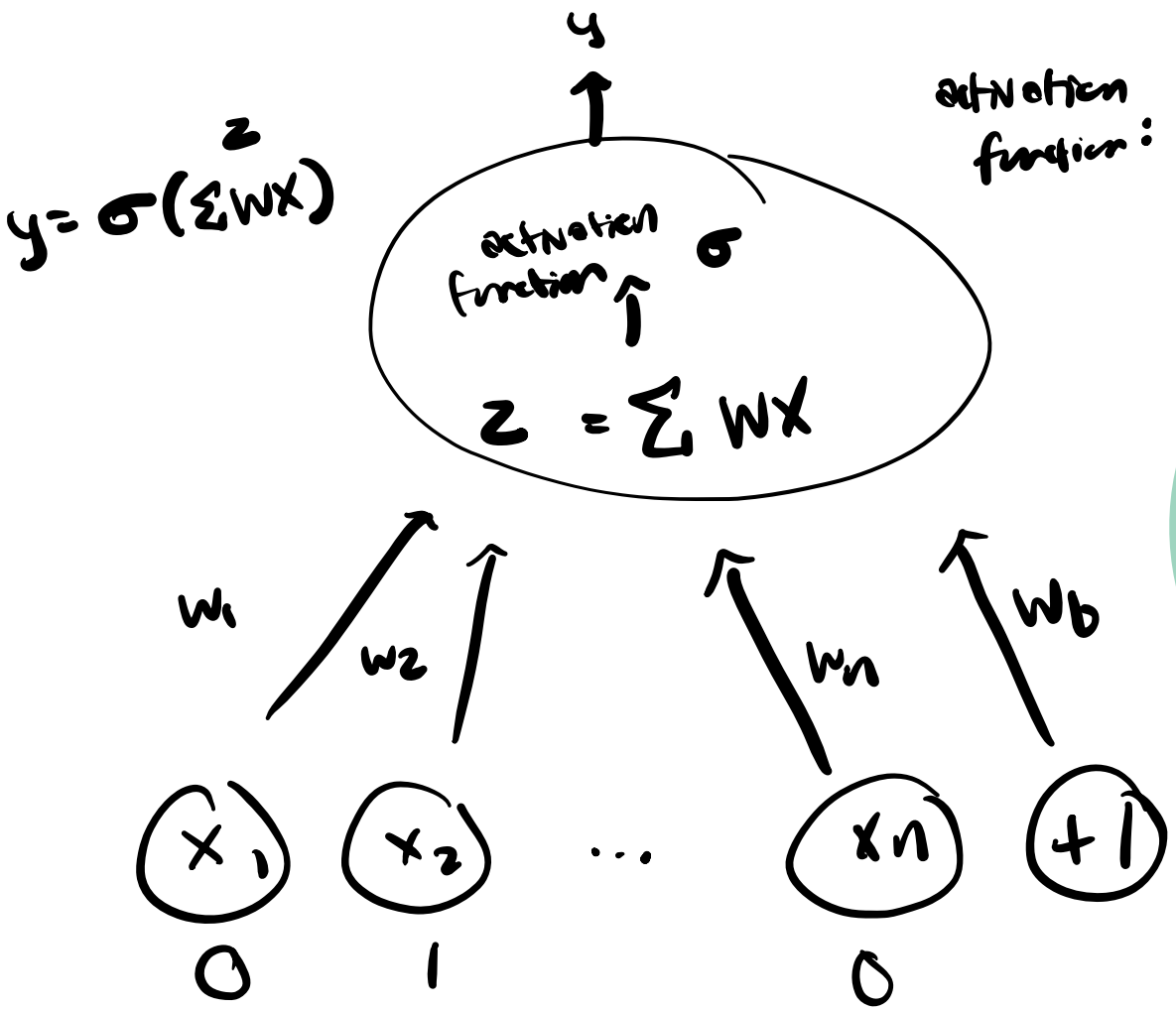
CS 232: Artificial Intelligence

Spring 2024

Prof. Carolyn Anderson
Wellesley College

Recap

Neural Network Unit



parameters

