Ch 10.1: Neural Nets

Lecture 30 - CMSE 381

Prof. Elizabeth Munch

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

Mon, Nov 20, 2023

Announcements

Last time:

- SVM

This lecture:

- Feed Forward Neural Nets

Announcements:

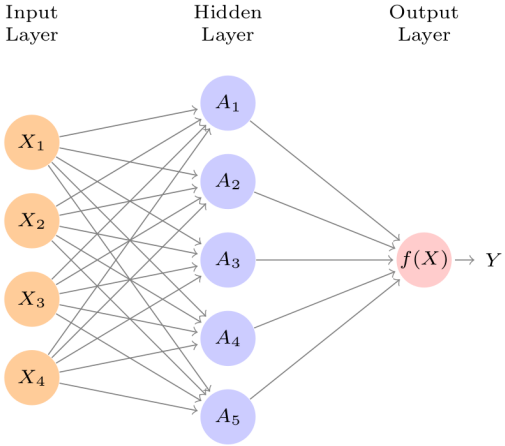
- Homework #7 due tonight
- No class wednesday. Virtual office hours by request.

Section 1

Neural Nets

The idea

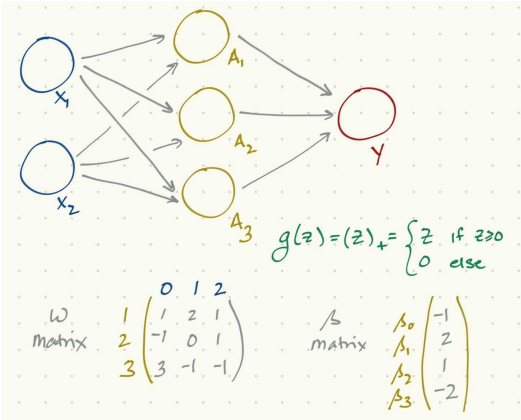
Feed Forward Neural Network: The cartoon



A very simple example

Computing A_k for $(1, 0)$

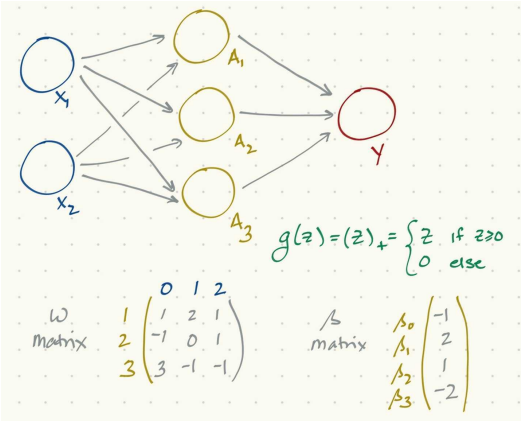
$$A_k = h_k(X) = g(w_{k0} + \sum_{j=1}^{p} w_{kj} X_j),$$



A very simple example

Computing Y for $(1, 0)$

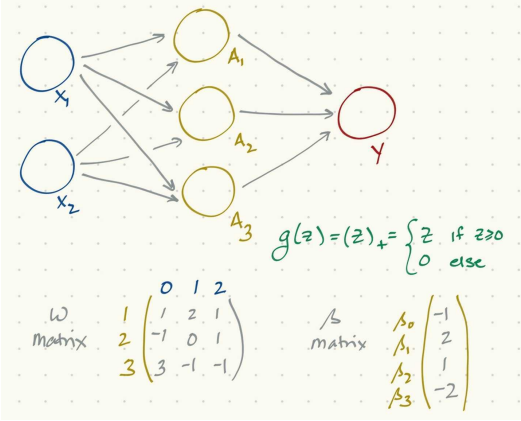
$$f(X) = \beta_0 + \sum_{k=1}^{K} \beta_k A_k$$



A very simple example

Computing Y for $(0, 1)$

$$A_k = h_k(X) = g(w_{k0} + \sum_{j=1}^{p} w_{kj} X_j), \qquad f(X) = \beta_0 + \sum_{k=1}^{K} \beta_k A_k$$



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 $(2, 0, 1)$?

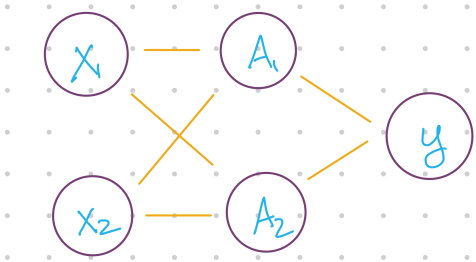
$$\omega = \begin{pmatrix} 1 & 0 & -2 & 2 \\ -3 & 1 & 0 & -1 \end{pmatrix} \qquad \beta = \begin{pmatrix} -1 \\ -2 \\ 1 \end{pmatrix}$$

- 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

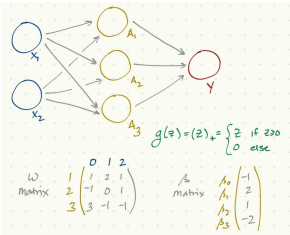
Sigmoid:

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

ReLU: Rectified linear unit

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

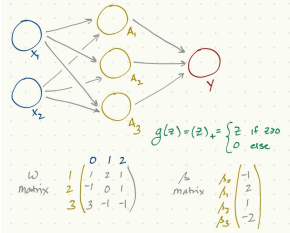
Matrix version: First layer



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

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

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	Homeworks
20	Fri	Oct 27	Dimension Reduction	6.3	
21	Mon	Oct 30	More dimension reduction; High dimensions	6.4	
22	Wed	Nov 1	Polynomial & Step Functions	7.1,7.2	
23	Fri	Nov 3	Step Functions; Basis functions; Start Splines	7.2 - 7.4	
24	Mon	Nov 6	Regression Splines	7.4	~~HW #6 Due~~
25	Wed	Nov 8	Decision Trees	8.1	HW #6 Due
26	Fri	Nov 10	Random Forests	8.2.1, 8.2.2	
27	Mon	Nov 13	Maximal Margin Classifier	9.1	
28	Wed	Nov 15	SVC	9.2	
29	Fri	Nov 17	SVM	9.3, 9.4	
30	Mon	Nov 20	Single layer NN	10.1	HW #7 Due
31	Wed	Nov 22	Virtual: Project office hours		
	Fri	Nov 24	No class - Thanksgiving		
	Mon	Nov 27	**Review**		
	Wed	Nov 29	**Midterm #3**		
32	Fri	Dec 1	Multi Layer NN	10.2	
33	Mon	Dec 4	CNN	10.3	
34	Wed	Dec 6	Unsupervised Learning & Clustering	12.1, 12.4	
35	Fri	Dec 8	Virtual: Project office hours		Project due