STATS 769

Neural Networks and Deep Learning

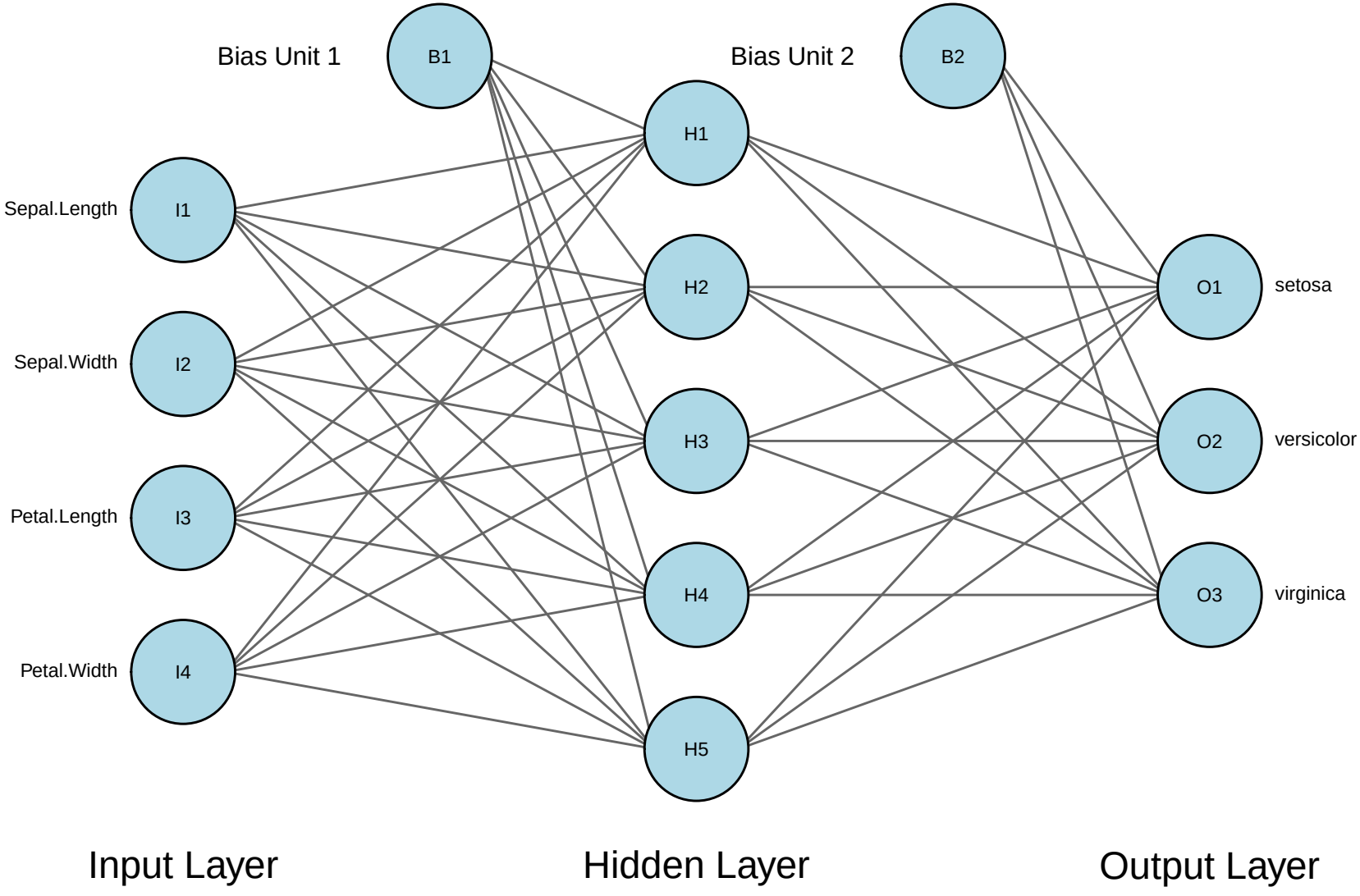
Yong Wang
Department of Statistics
The University of Auckland