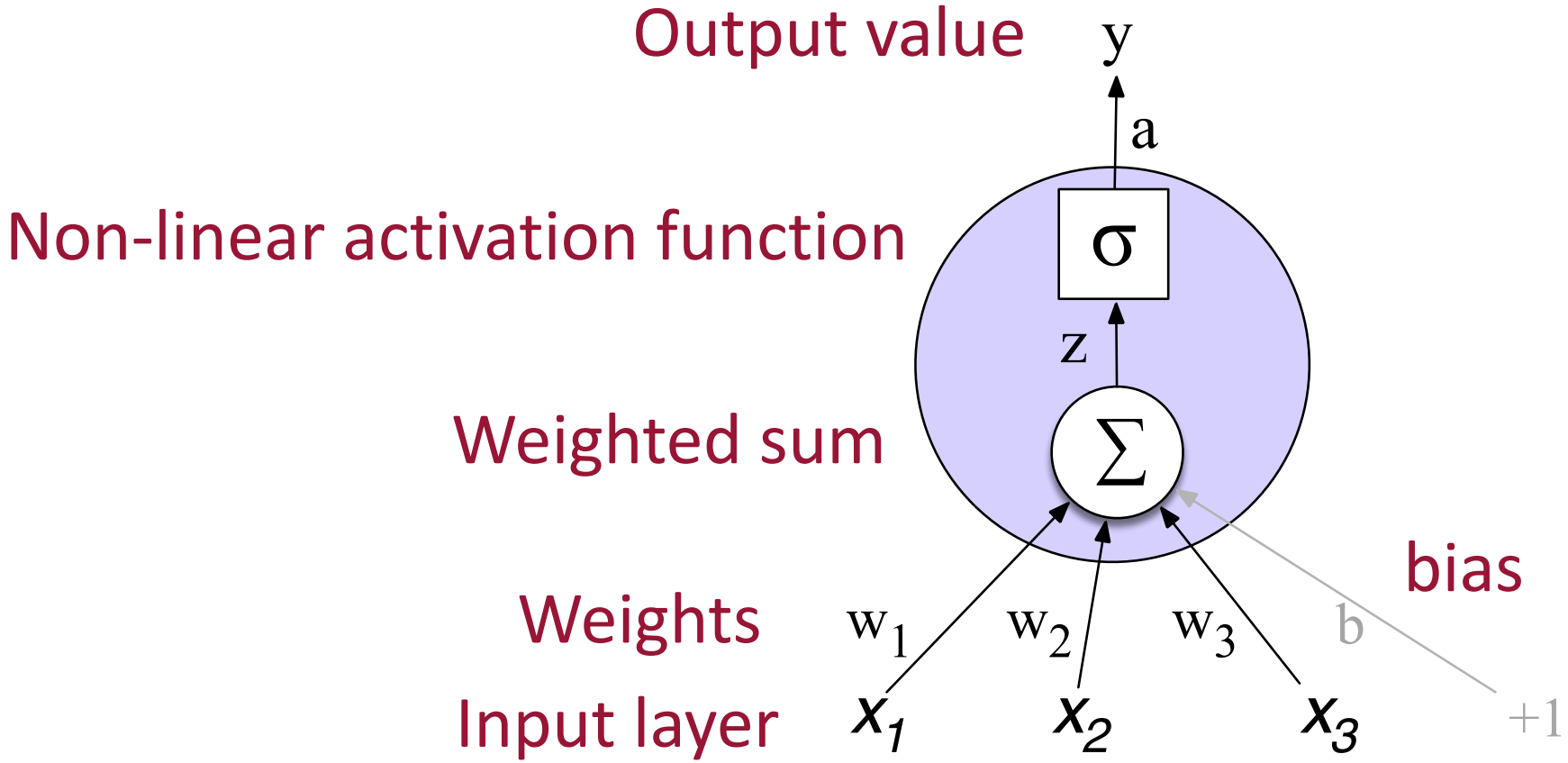
$\theta = W$

activation function: sigmoid, softmax, tanh, relu

$= W, b$

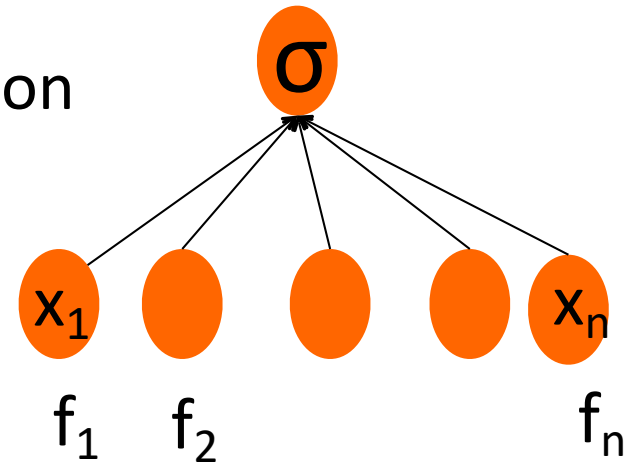
????

