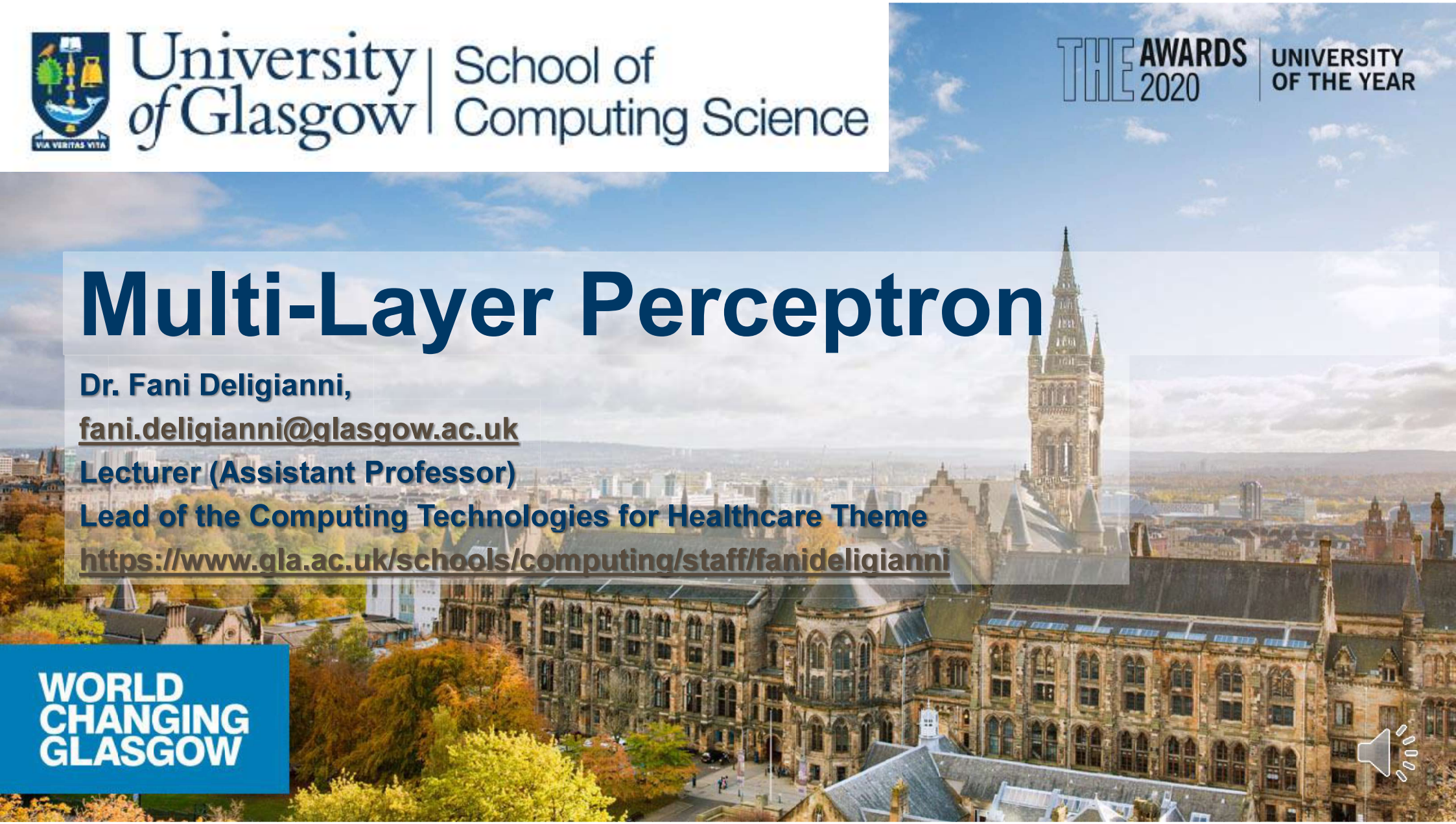
University of Glasgow | School of Computing Science

