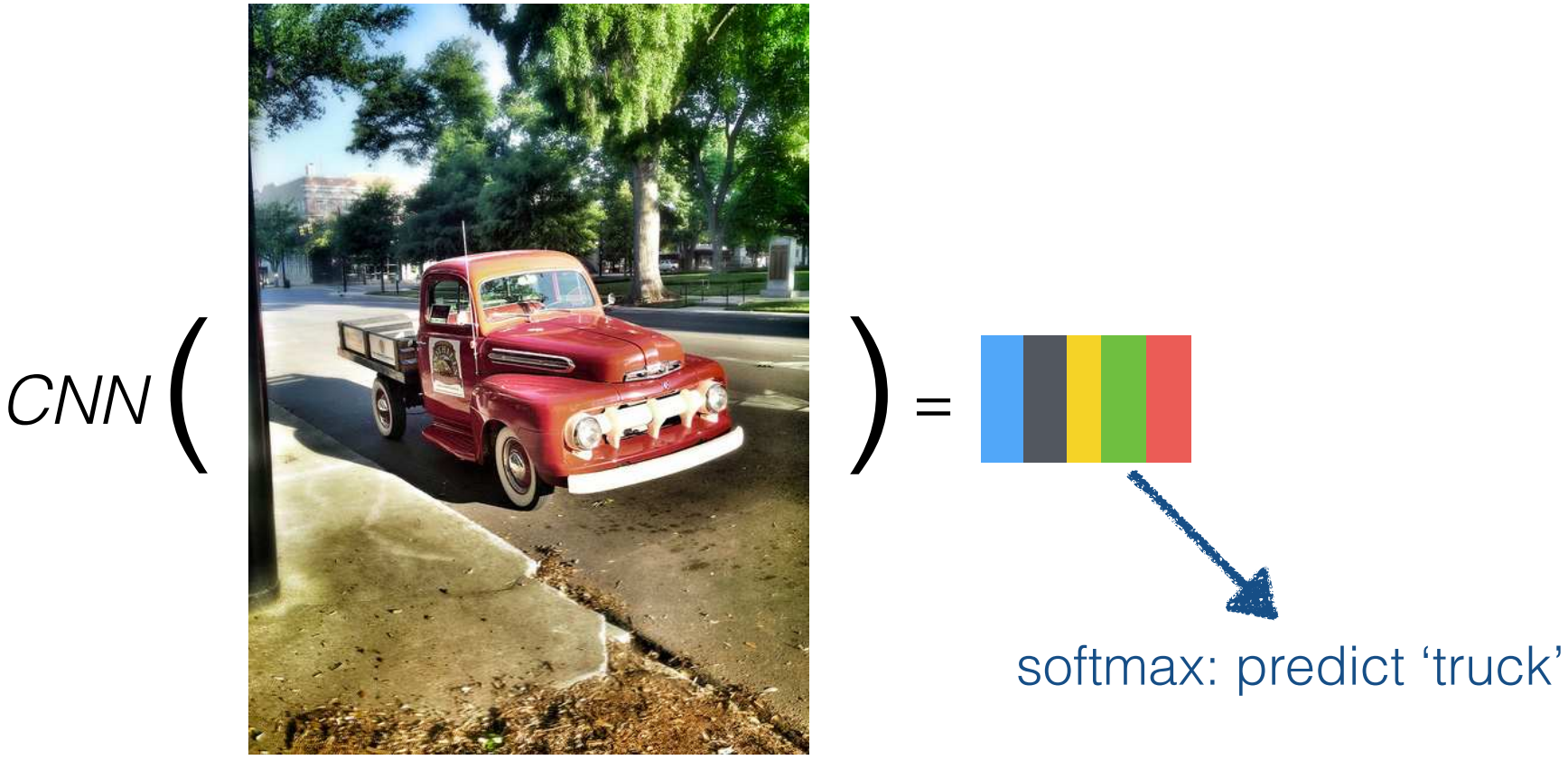
What is happening?