2021

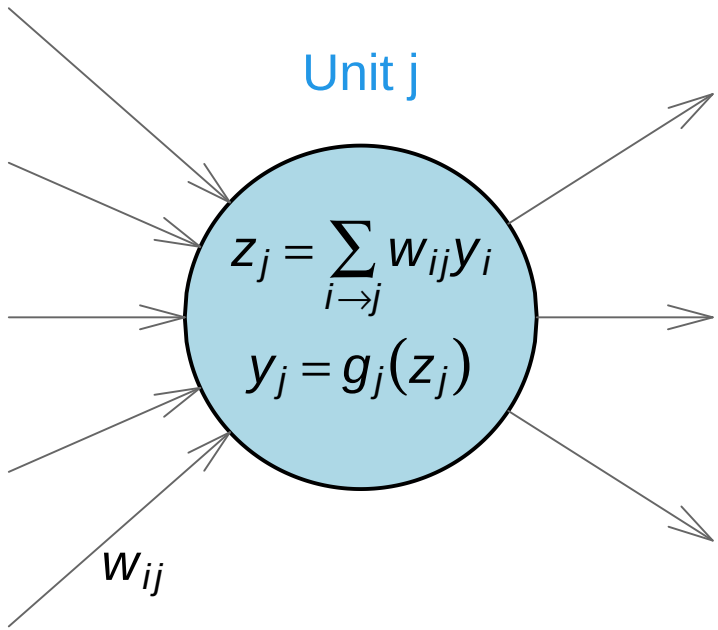
Neural Networks

- An (artificial) neural network mimics how human brains (seemingly) work.
- It takes predictor variables as its inputs and produces a predicted value of Y as its output.
  - For a classification problem, the outputs can first be the predicted posterior probabilities of all classes, from which a class label can be produced.
- It has the form of a network with connected layers of units (or nodes), from the input ones to the output ones.

Single Layer Neural Networks



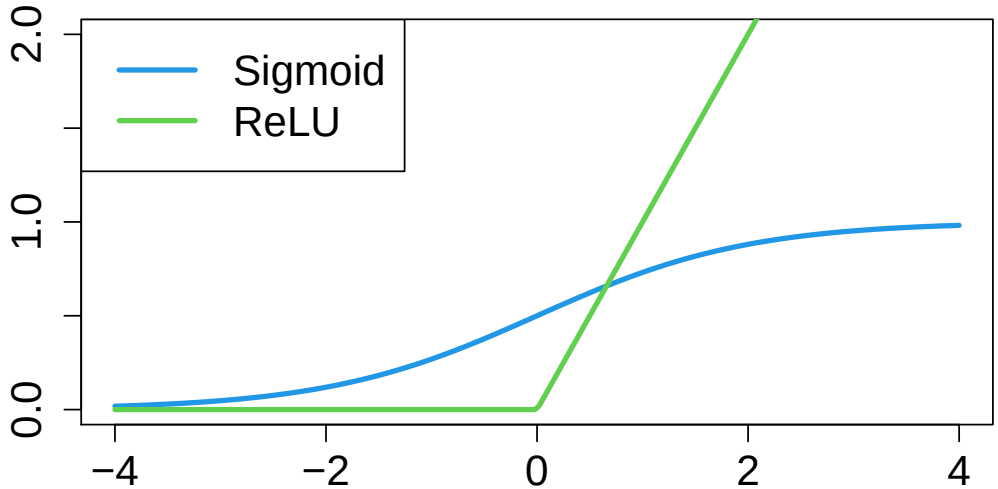
Inside a Unit



- Each unit (hidden/output) takes as its inputs the outputs of the units in the layer before, each associated with a weight w_{ij}.
- A weighted sum of all inputs is computed, where a bias unit has its $y_0 = 1$ (i.e., w_{0j} is the intercept of the linear model).
- The sum is then transformed by an activation function g_j to produce the output.

