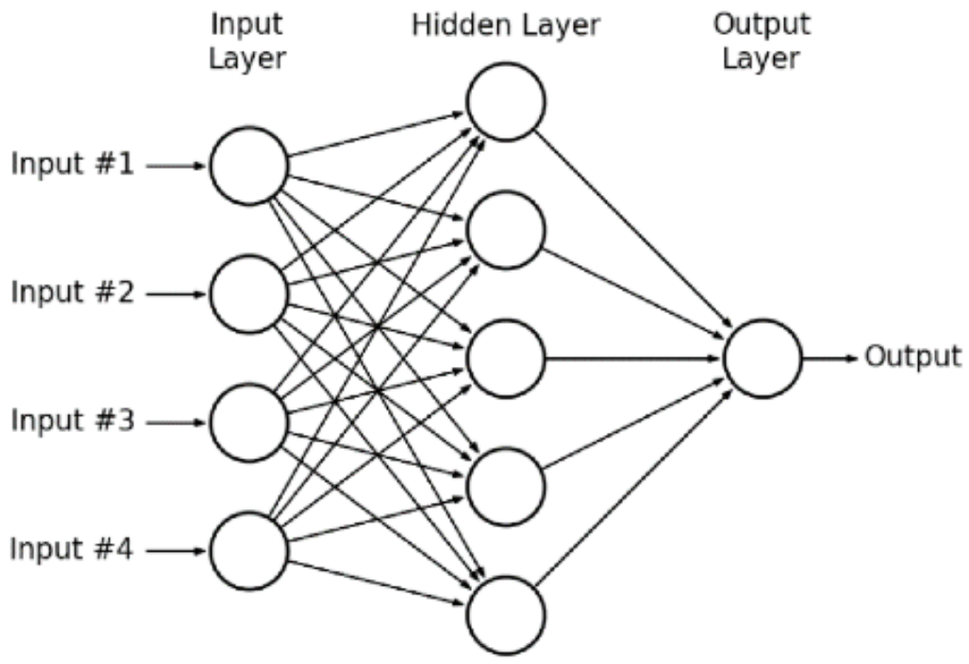
Multilayer Perceptron



The Multilayer Perceptron (MLP) stacks Perceptrons in sequential layers, feeding the outputs of one layer to the inputs of the next layer. In the case of a 2 layer MLP,

$$\hat{y} = f_2\left(w_0^2 + w_1^2 h_1 + \ldots + w_{m_h}^2 h_{m_h}\right)$$

$$h = f_1\left(w_0^1 + w_1^1 x_1 + \ldots + w_m^1 x_m\right)$$

Training an MLP

Gradient descent on the perceptron

- Compute the output, $\hat{y}$
- Compute the gradient of the loss function wrt the parameters, $\frac{\partial l}{\partial w}$
- Update parameters, $w \leftarrow w - \eta \frac{\partial l}{\partial w}$
- Repeat until convergence