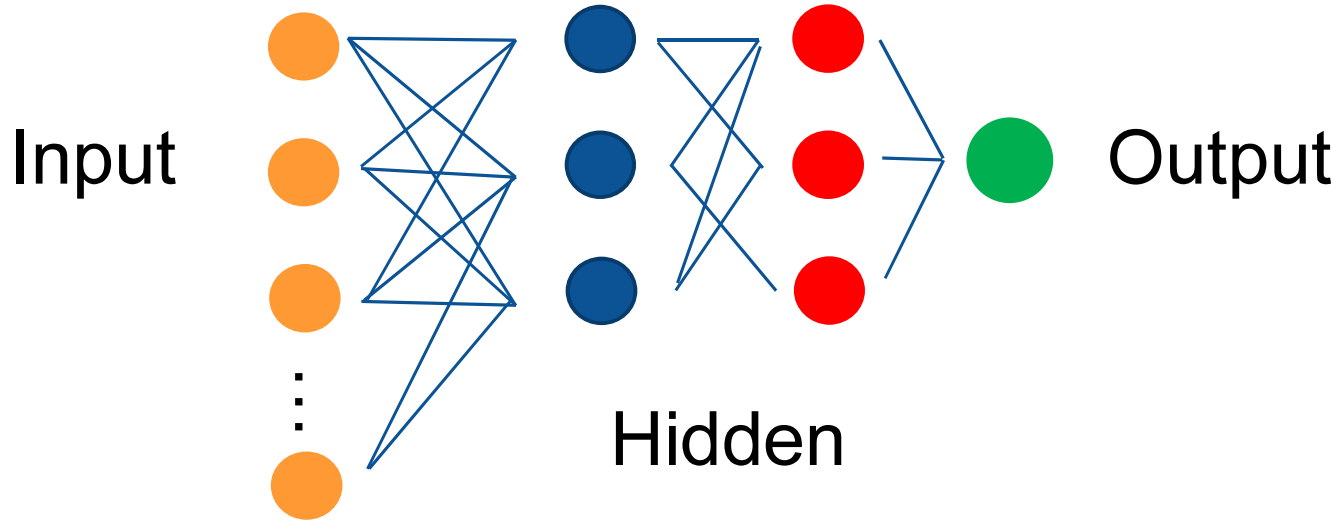


Multi-Layer Perceptron

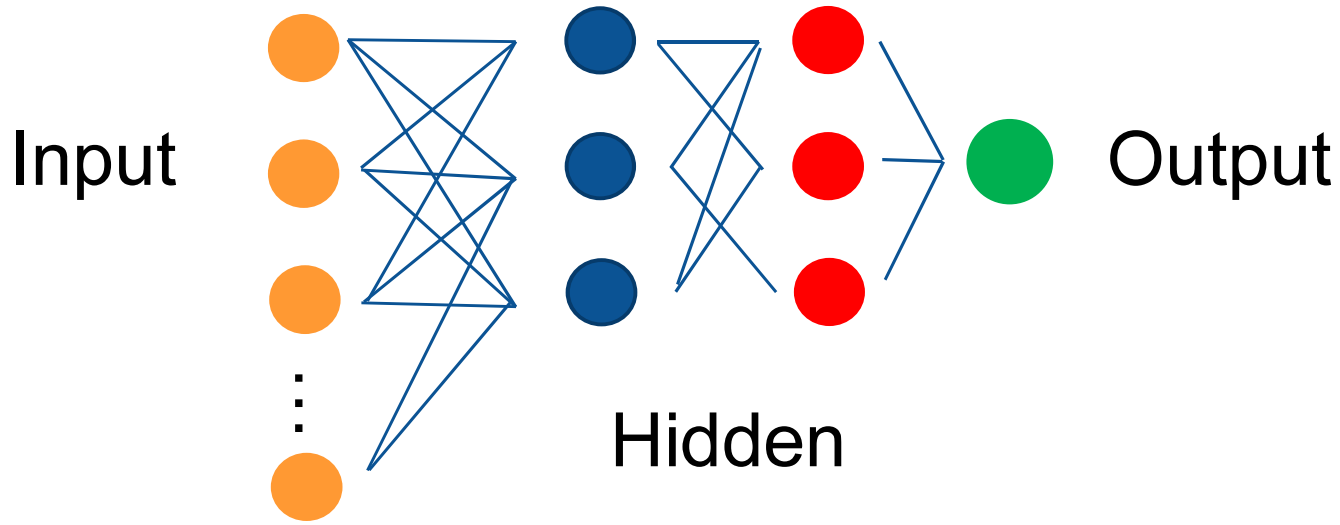
Dr. Fani Deligianni,
fani.deligianni@glasgow.ac.uk
Lecturer (Assistant Professor)
Lead of the Computing Technologies for Healthcare Theme
https://www.gla.ac.uk/schools/computing/staff/fanideligianni

Multi Layer Perceptron

- The neural network multi-layer perceptron (NNMLP) are **feed-forward neural networks.**
- They consist of **minimum three layers of nodes**, which are input, hidden and output layer.