Neural Network Unit

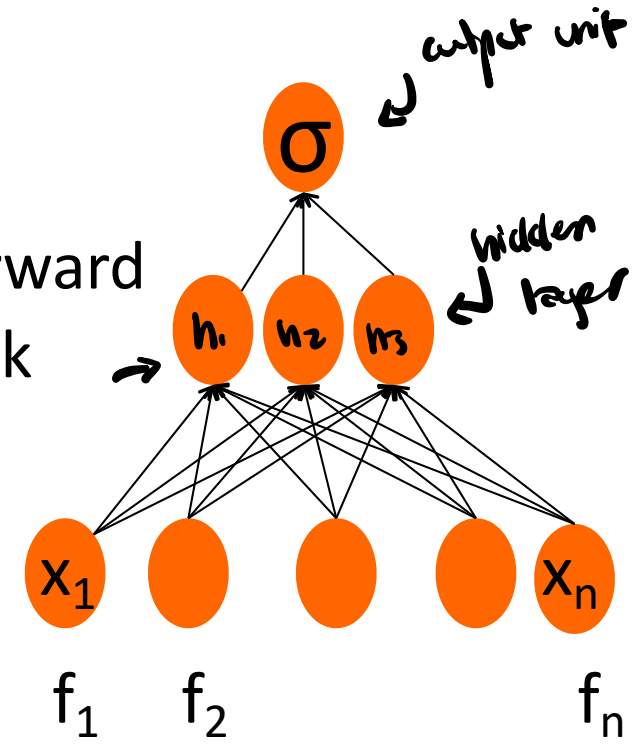


Feedforward nets for simple classification

Logistic Regression



