Ch 10.1: Neural Nets

Lecture 29 - CMSE 381

Prof. Mengsen Zhang

Michigan State University
::
Dept of Computational Mathematics, Science & Engineering

Mon, April 7, 2025

Announcements

Last time:

- SVM

This lecture:

- Feed Forward Neural Nets

Announcements:

- Homework #9 due Sunday, April 13
- detailed project content rubrics
- No class wednesday 4/16. Virtual office hours.

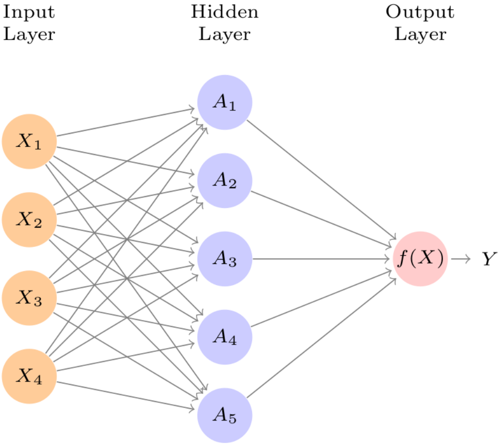
	M	3/17	**Midterm #2**		Sun 3/16
21	W	3/19	Polynomial & Step Functions	7.1-7.2	
22	F	3/21	Step Functions; Basis functions; Start Splines	7.2-7.4	
23	M	3/24	Regression Splines	7.4	
24	W	3/26	Decision Trees	8.1	HW #6 Due Wed 3/26
25	F	3/28	Random Forests	8.2.1, 8.2.2	HW #7 Due Sun 3/30
26	M	3/31	Maximal Margin Classifier	9.1	
27	W	4/2	SVC	9.2	
28	F	4/4	SVM	9.3, 9.4	HW #8 Due Sun 4/6
29	M	4/7	Single Layer NN	10.1	
30	W	4/9	Multi Layer NN	10.2	
31	F	4/11	CNN	10.3	HW #9 Due Sun 4/13
32	M	4/14	Unsupervised learning / clustering	12.1, 12.4	
33	W	4/16	Virtual: Project Office Hours		
	F	4/18	***Review***		
	M	4/21	**Midterm #3**		
	W	4/23			
	F	4/25			**Project Due**

Section 1

Neural Nets

The idea

Feed Forward Neural Network: The cartoon



What is activation? Neuroscience 101.

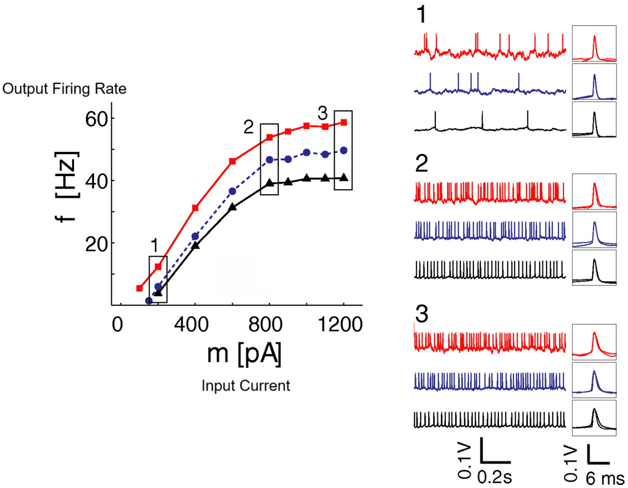
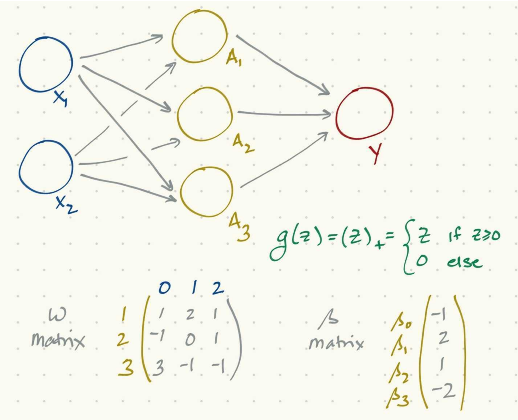


Figure adapted from Fig 1 of Arsiero et al. 2007 (J of Neurosci.)

A very simple example

Computing A_k for $(1, 0)$

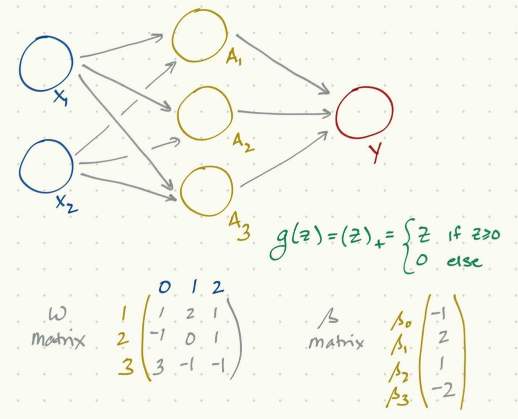
$$A_k = h_k(X) = g(w_{k0} + \sum_{j=1}^{p} w_{kj} X_j),$$



A very simple example

Computing Y for $(1, 0)$

$$f(X) = \beta_0 + \sum_{k=1}^{K} \beta_k A_k$$

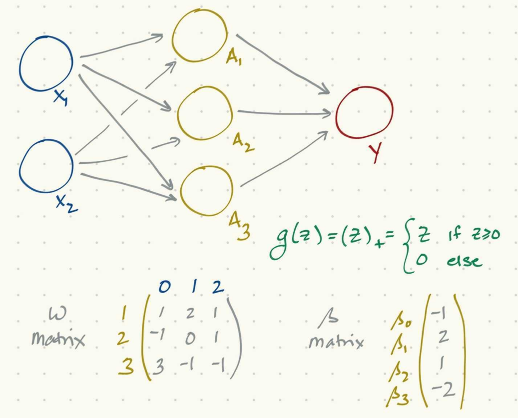


A very simple example

Computing Y for $(0, 1)$

$$A_k = h_k(X) = g(w_{k0} + \sum_{j=1}^{p} w_{kj} X_j),$$

$$f(X) = \beta_0 + \sum_{k=1}^{K} \beta_k A_k$$



A different example

- Draw the diagram for a neural net with input data points with $p = 3$ (i.e., (X_1, X_2, X_3)) and two units in the hidden layer.
- Using the ω and β matrices, what is the output predicted Y for the point $(2, 0, 1)$?