Model Layers: Dense



- Each neuron in a layer receives an input from all the neurons present in the previous layer—thus, they're densely connected.
- In other words, the **dense layer is a fully connected layer**, meaning all the neurons in a layer are connected to those in the previous layer.

Activation Functions

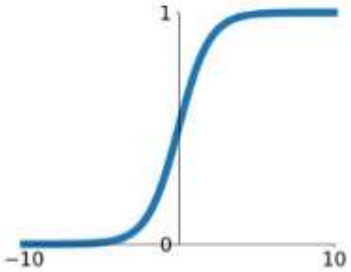
- The **sigmoid activation function**, also called the logistic function, is a very popular activation function for neural networks. The input to the function is transformed into a value between 0.0 and 1.0.

- The advantage of the **Hyperbolic Tangent (tanh) function** is that it produces a zero-centered output, thereby supporting the backpropagation process. The tanh function has been mostly used in recurrent neural networks.

- ReLU stands for **rectified linear unit** and is a type of activation function. Mathematically, it is defined as y = max(0, x). Its a simple function which allows model to account for non-linearities and interactions.

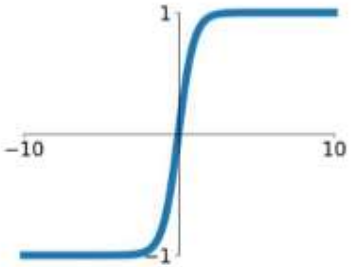
Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



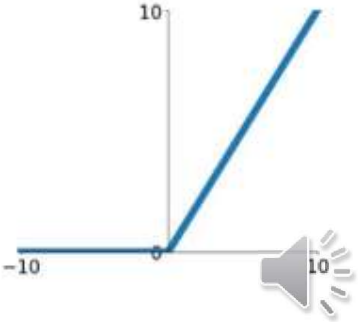
tanh

$$\tanh(x)$$



ReLU

$$\max(0, x)$$

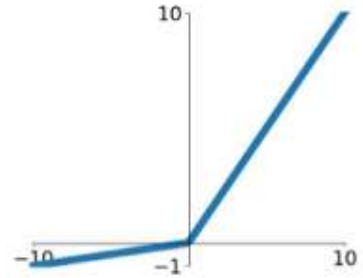


Activation Functions

- **Leaky ReLU** function is an improved version of the ReLU activation function. For the ReLU activation function, the gradient is 0 for all the values of inputs that are less than zero, which would deactivate the neurons in that region and may cause dying ReLU problem.

Leaky ReLU

$$\max(0.1x, x)$$



- **The Maxout Unit** is a generalization of the ReLU and the leaky ReLU functions. It is a piecewise linear function that returns the maximum of the inputs. Both ReLU and leaky ReLU are special cases of Maxout.

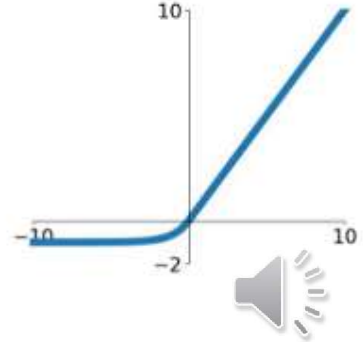
Maxout

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

- **Exponential Linear Unit** or its widely known name ELU is a function that tend to converge cost to zero faster and produce more accurate results. Different to other activation functions, ELU has an extra alpha constant which should be positive number.

ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



Activation Functions

- **Softmax function** is useful when the total sum of the output of a layer needs to be equal to one. It is used to produce a vector of probabilities and it is common in the last layer of deep neural networks for multi-class classification.

$$y_i \frac{e^{x_i}}{\sum_{j=1}^{J} e^{x_j}}$$

- **Hierarchical Softmax** is useful when the number of classes is high, and it is computationally expensive to use the softmax function. It approximates probability distributions as a binary tree with larger depth corresponding to less probable class activations.

Summary

- A multi-layer perceptron consists of at least three layers
- Internal nodes of a network are processing functions
- Activation functions are required to be differentiable functions
- Activations functions are monotonic but their derivates are not

References

- Ravi et al. Deep Learning for Health Informatics, IEEE Journal of Biomedical and Health Informatics, 21(1), 2017
- Kamath, Deep Learning for NLP Applications, Springer, 2019
- Foster, Generative Deep Learning – Teaching Machines to Paint, Write, Compose and Play, O'Reilly, 2019