Applied Machine Learning!!!

W207 Section 9
Rasika Bhalerao
rasikabh@berkeley.edu

Aug 23: Welcome!
Nov 8 and 22: No classes

Schedule

Supervised learning methods

	Sync	Topic
2	Aug 30	Linear Regression / Gradient Descent
3	Sep 6	Feature Engineering Bonus: Naive Bayes
4	Sep 13	Logistic Regression
5	Sep 20	Multiclass classification / Eval Metrics Bonus: Reinforcement learning
6	Sep 27	Neural Networks
7	Oct 4	KNN, Decision Trees, Ensembles

Unsupervised learning methods

	Sync	Topic
8	Oct 11	KMeans and PCA Bonus: LDA
9	Oct 18	Text Embeddings Bonus: Language models
10	Oct 25	CNNs Bonus: GANs
11	Nov 1	EDA, Real data, Baselines
12	Nov 15	Fairness / Ethics
13	Nov 29	Fancy Neural Networks
14	Dec 6	Final Presentations

Assignment Schedule

Due Date	Assignment
Aug 28	HW1
Sep 4	HW2
Sep 11	HW3
Sep 18	HW4
Sep 25	HW5
Oct 2	HW6
Oct 16	Group project baseline
Oct 23	HW8
Nov 6	HW9
Nov 20	HW10
Dec 4	Final project notebook + presentation

Behavior expectations

- Healthy disagreement is expected
- Be mindful of one another’s schedules
- Be a good listener
- Have fun in a professional manner
- Share related real-world experience
- Ask questions when something is confusing
- Keep it 100 but be respectful
- Be open-minded to new ideas in the real world and when coding
- On time for group meetings

How are final projects going?

Guidelines: https://docs.google.com/document/d/1R7mIHOtYXKU8vEQzw10uofb_iK3sgimw8iZLWSTzdgg/edit?usp=sharing

