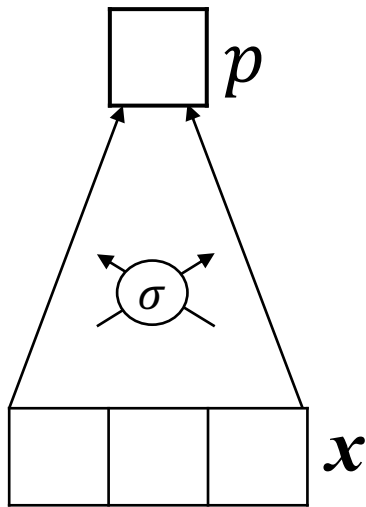


Activity: MLP Structure

ML for Health, Week 3

ANSWER KEY

Logistic Regression

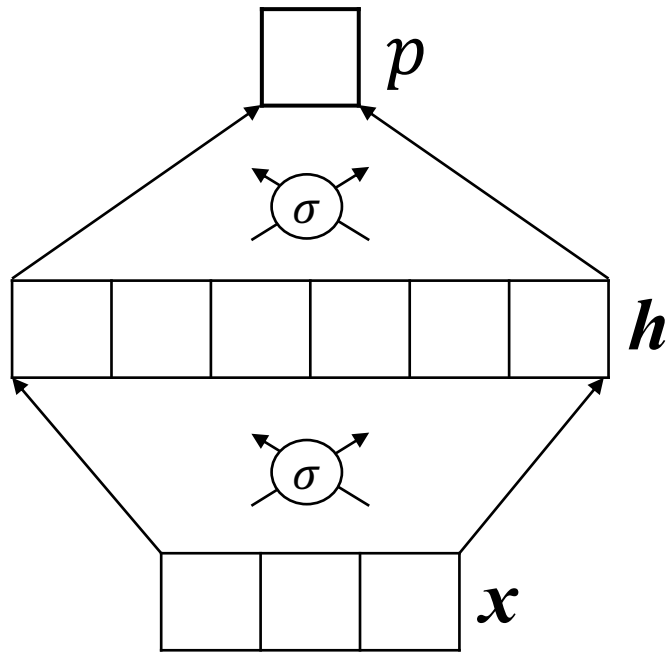


- 0 hidden layers
- $3 + 1 = 4$ parameters
- $p = \sigma(\mathbf{w}^T \mathbf{x} + b)$
- Note that $\mathbf{w}^T \mathbf{x}$ is another way to write the inner product between the vectors \mathbf{w} and \mathbf{x} . The idea here is that we're viewing \mathbf{x} as a 3-by-1 matrix and \mathbf{w}^T , the transpose of \mathbf{w} , as a 1-by-3 matrix.