https://www.saagie.com/fr/blog/object-detection-part1

Slides adapted from Mohit Iyyer

at the end of the day, we generate a fixed size vector from an image and run a classifier over it

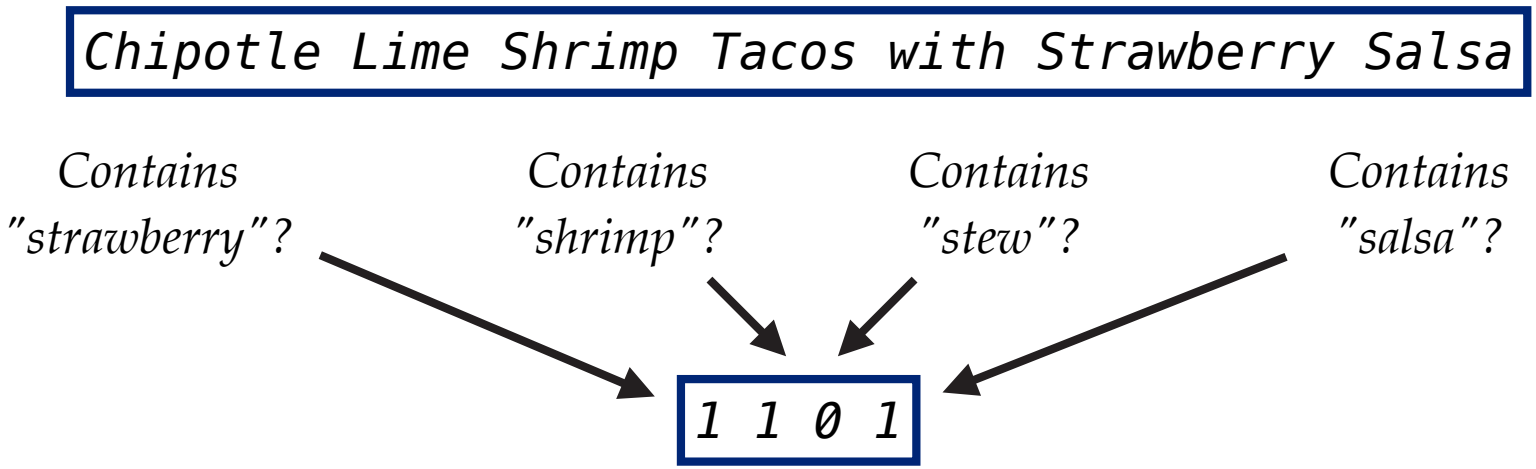


Slides adapted from Mohit Iyyer

Text Representations

How Do We Represent Text?

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

In our neural network version, we had the model **learn representations** for us.

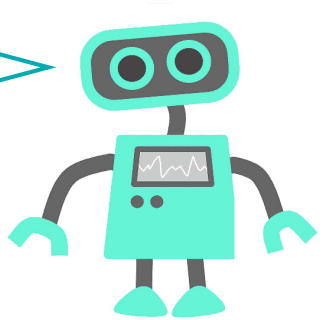
1 17 2 37 4025 101

`Chipotle Lime Shrimp Tacos with Strawberry Salsa`

V = 8

{ Chipotle: 1
Shrimp: 2
with: 4025

`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**"

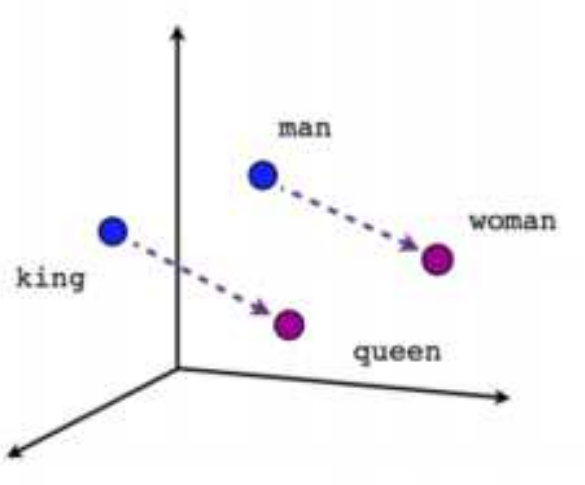
We build this space automatically by seeing which words are **nearby in text**