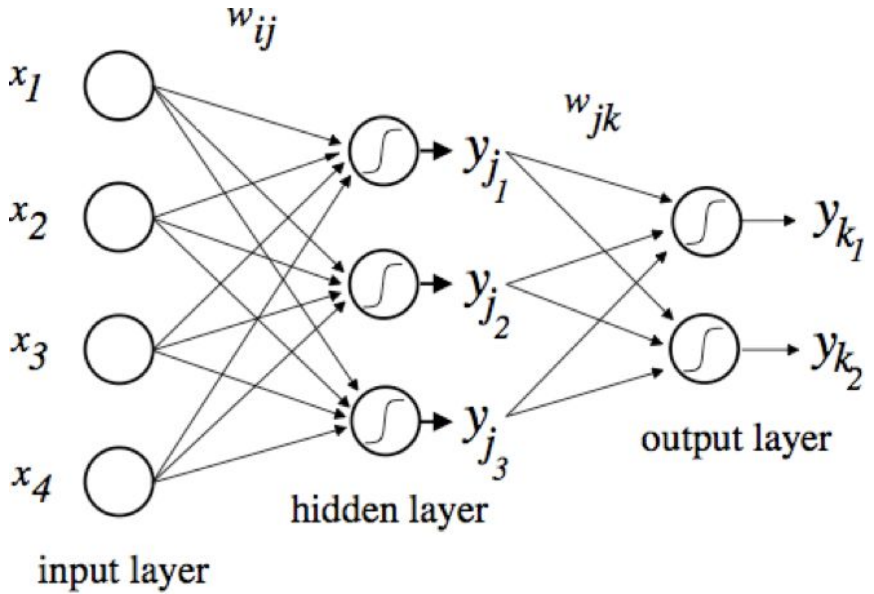
Hyperparameters

- Learning rate
- Batch size
- Epochs

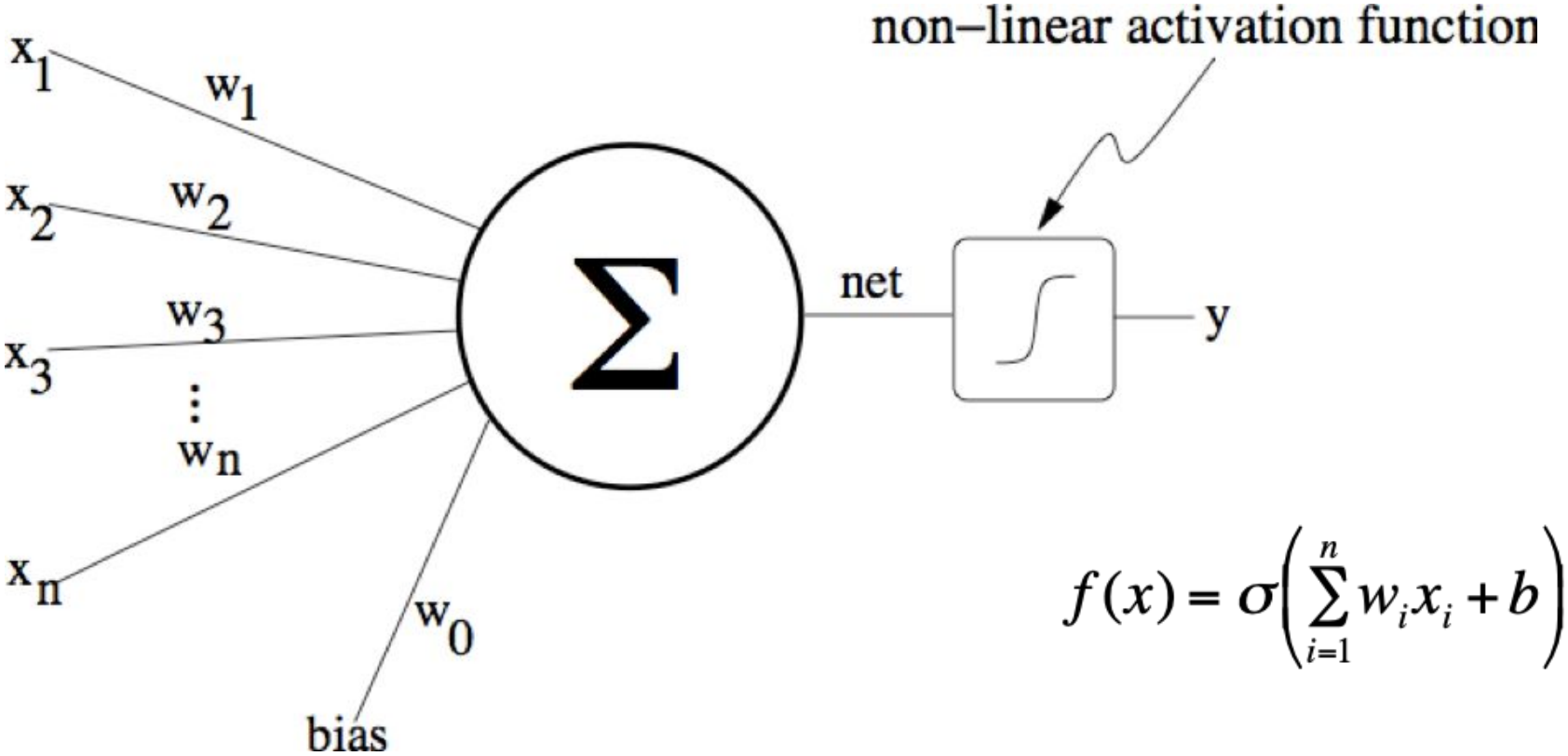
What are they? How do we pick values for them?

Training a neural net: Backpropagation

- **Forward Pass:** present training input to network and calculate output
- **Backward Pass:** calculate error gradient and update weights starting at output layer going backwards

