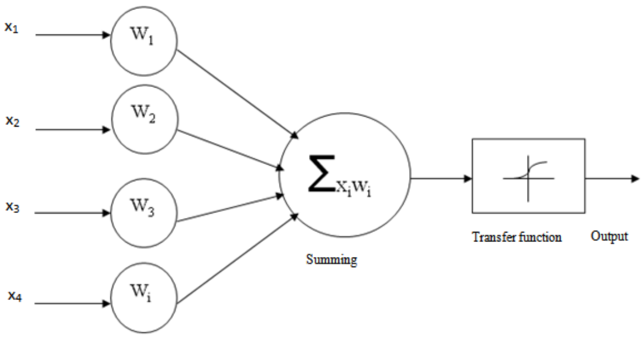
MT Summer Term 2021 Ex8: A Bluffer's Guide to NNs

1. In your own words, please describe supervised machine learning. Think about data, descriptive features, labels to be predicted, finding a function that describes the data, models, model bias, parameters, error functions, what is adjusted by learning, how learning is driven by the error, how learning tries to minimise error, overfitting/underfitting the data, generalising to new unseen data, training test and development data etc. Use an example, if you find this useful.

2. Given the following AN:



where $x_1 = 0.9$, $x_2 = -0.3$, $x_3 = 1.2$, $x_4 = 0.1$ and $w_1 = 0.5$, $w_2 = 0.8$, $w_3 = -1.1$, $w_2 = 0.2$, please compute

$$z = \sum_{i=1}^{4} x_i w_i$$

$$a = \text{sigmoid}(z)$$

3. Please write the input in (2) above as a column vector $\mathbf{x}.$ What is the transpose $\mathbf{x}^{\mathbf{T}}$ of $\mathbf{x}$? Please write the weights in (2) above as a column vector $\mathbf{w}.$ What is the transpose $\mathbf{w}^{\mathbf{T}}$ of $\mathbf{w}$?

4. Vectorise the AN in (2) above: please express the same AN as in (2) above in terms of an input column vector $\mathbf{x}$ and a weight column vector $\mathbf{w}$ (with the same values as in (2) above) in terms of an inner product between the vectors and an element-wise application of the sigmoid to the result of the inner product. Please do this once with $\mathbf{w}$ followed by $\mathbf{x}$, and once the other way around. (Make sure you remember how to place the transpose to make this work).

5. Please complete the definition of the sigmoid below

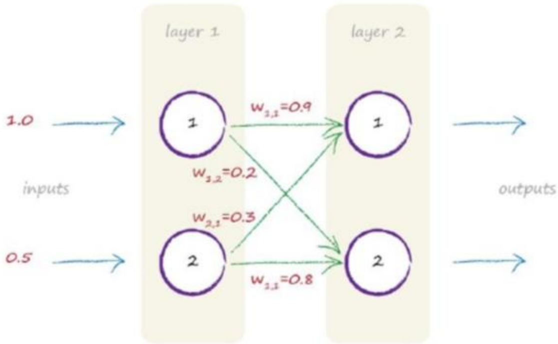
$$\text{sigmoid}(z) = \cdots$$

and draw the sigmoid function in a coordinate system where $z \in [-5, 5]$ and $\text{sigmoid}(z) \in [-0.5, 1.5]$. What are $\text{sigmoid}(0)$, $\text{sigmoid}(-4)$ and $\text{sigmoid}(4)$?

6. Please work out the inner product between the following two matrices:

$$\begin{pmatrix}1 & 2\\ 3 & 4\end{pmatrix}\begin{pmatrix}5 & 6\\ 7 & 8\end{pmatrix} =$$

7. Given the following simple FFNN, please compute the activations (the outputs) of the two ANs in layer 2. Please assume that layer 1 ANs are just input neurons ("sensors") without activation functions, and that layer 2 ANs have sigmoid activation functions.

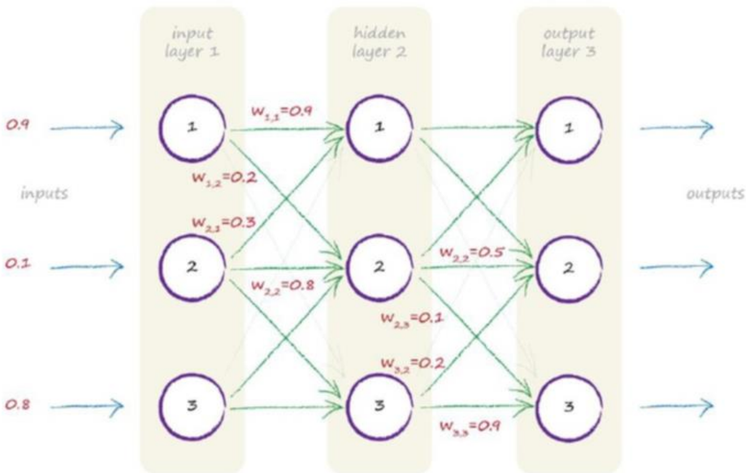


8. Please vectorise the simple FFNN in (7) above: define an input column vector $\mathbf{I}$ and a weight matrix $\mathbf{W}$ such that output column vector $\mathbf{O} = \text{sigmoid}(\mathbf{W} \cdot \mathbf{I})$, where "$\cdot$" is the inner product, and $\mathbf{I}$ and $\mathbf{O}$ are column vectors.

9. Can you do the same thing as in (8 above) but where $\mathbf{O} = \text{sigmoid}(\mathbf{I} \cdot \mathbf{W}^{\mathbf{T}})$, where "$\cdot$" is the inner product, and $\mathbf{I}$ and $\mathbf{O}$ are row vectors. (Please make sure you adapt $\mathbf{W}$ from (8) above in the right way for this to work).

10. Why do we vectorise ANs and ANNs?

11. Given the following FFNN



with

$$I = \begin{pmatrix} 0.9 \\ 0.1 \\ 0.8 \end{pmatrix} \quad W_{input\_hidden} = \begin{pmatrix} 0.9 & 0.3 & 0.4 \\ 0.2 & 0.8 & 0.2 \\ 0.1 & 0.5 & 0.6 \end{pmatrix} \quad W_{hidden\_output} = \begin{pmatrix} 0.3 & 0.7 & 0.5 \\ 0.6 & 0.5 & 0.2 \\ 0.8 & 0.1 & 0.9 \end{pmatrix}$$

please work out :

$$\mathbf{X}_{\text{hidden}} = \mathbf{W}_{\text{input\_hidden}} \cdot \mathbf{I} \quad \mathbf{O}_{\text{hidden}} = \text{sigmoid}(\mathbf{X}_{\text{hidden}}) \quad \mathbf{X}_{\text{output}} = \mathbf{W}_{\text{hidden\_output}} \cdot \mathbf{O}_{\text{hidden}}$$

and:

$$\mathbf{O}_{\mathbf{output}} = \text{sigmoid}(\mathbf{X}_{\mathbf{output}})$$