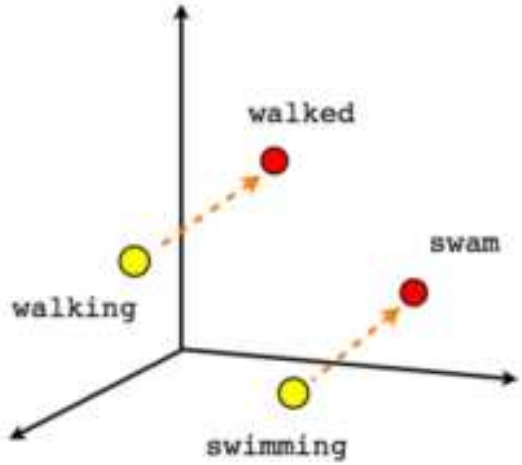
Word Embeddings

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

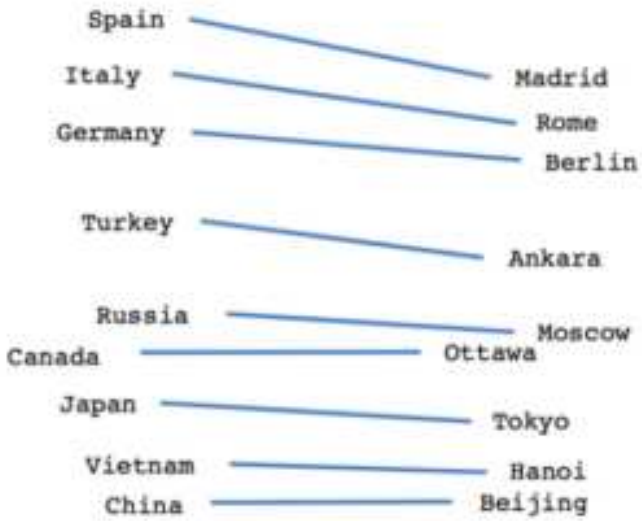
king =
[0.23, 1.3, -0.3, 0.43]