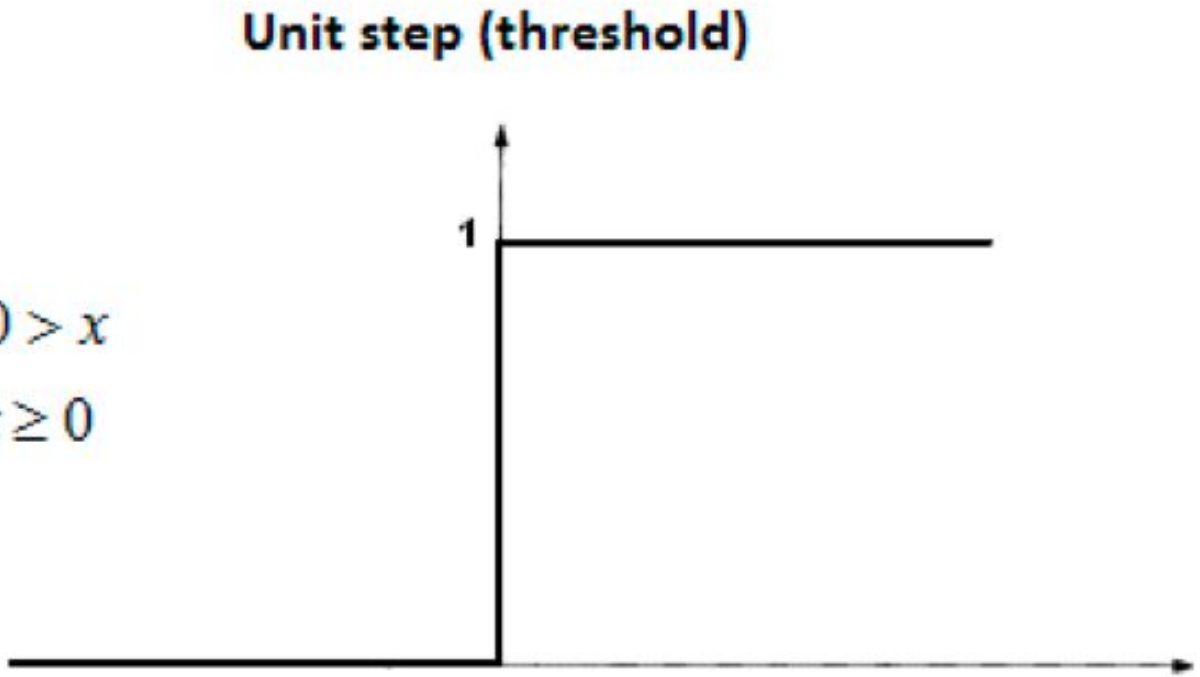


Activation functions

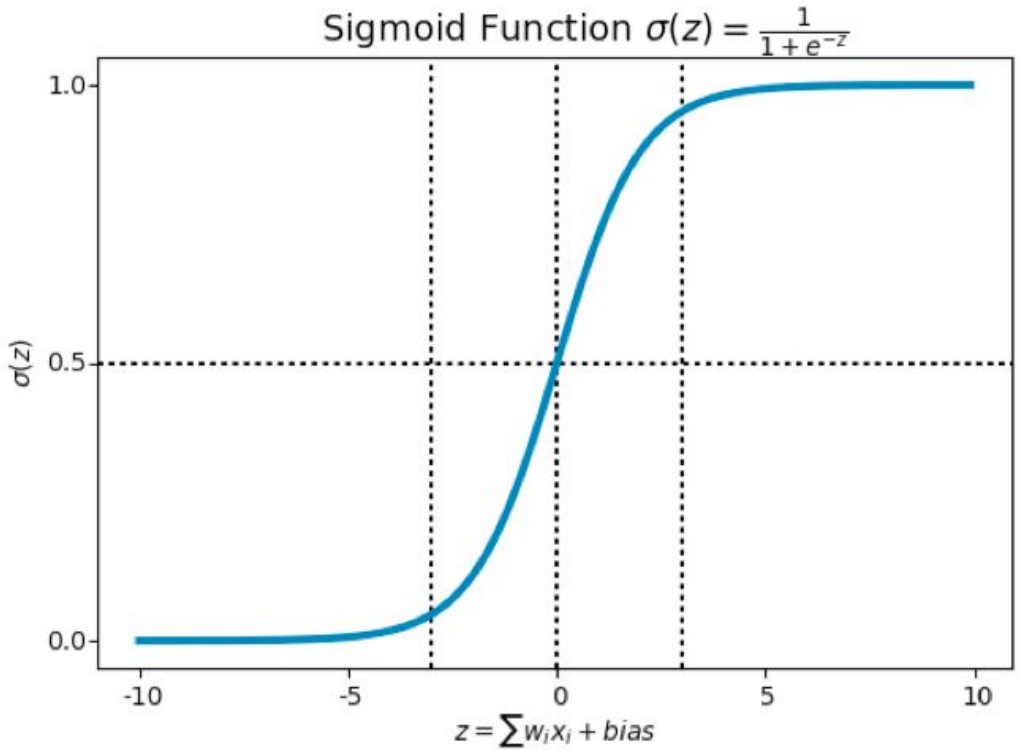


Activation functions

$$f(x) = \begin{cases} 0 \text{ if } 0 > x \\ 1 \text{ if } x \geq 0 \end{cases}$$