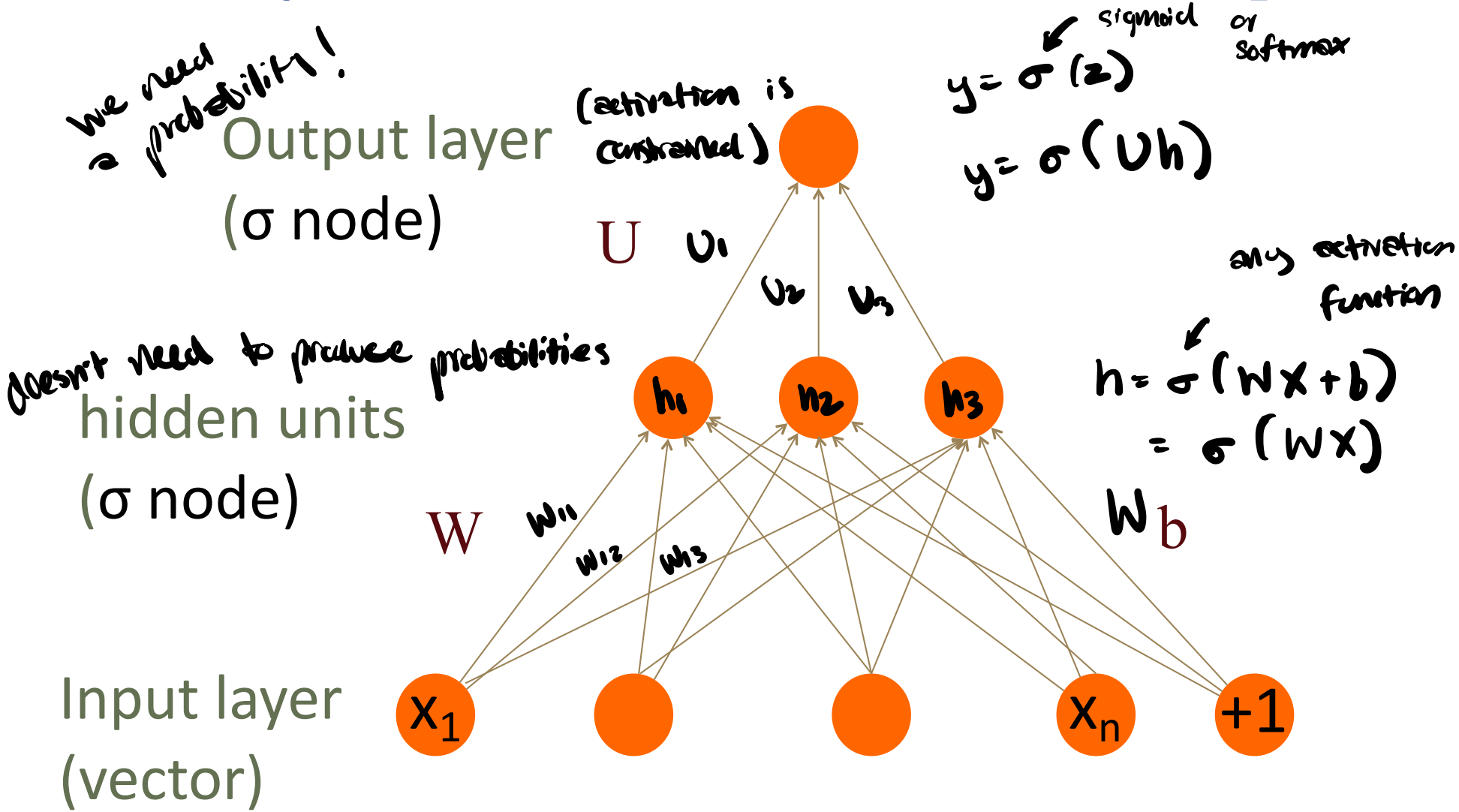
2-layer feedforward network



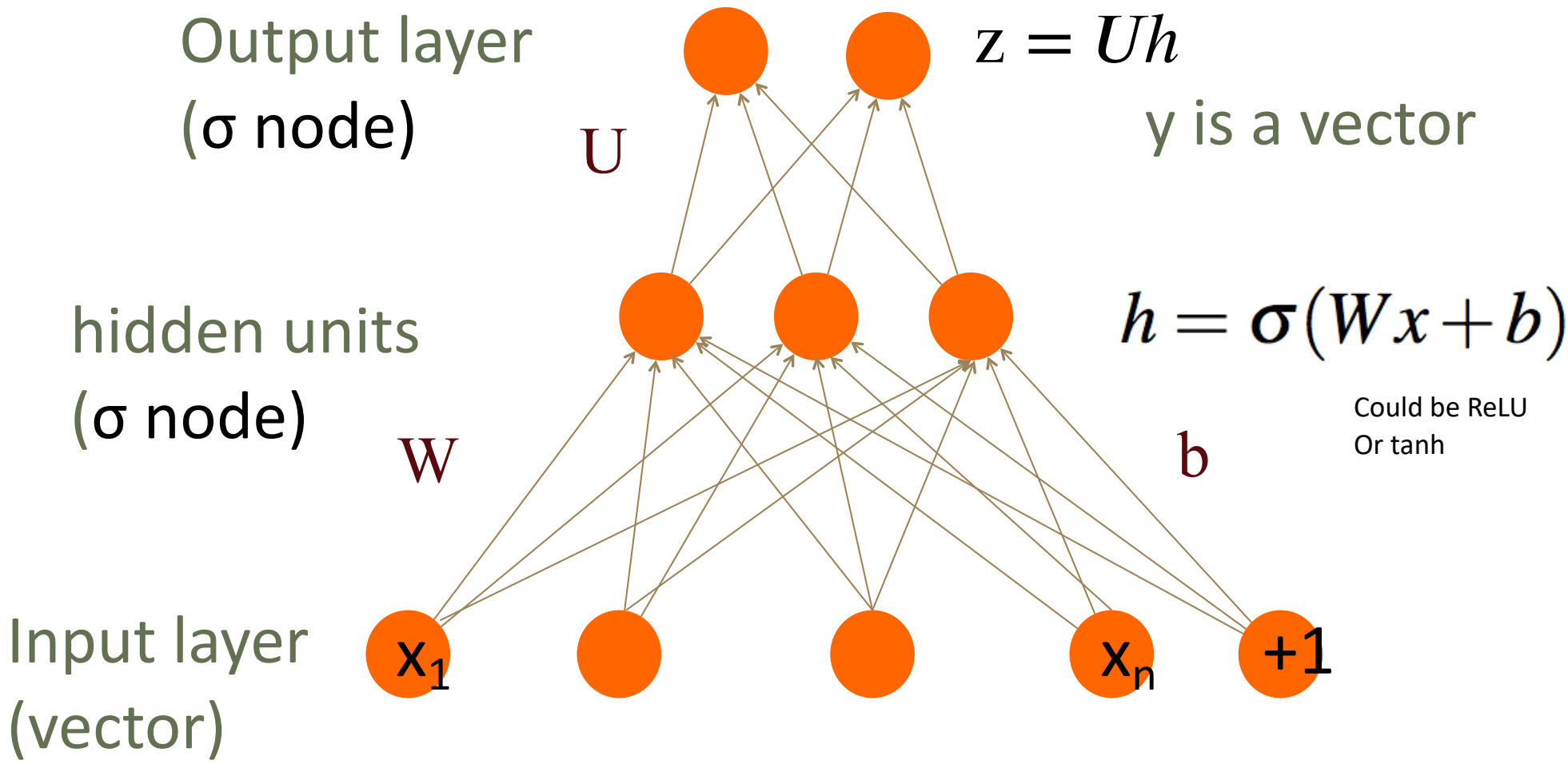
Just add a hidden layer to logistic regression!