Male-Female

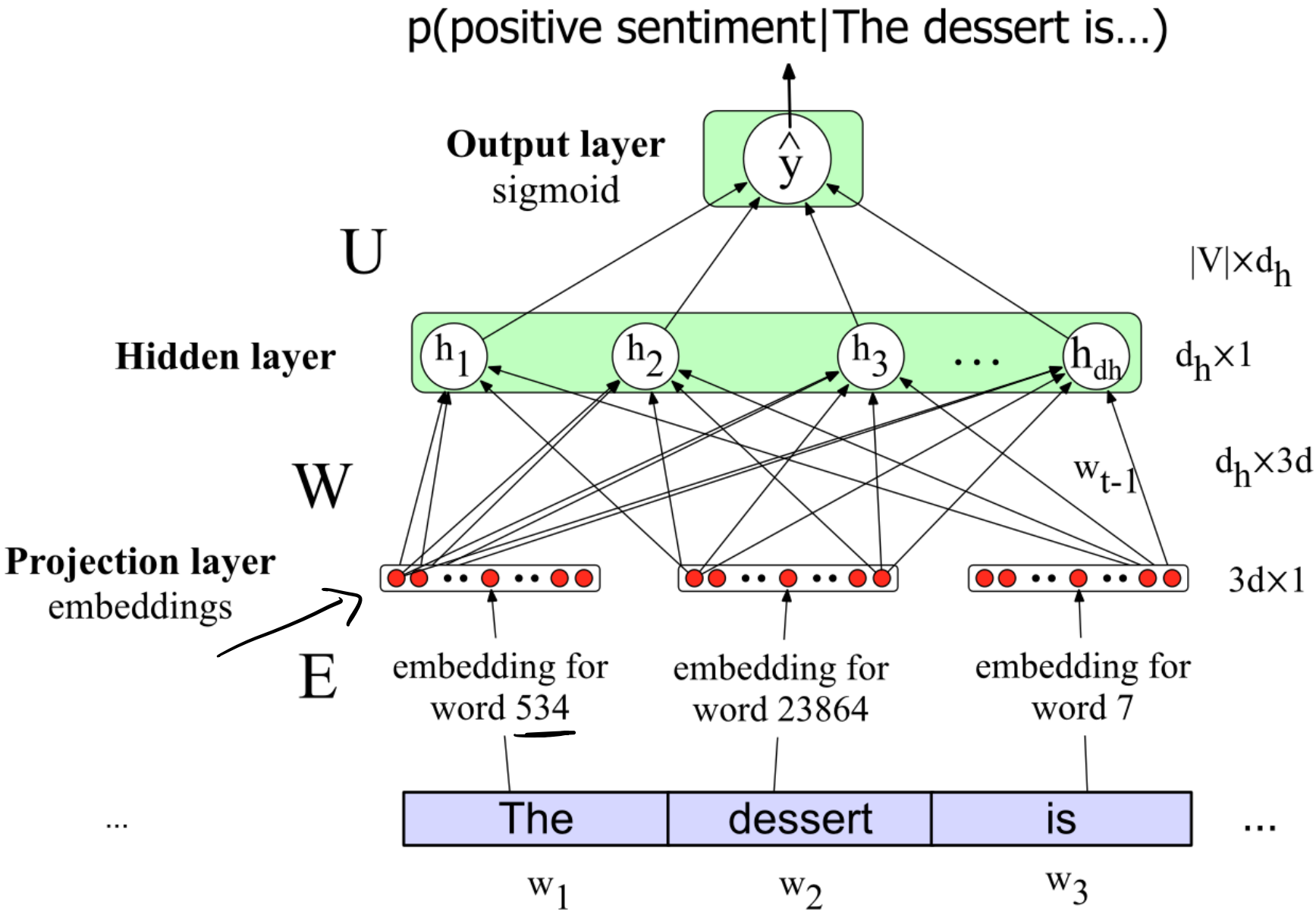


Verb tense



Country-Capital

Neural Net Classification with embeddings as input features!



Language Modeling

Language Model

A language model is a model that computes the probability of a sentence in a language:

$p(w_0, \ldots, w_n)$

A language model can also be used to compute the probability of the next word in a sentence:

$p(w_n \mid w_0, \ldots, w_{n-1})$

Sentence probability

How do we guess what the next word is?

I write this sitting in the kitchen ...

Answer: try out different words and compare their likelihood.

p(I write this sitting in the kitchen sink)

versus

p(I write this sitting in the kitchen knife)

versus

p(I write this sitting in the kitchen chair)

Sentence probability

But, since we never saw "I write this sitting in the kitchen sink", we still don't have a way to calculate

p("sink" | "I write this sitting in the kitchen")

What can we do?

Simple Language Model

For a very simple language model, we could **estimate** this conditional probability by looking at a smaller context window: a single word.

p(sink | "I wish this sitting in the kitchen")