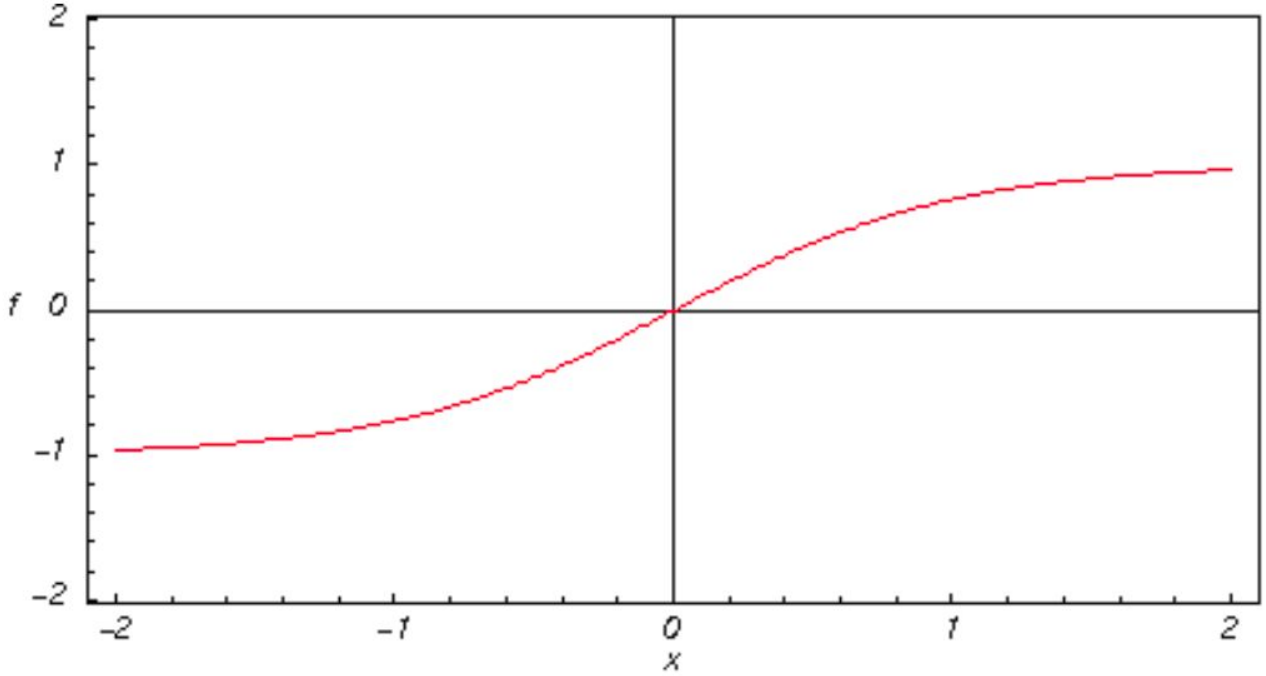


Activation functions



More useful if you don’t just want {0,1} output, and you want numbers between 0 and 1 (for example, if you have a classifier with multiple classes and want to compare numbers for each)

Activation functions

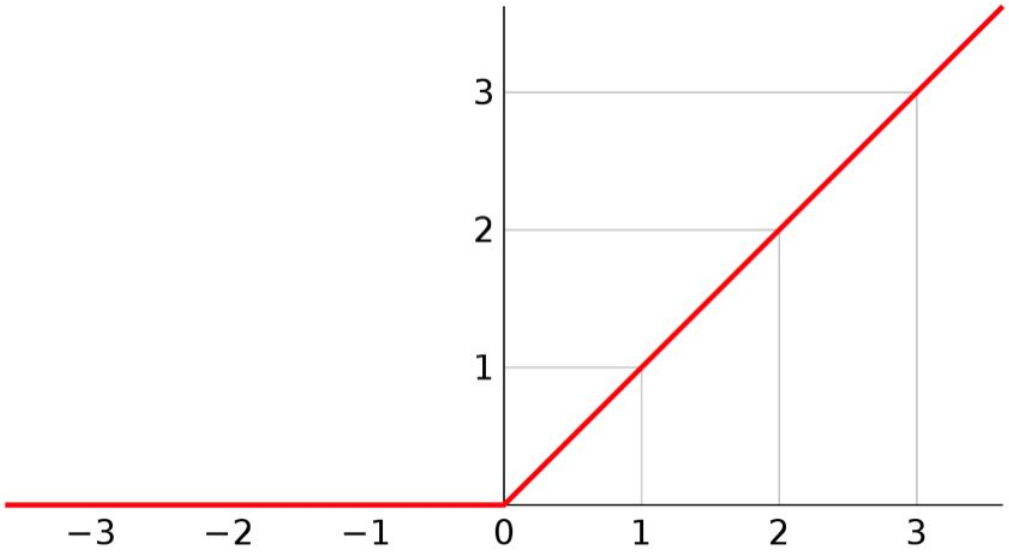


Hyperbolic Tangent

Scaled version of sigmoid. It just has a “stronger” derivative.