Shallow MLP

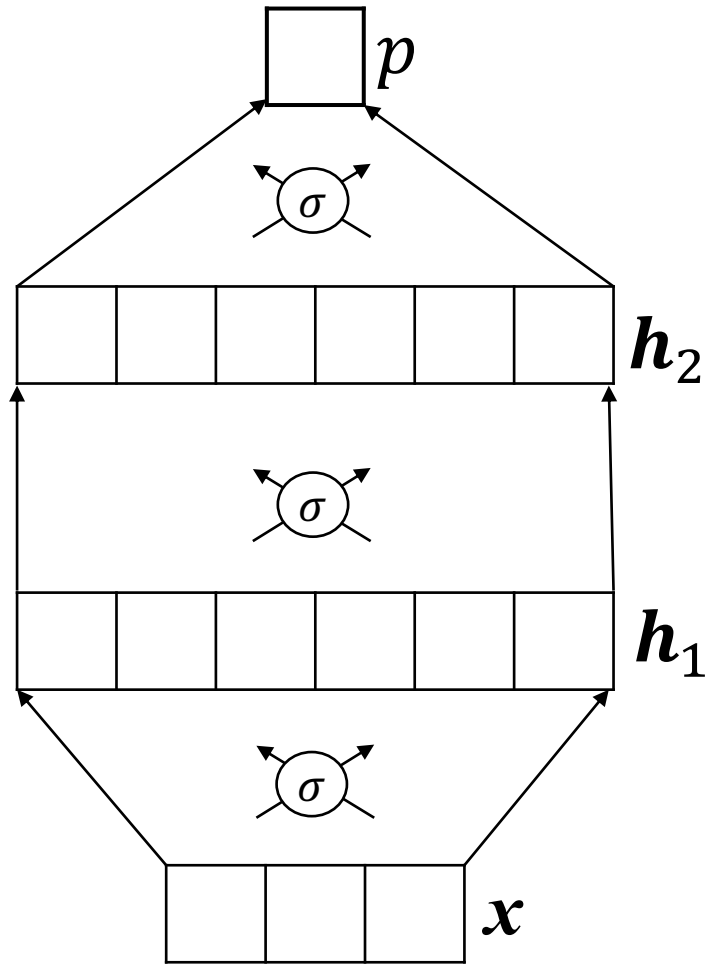
(with σ activations)

- 1 hidden layer of size 6
- $(3 \times 6) + 6 + (6 \times 1) + 1 = 31$ parameters
- $p = \sigma(\mathbf{w}_2^T * \sigma(W_1 \mathbf{x} + \mathbf{b}_1) + b_2)$



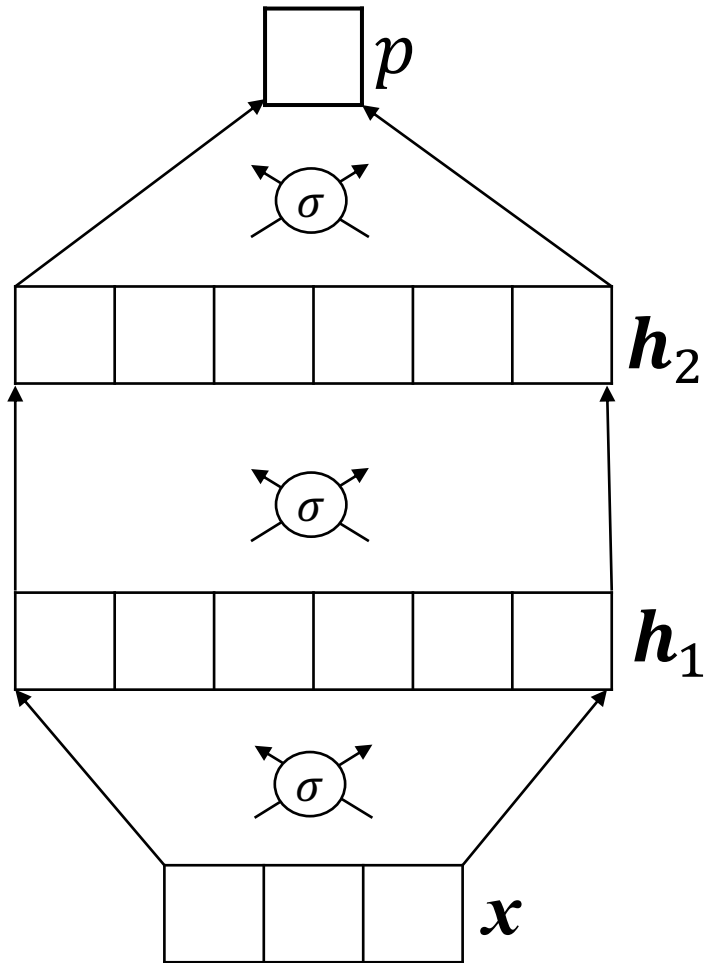
Slightly Deeper MLP

(with σ activations)



- 2 hidden layers, each of size 6
- $(3 \times 6) + 6 + (6 \times 6) + 6 + (6 \times 1) + 1 = 73$ parameters
- $p = \sigma(\mathbf{w}_3^T * \sigma(W_2 * \sigma(W_1 \mathbf{x} + \mathbf{b}_1) + \mathbf{b}_2) + b_3)$

Which model contains more parameters?