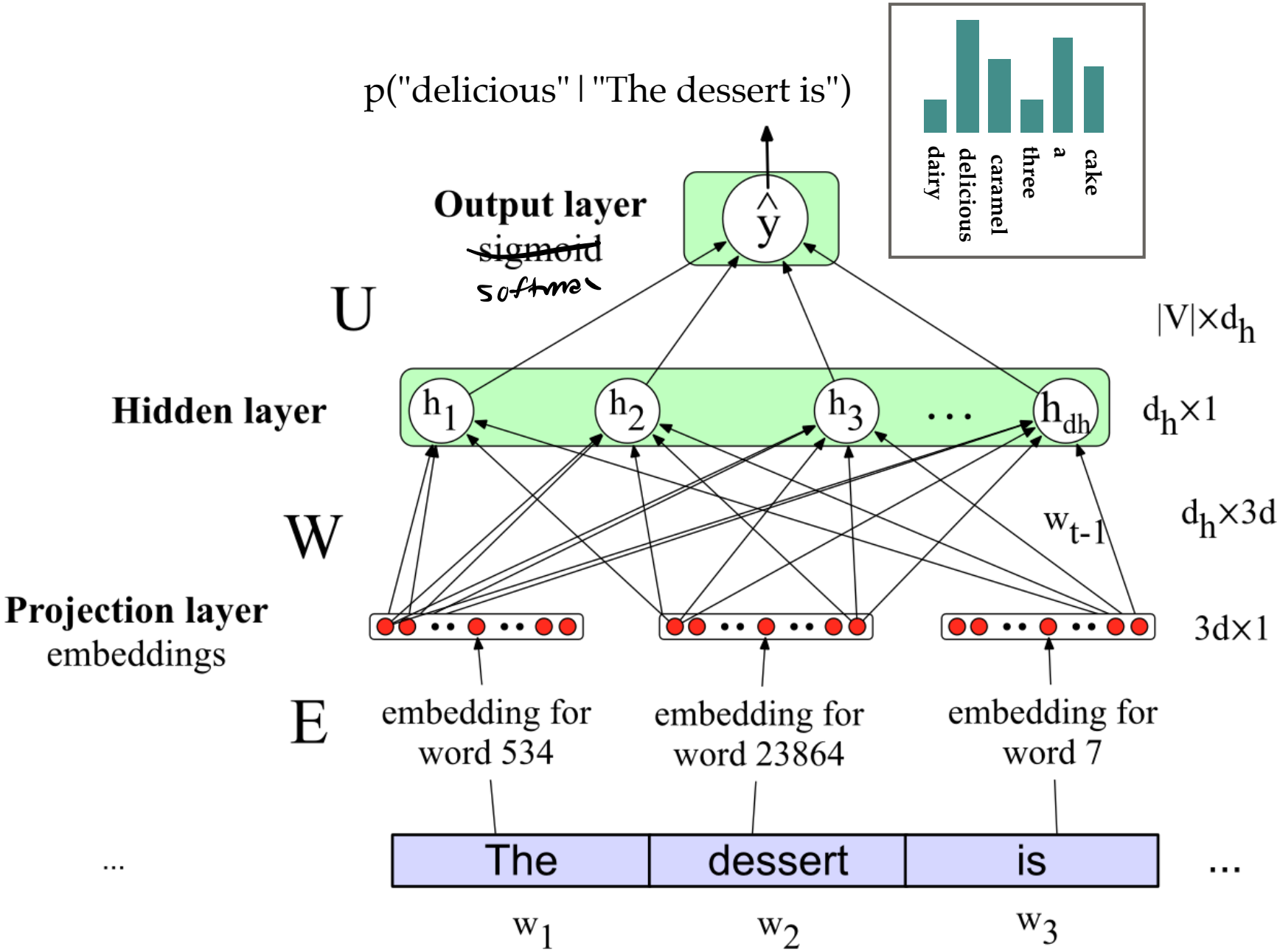
$\approx$

p (sink | "the kitchen")

n-gram

n : context window

Neural Net Classification with embeddings as input features!



Language Generation

How do we **generate** sentences?

Language Generation

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

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

Evaluation: How good is our model?

Does our language model prefer good sentences to bad ones?

Does it assign higher probability to “real” or “frequently observed” sentences than “ungrammatical” or “rarely observed” sentences?

AI Needs You Discussion: the Internet

- How familiar were you with the history of the Internet? Did you learn anything that surprised you?

- This section explores the tradeoff between freedom and safety on the Internet. Do you think the current balance makes sense? Should the Internet be more closely regulated? Less closely regulated?