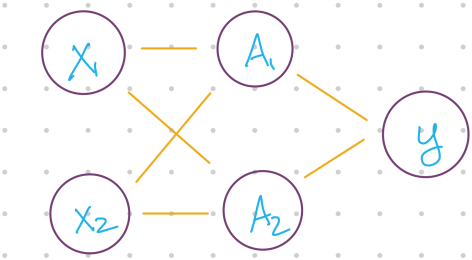
$$\omega = \begin{pmatrix} 1 & 0 & -2 & 2 \\ -3 & 1 & 0 & -1 \end{pmatrix} \qquad \beta = \begin{pmatrix} -1 \\ -2 \\ 1 \end{pmatrix}$$

- Use the activation function

$$g(z) = (z)_+ = \begin{cases} 0 & \text{if } z < 0 \\ z & \text{else.} \end{cases}$$

Extra space

What if there's no activation function?



$$\omega = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \qquad \beta = \begin{pmatrix} -1 \\ -2 \\ 1 \end{pmatrix}$$

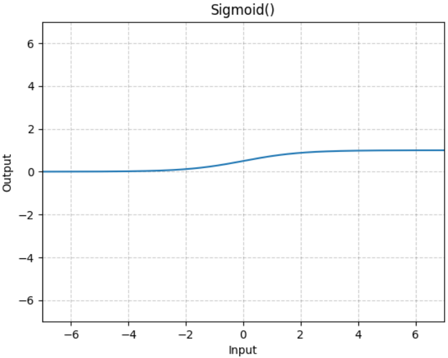
Choices for activation function

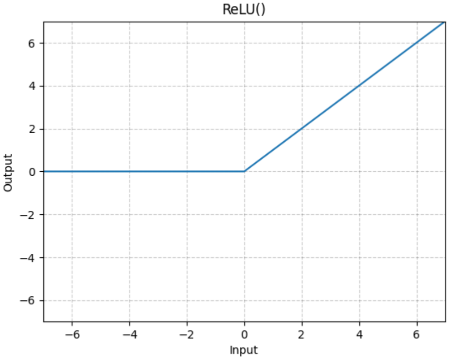
Sigmoid:

$$g(z) = \frac{e^z}{1+e^z} = \frac{1}{1+e^{-z}}$$

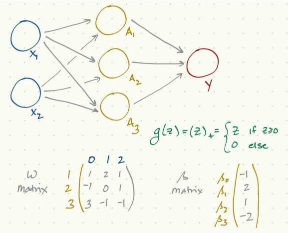
ReLU: Rectified linear unit

$$g(z) = (z)_+ = \begin{cases} 0 & \text{if } z < 0 \\ z & \text{else.} \end{cases}$$



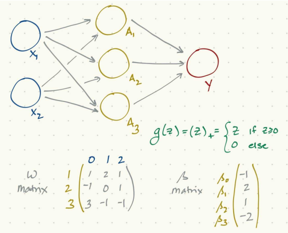


Matrix version: First layer



$$A_k = h_k(X) = g(w_{k0} + \sum_{j=1}^{p} w_{kj} X_j),$$

$$A = g(\mathbf{W} \cdot \mathbf{X}) \qquad \mathbf{X}^T = (1 \; X_1 \; X_2 \; \cdots \; X_p)$$



$$f(X) = \beta_0 + \sum_{k=1}^{K} \beta_k A_k$$

$$Y = \beta \cdot \mathbf{A} \qquad \mathbf{A}^T = (1 \; A_1 \; A_2 \; \cdots \; A_K)$$

Now what?

Choose parameters by minimizing RSS, $\sum_{i=1}^{n}(y_i - f(x_i))^2$

Chosen in advance:

Tuned by the model:

Coding

Next time

	M	3/17	**Midterm #2**		Sun 3/16
21	W	3/19	Polynomial & Step Functions	7.1-7.2	
22	F	3/21	Step Functions; Basis functions; Start Splines	7.2-7.4	
23	M	3/24	Regression Splines	7.4	
24	W	3/26	Decision Trees	8.1	HW #6 Due Wed 3/26
25	F	3/28	Random Forests	8.2.1, 8.2.2	HW #7 Due Sun 3/30
26	M	3/31	Maximal Margin Classifier	9.1	
27	W	4/2	SVC	9.2	
28	F	4/4	SVM	9.3, 9.4	HW #8 Due Sun 4/6
29	M	4/7	Single Layer NN	10.1	
30	W	4/9	Multi Layer NN	10.2	
31	F	4/11	CNN	10.3	HW #9 Due Sun 4/13
32	M	4/14	Unsupervised learning / clustering	12.1, 12.4	
33	W	4/16	Virtual: Project Office Hours		
	F	4/18	***Review***		
	M	4/21	**Midterm #3**		
	W	4/23			
	F	4/25			**Project Due**

Q of the Day:

What is the relationship between Neural Network and Logistic regression?