This allows the network to use non-linear interactions between features (which *hopefully* improves performance).

Two-Layer Network with scalar output

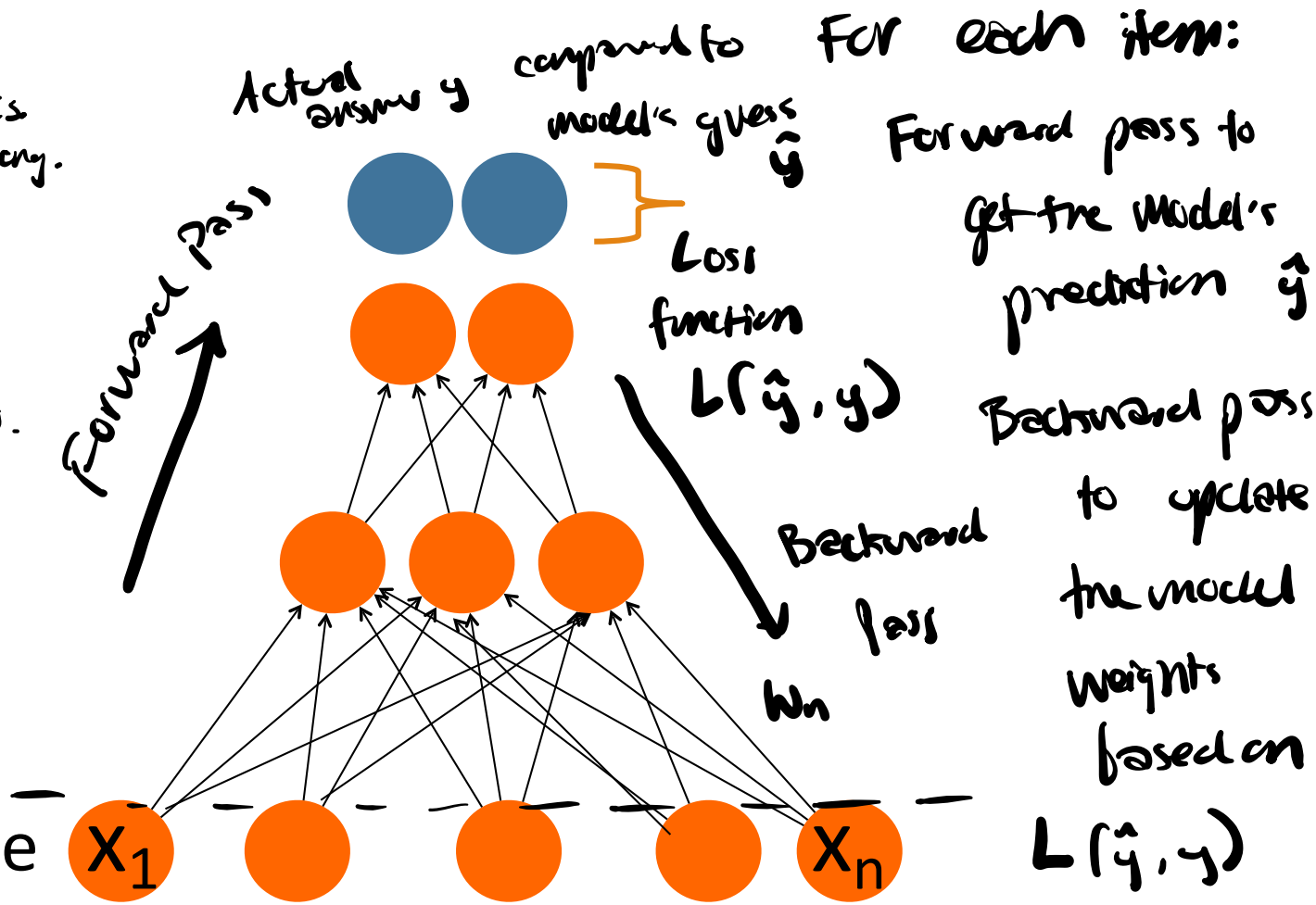


Two-Layer Network with softmax output



Intuition: training a 2-layer Network

1. Training
   Take a training set.
   Make the model guess labels
   Update the weights if wrong.
2. Testing
   Take a test set.
   Use the model to
   make new predictions.



Intuition: Training a 2-layer network

For every training tuple (x, y)

- Run *forward* computation to find our estimate $\hat{y}$
- Run *backward* computation to update weights:
  - For every output node
    - Compute loss L between true y and the estimated $\hat{y}$
    - For every weight w from hidden layer to the output layer
      - Update the weight
  - For every hidden node
    - Assess how much blame it deserves for the current answer
    - For every weight w from input layer to the hidden layer
      - Update the weight

Gradient descent for weight updates

