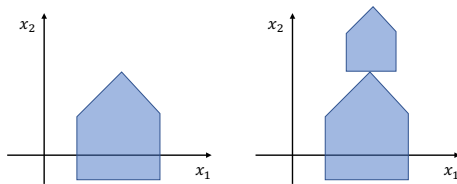


1 Decision Space

Let's further the intuition about how we can compose arbitrarily complex decision boundaries with a neural network. Consider the images below. For each one, build a network of units with a single output that fires if the input is in the shaded area.



Take-away: MLPs can capture any classification boundary. MLPs are universal classifiers. Note that we haven't said anything yet about their ability to generalize.

2 Backprop in Practice: Staged Computation

For the function $f(x, y, z) = (x + y)z$:

- Decompose f into two simpler functions.
- Draw the network that represents the computation of f .
- Write the forward pass and backward pass (backpropagation) in the network.
- Update your network drawing with the intermediate values in the forward and backward pass. Use the inputs $x = -2$, $y = 5$, and $z = -4$.

3 Backpropagation Practice

- Chain rule of multiple variables: Assume that you have a function given by $f(x_1, x_2, \dots, x_n)$, and that $g_i(w) = x_i$ for a scalar variable w . How would you compute $\frac{d}{dw}f(g_1(w), g_2(w), \dots, g_n(w))$? What is its computation graph?
- Let $Z = XW + \mathbf{1}b$, where $Z \in \mathbb{R}^{d_n \times d_{out}}$, $X \in \mathbb{R}^{d_n \times d_{in}}$, $W \in \mathbb{R}^{d_{in} \times d_{out}}$, b is a row vector in $\mathbb{R}^{d_{out}}$, and $\mathbf{1}$ is a column vector in $\mathbb{R}^{d_{in}}$. Given $\frac{\partial L}{\partial Z} \in \mathbb{R}^{d_n \times d_{out}}$, where l is a scalar loss, calculate $\frac{\partial L}{\partial W}$ and $\frac{\partial L}{\partial b}$.

4 Model Intuition

- (a) What can go wrong if you just initialize all the weights in a neural network to exactly zero? What about to the same nonzero value?
- (b) Adding nodes in the hidden layer gives the neural network more approximation ability, because you are adding more parameters. How many weight parameters are there in a neural network with architecture specified by $d = [d^{(0)}, d^{(1)}, \dots, d^{(N)}]$, a vector giving the number of nodes in each of the N layers? Evaluate your formula for a 2 hidden layer network with 10 nodes in each hidden layer, an input of size 8, and an output of size 3.
- (c) Consider the two networks in the image below, where the added layer in going from Network A to Network B has 10 units with linear activation. Give one advantage of Network A over Network B, and one advantage of Network B over Network A.