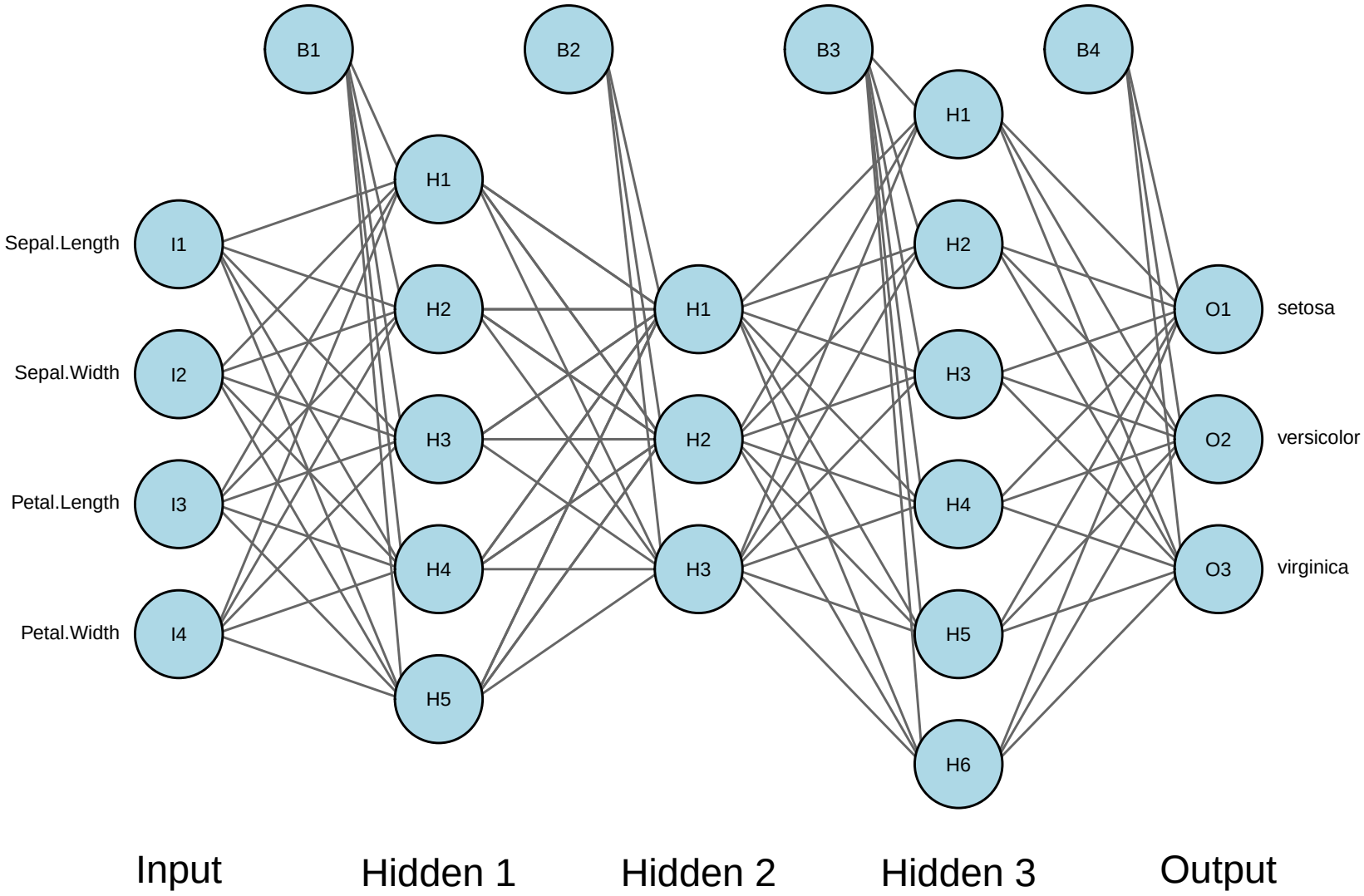
Activation Functions

- An activation function is an increasing, nonlinear function.
- Popular ones for internal units:
  - Sigmoid: $g(z) = 1/(1+e^{-z})$.
  - ReLU (rectified linear unit): $g(z) = (z)_+ = zI(z > 0)$.



- There can be a final transformation of the output vector z:
  - Identity (for regression): $g(z) = z$
  - Softmax (for classification): $g_j(z) = e^{z_j}/(\sum_l e^{z_l})$