Note: $tanh(x) = 2\,sigmoid(2x) - 1$

Activation functions



Rectified Linear Units (ReLu)

Since the range is [0, inf), it can “blow up” the activation.

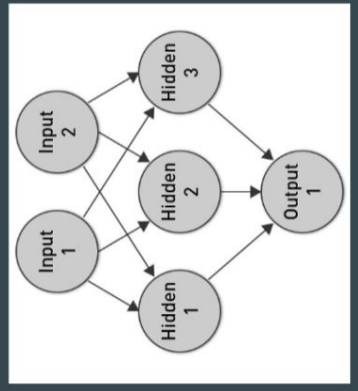
Not as easy for classification as the functions with range [0,1].

Since the derivative in negative section is 0, it can cause neurons to “die” (no update to weights). Solution: make the horizontal line into a very slightly tilted line → “leaky ReLu”

Since negative values → 0, ReLu is less costly and overtrains less.

Goal: Classify digit 0-9 Question: Which ways make sense?

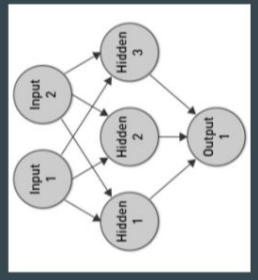
1



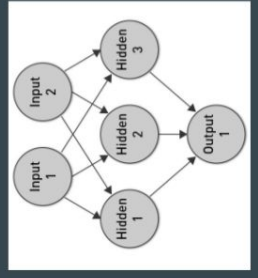
0-9

good or bad?

2

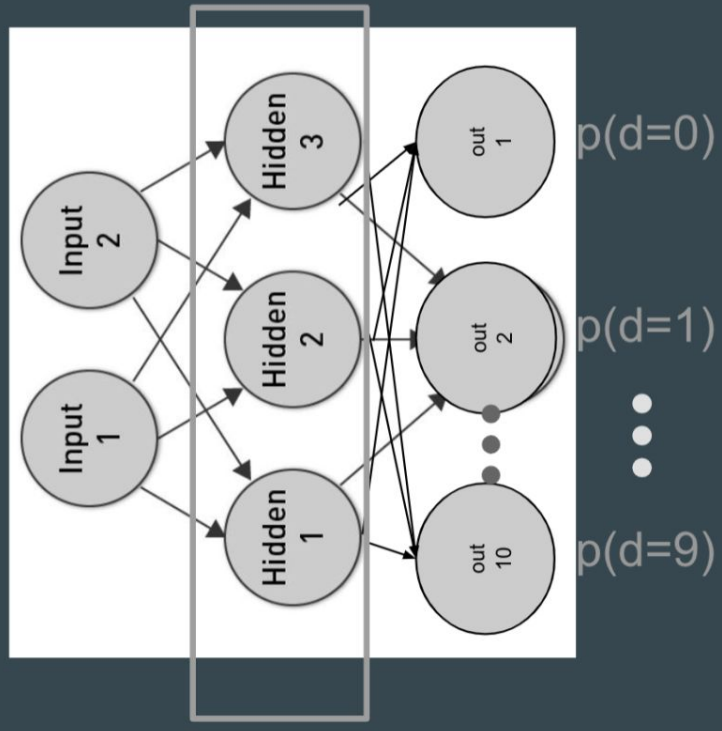


0?



9?

good or bad?



good or bad?

View options (at the top) → Annotate

Async Practice Quiz Questions (vote!)

Our neural network solution to the XOR problem had how many parameters?	9 / Nine	Something else
With inputs X1 and X2, the neurons in the first layer can only learn diagonal lines.	True	False
Computing the gradient of the loss in a neural network involves a partial derivative for just the parameters in the last layer.	True	False
Back propagation allows us to compute a partial derivative for all parameters using the chain rule.	True	False
Neural networks do not need nonlinearities to learn nonlinear functions.	True	False
Both ANNs and brains appear to arrange neurons in layers, with sensitivity to increasingly complex patterns.	True	False

Notebook!

To access later:
https://github.com/MIDS-W207/rasikabh/blob/main/live_sessions/Week6.ipynb

Also, if you want last semester's assignments:
https://github.com/MIDS-W207/coursework