



Georgetown
University

Refresher: Neural Networks

Artificial neural networks

- Feedforward NN with one hidden layer:

$$\hat{\mathbf{y}} = \varphi(\mathbf{W}^{(2)}\sigma^{(1)}(\mathbf{W}^{(1)}\mathbf{x} + \mathbf{b}^{(1)}) + \mathbf{b}^{(2)}) = \varphi\left(\sum_{k=1}^K W_{kl}^{(2)} \sigma^{(1)}\left(\sum_{j=1}^J W_{jk}^{(1)} x_j + b_j^{(1)}\right) + b_k^{(2)}\right)$$

\mathbf{x} = input layer

$\hat{\mathbf{y}}$ = output prediction layer

θ = parameters to estimate = $\{\mathbf{W}^{(1)}, \mathbf{b}^{(1)}, \mathbf{W}^{(2)}, \mathbf{b}^{(2)}\}$

$$\sigma(\mathbf{z}) = \begin{cases} \max(\mathbf{0}, \mathbf{z}) & \text{relu, defacto standard} \\ (1 + e^{-\mathbf{z}})^{-1} & \text{sigmoid, old school} \\ \text{many} & \text{variations on these and others} \end{cases}$$

$$\varphi(\mathbf{z}) = \begin{cases} h \tan \mathbf{z} & \text{regression} \\ \frac{e^{\mathbf{z}}}{\sum_{\mathbf{z}} e^{\mathbf{z}}} & \text{classification} \end{cases}$$