Multilayer Neural Networks



Fitting a Neural Network

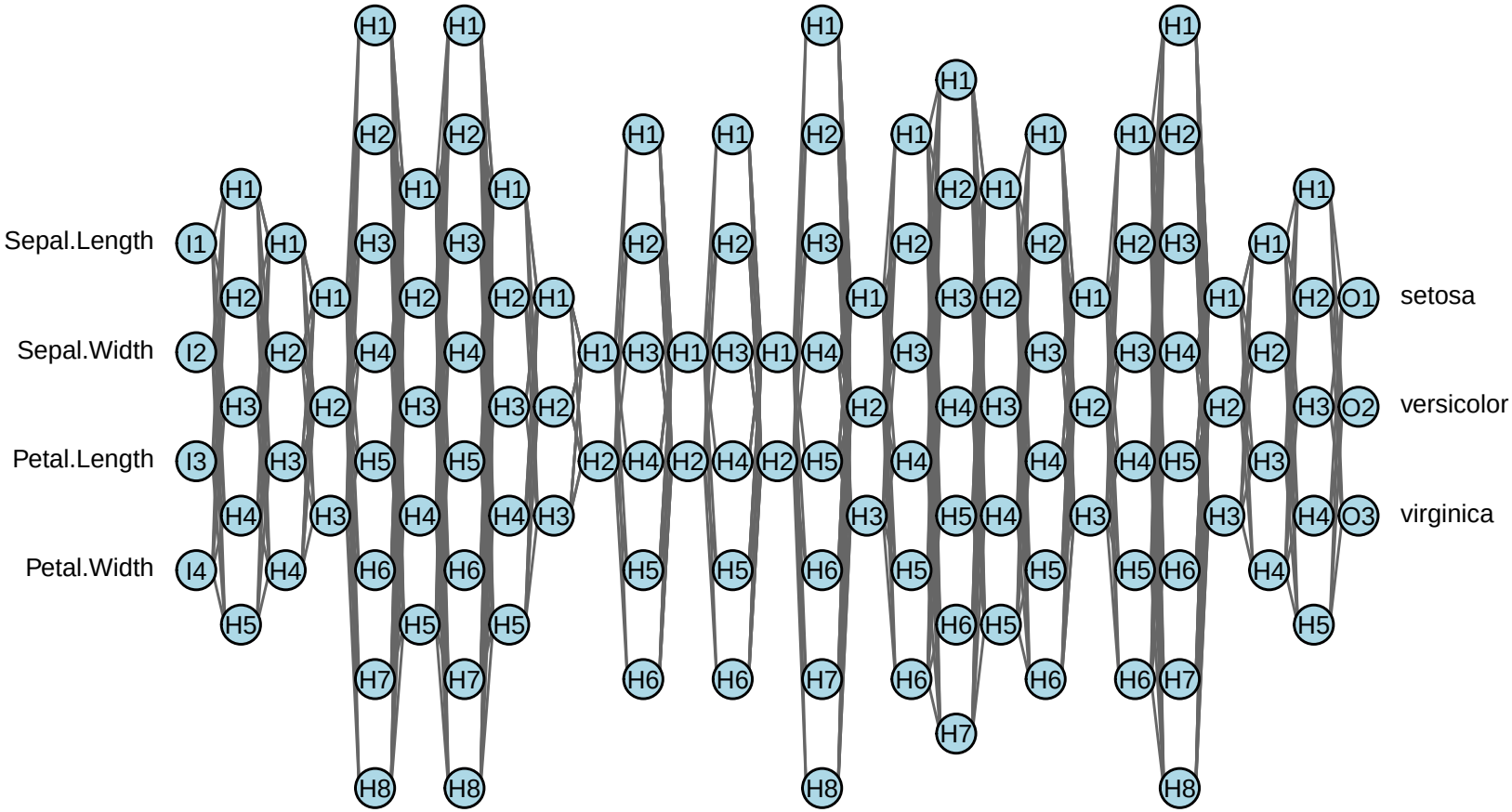
- Parameters are the weights (linear coefficients).
- Minimise the sum of squared errors (same as least squares) for regression.
- Minimise the cross-entropy (same as maximum likelihood) for classification.
- May converge to a local optimum.
  - Fit the network with several different sets of initial parameter values, and choose the best one.

Fitting a Neural Network II

- Neural networks can be overfitted, so choose a simpler network or perform some pruning/regularisation (availability depends on implementations/packages).
- Bagging can also be used for neural networks.
- It may take a long time to train a neural network, but it is straightforward for a trained model to make predictions.
- Neural networks are difficult for human interpretation, but they tend to have fairly good performance, especially with the extension of deep learning.,

Deep Learning

(Bias units not shown)



Tens or hundreds of hidden layers

Deep Learning

- Why so many layers?
- Convolutional neural networks:
  - For image classification
  - Types of layers: convolution, pooling, flatten, dense
- Recurrent neural networks:
  - For time series forecasting

Recommended Readings

ISLv2 (basics):

- Sections 10.1–10.3
- Labs: Sections 10.9.1–10.9.3

ESL (advanced):

- Sections 11.3, 11.4

See also:

- Chollet and Allaire (2007). *Deep Learning With R*. Manning Publications.