- Name three popular activation functions. Can you draw them?
- Suppose you have an MLP composed of one input layer with 10 passthrough neurons, followed by one hidden layer with 50 artificial neurons, and finally one output layer with 3 artificial neurons. All artificial neurons use the ReLU activation function

  1. PWhat is the shape of the input matrix $\mathbf{X}$?
  2. What are the shapes of the hidden layer's weight matrix $\mathbf{W}h$ and bias vector $\mathbf{b}h$?
  3. What are the shapes of the output layer's weight matrix $\mathbf{W}o$ and bias vector $\mathbf{b}o$?
  4. What is the shape of the network's output matrix $\mathbf{Y}$?
  5. Write the equation that computes the network's output matrix $\mathbf{Y}$ as a function of $\mathbf{X}$, $\mathbf{W}h$, $\mathbf{b}h$, $\mathbf{W}o$, and $\mathbf{b}o$.

- How many neurons do you need in the output layer if you want to classify email into spam or ham? What activation function should you use in the output layer? If instead you want to tackle MNIST, how many neurons do you need in the output layer, and which activation function should you use? What about for getting your network to predict housing prices.