The derivative of the loss function with respect to weights tells us how to adjust the weights to make better predictions.

Derivative of the loss function: $\frac{\partial L(f(x;w),y)}{\partial w}$

We want to move the weights in the opposite direction of the gradient:

θ : parameter/weight: something the model learns

hyperparameter: something that we set

$$w_{t+1} = w_t - \gamma \frac{\partial L(f(x;w),y)}{\partial w}$$

↑ learning rate controls how much we update (common settings: 0.0001 or even smaller)

For logistic regression:

$$\frac{\partial L_{CE}(w,b)}{\partial w_j} = (\hat{y} - y)x_j$$

Activity

Mid-Semester Feedback

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

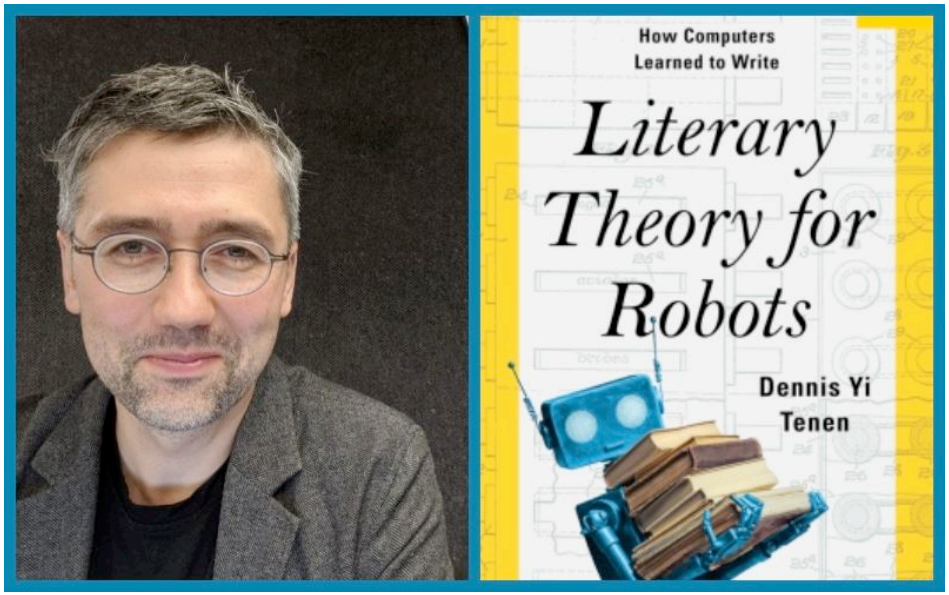
Dr. Rumman Chowdhury

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



WELLESLEY FREE
LIBRARY

Date:	Tuesday, March 26, 2024
Time:	7:00pm - 8:00pm
Time Zone:	Eastern Time - US & Canada (change)
Location:	Online- Zoom
Audience:	Adult
Categories:	Author Talk

https://wellesleyfreelibrary.libcal.com/event/11711295