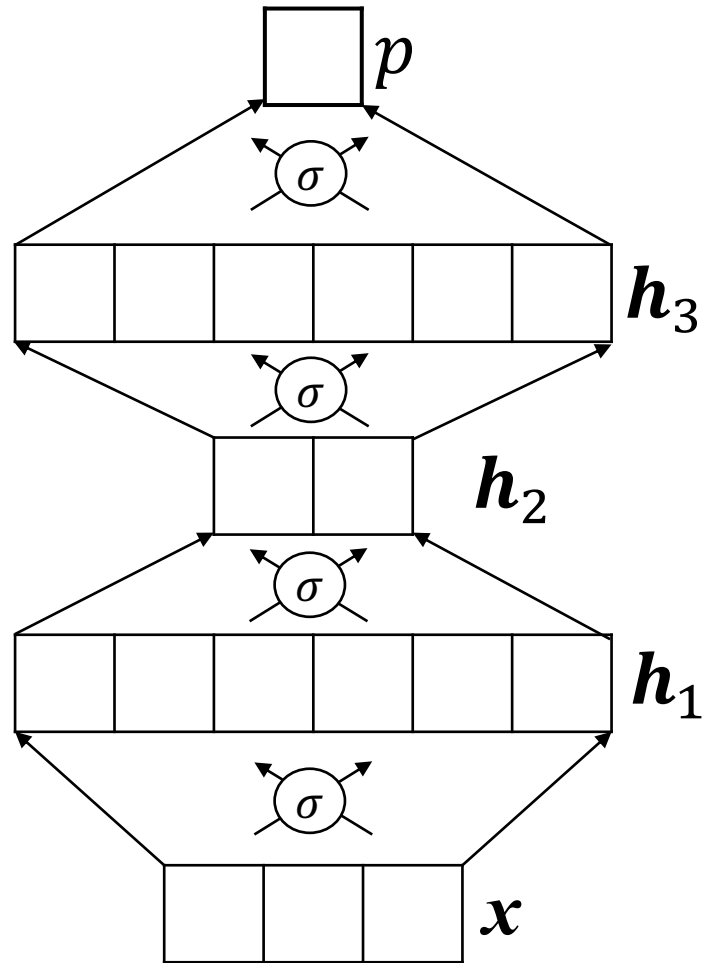


$$(6 \times 1) + 1$$

$$+ (6 \times 6) + 6$$

$$+ (3 \times 6) + 6$$

$$= 73 \text{ params}$$



$$(6 \times 1) + 1$$

$$+ (2 \times 6) + 6$$

$$+ (6 \times 2) + 2$$

$$+ (3 \times 6) + 6$$

$$= 63 \text{ params}$$

- Effect of x_1 on p (LogReg): you need to know the value of b_1 , the parameter associated with x_1 ; and since σ is non-linear, you also need to know the value of z
- Effect of $h_{2,1}$ on p (MLP): similar to the previous answer, you need to know the parameter connecting $h_{2,1}$ to z and $p = \sigma(z)$; as well as the value of z
- Effect of x_1 on p (MLP): here you need to know (a) the values of all parameters connecting x_1 to h_1 ; (b) the current value of h_1 ; (c) the values of all parameters connecting h_1 to h_2 ; (d) the current value of h_2 ; (e) the values of all parameters connecting h_2 to z ; and the current value of z . In short, you need to know all current parameters and all current values in all hidden layers, because x_1 exerts an effect on p through all possible paths connecting x_1 to p in the graph at right. In contrast, in the graph at left, there is only one path connecting x_1 to p .

