Ch 10.1: Neural Nets

Lecture 29 - CMSE 381

Prof. Elizabeth Munch

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

Mon, Nov 18, 2024

Announcements

Last time:

- SVM

This lecture:

- Feed Forward Neural Nets

Announcements:

- Homework #9 due Sunday, Nov 24
- No class wednesday Nov 27. Virtual office hours by request.

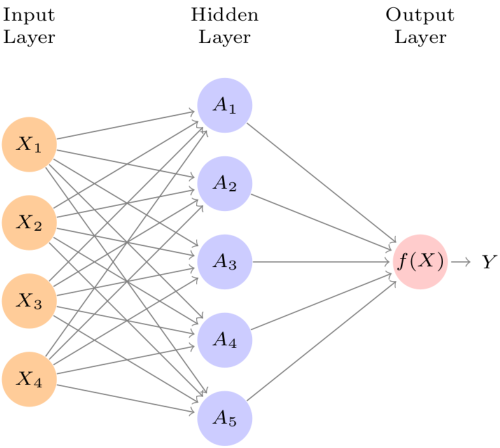
Lec #	Date			Reading	HW
21	Mon	10/28	Polynomial & Step Functions	7.1,7.2	
22	Wed	10/30	Step Functions; Basis functions; Start Splines	7.2 - 7.4	
23	Fri	11/1	Regression Splines	7.4	HW #6 Due Sun 11/3
24	Mon	11/4	Decision Trees	8.1	
25	Wed	11/6	Class Cancelled (Dr Munch out of town)		
26	Fri	11/8	Random Forests	8.2.1, 8.2.2	HW #7 Due Sun 11/10
27	Mon	11/11	Maximal Margin Classifier	9.1	
28	Wed	11/13	SVC	9.2	
29	Fri	11/15	SVM	9.3, 9.4	HW #8 Due Sun 11/17
30	Mon	11/18	Single layer NN	10.1	
31	Wed	11/20	Multi Layer NN	10.2	
32	Fri	11/22	CNN	10.3	HW #11 Due Sun 11/24
33	Mon	11/25	TBD: Unsupervised learning/clustering	12.1, 12.4?	
	Wed	11/27	Virtual: Project office hours		
	Fri	11/29	No class - Thanksgiving		
	Mon	12/2	Review		
	Wed	12/4	**Midterm #3**		
	Fri	12/6	No class - EGR Design Day		Project due

Section 1

Neural Nets

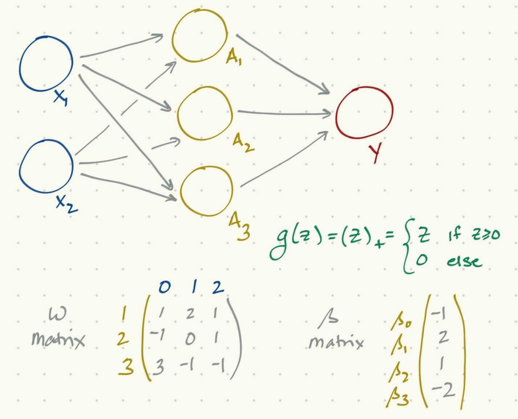
The idea

Feed Forward Neural Network: The cartoon



Computing A_k for $(1, 0)$

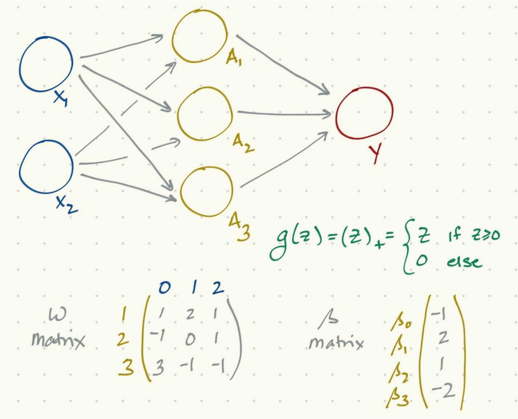
$$A_k = h_k(X) = g(w_{k0} + \sum_{j=1}^{p} w_{kj} X_j),$$



A very simple example

Computing Y for $(1, 0)$

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

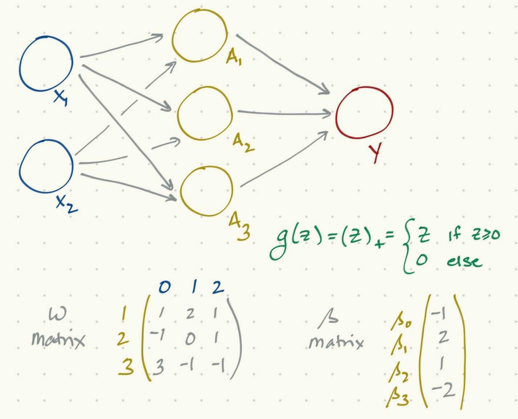


A very simple example

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 $(2, 0, 1)$?

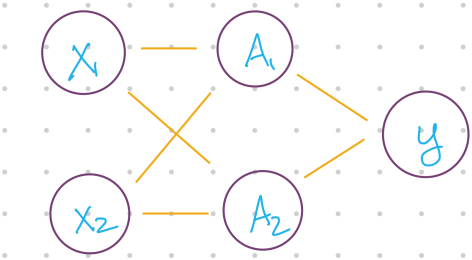
$$\omega = \begin{pmatrix} 1 & 0 & -2 & 2 \\ -3 & 1 & 0 & -1 \end{pmatrix} \qquad \beta = \begin{pmatrix} -1 \\ -2 \\ 1 \end{pmatrix}$$

- 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

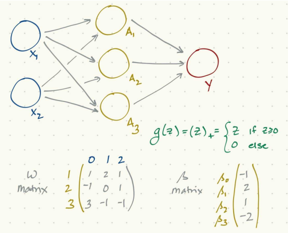
Sigmoid:

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

ReLU: Rectified linear unit

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

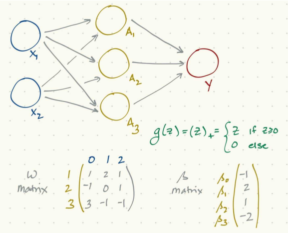
Matrix version: First layer



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

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

Matrix version: Output



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

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

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

Lec #	Date			Reading	HW
21	Mon	10/28	Polynomial & Step Functions	7.1,7.2	
22	Wed	10/30	Step Functions; Basis functions; Start Splines	7.2 - 7.4	
23	Fri	11/1	Regression Splines	7.4	HW #6 Due Sun 11/3
24	Mon	11/4	Decision Trees	8.1	
25	Wed	11/6	Class Cancelled (Dr Munch out of town)		
26	Fri	11/8	Random Forests	8.2.1, 8.2.2	HW #7 Due Sun 11/10
27	Mon	11/11	Maximal Margin Classifier	9.1	
28	Wed	11/13	SVC	9.2	
29	Fri	11/15	SVM	9.3, 9.4	HW #8 Due Sun 11/17
30	Mon	11/18	Single layer NN	10.1	
31	Wed	11/20	Multi Layer NN	10.2	
32	Fri	11/22	CNN	10.3	HW #11 Due Sun 11/24
33	Mon	11/25	TBD: Unsupervised learning/clustering	12.1, 12.4?	
	Wed	11/27	Virtual: Project office hours		
	Fri	11/29	No class - Thanksgiving		
	Mon	12/2	Review		
	Wed	12/4	**Midterm #3**		
	Fri	12/6	No class - EGR Design Day		Project due