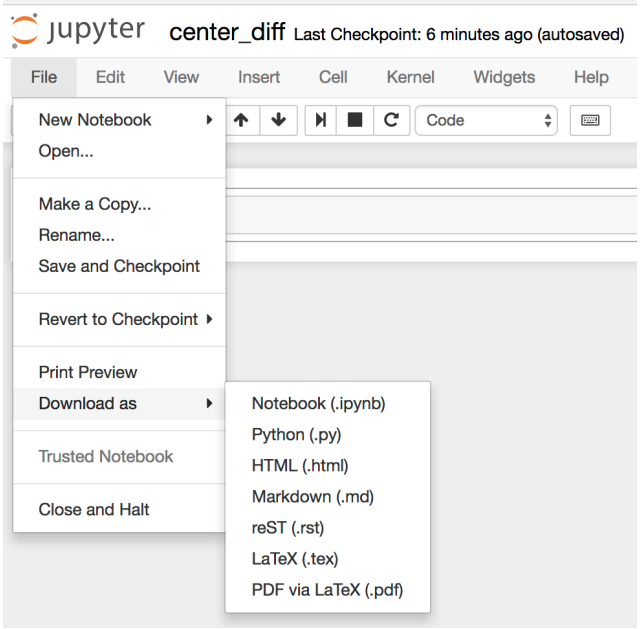
Homework 1
Concept, Derivation and Programming, Due 18:00, Wednesday, Noverber 3, 2021

Late submission within 24 hours: score*0.9;
Late submission before post of solution: score*0.8 (the solution will usually be posted within a week); no late submission after the post of solution)

HW Submission Procedure (請仔細閱讀)

1. For concept and derivation, please write them in a professional format and submit a pdf file. Name your pdf file HW1_yourID, for example, HW1_ n96081494.pdf

2. You should submit your Jupyter notebook and Python script (`*.py`, in Jupyter, click File, Download as, Python (`*.py`)).



3. Name a folder using your student id and HW number (e.g., n96081494-HW4), put the pdf and all the Jupyter notebooks and python scripts into the folder and zip the folder (e.g., n96081494-HW4.zip).

4. Submit your HW directly through the course website.

Total 120%

Concept and Derivation (Name your pdf file HW1_yourID, for example, HW1_ n96081494.pdf)

1. (120%) Consider the simple network example with a single input $x = 2$ and a single output $y = 1$ shown in Figure 1 below.

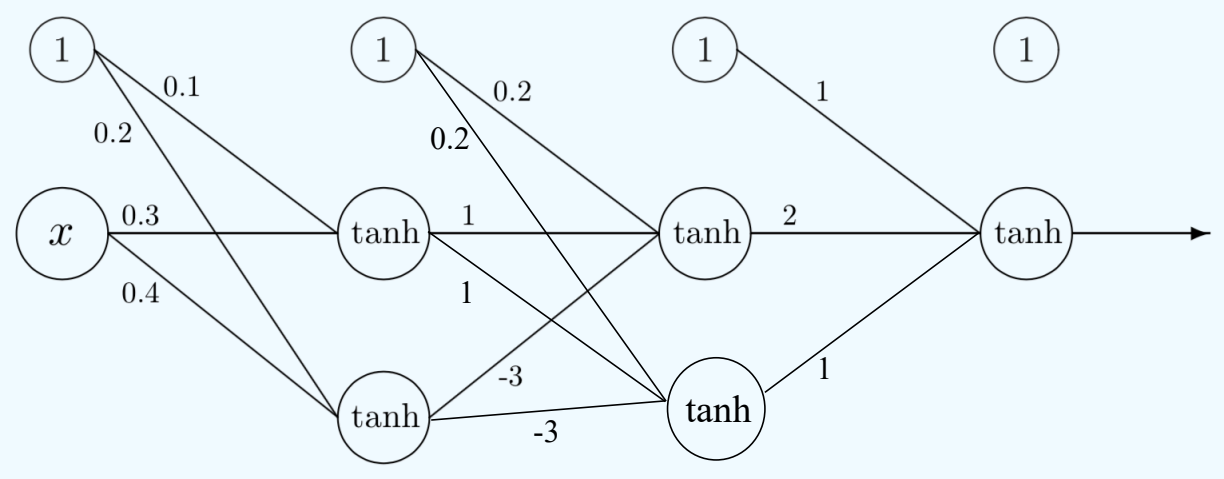


Figure 1

The weight matrices are:

$$\mathbf{W}^{(1)} = \begin{bmatrix} 0.1 & 0.2 \\ 0.3 & 0.4 \end{bmatrix};\ \mathbf{W}^{(2)} = \begin{bmatrix} 0.2 & 0.2 \\ 1 & 1 \\ -3 & -3 \end{bmatrix};\ \mathbf{W}^{(3)} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

and the summation of weighted nodes for layer 1 can be expressed as $\mathbf{u}^{(1)} = (\mathbf{W}^{(1)})^T \mathbf{x}^{(0)}$; you can perform similar operation for other layers.

(a) (25%) Derive and compute $\mathbf{u}^{(1)}$, $\mathbf{z}^{(1)}$, $\mathbf{u}^{(2)}$, $\mathbf{z}^{(2)}$, and $\mathbf{y}^{(3)}$.
(b) (40%) Using the half of the sum square as our error function, derive and compute $\boldsymbol{\delta}^{(3)}$, $\boldsymbol{\delta}^{(2)}$, $\boldsymbol{\delta}^{(1)}$.
(c) (25%) Compute $\frac{\partial E_n}{\partial \mathbf{W}^{(1)}}$, $\frac{\partial E_n}{\partial \mathbf{W}^{(2)}}$, $\frac{\partial E_n}{\partial \mathbf{W}^{(3)}}$.
(d) (30%) Update the weight matrices using learning rate $\eta = 0.5$, repeat the forward propagation and compute $\mathbf{u}^{(1)}$, $\mathbf{z}^{(1)}$, $\mathbf{u}^{(2)}$, $\mathbf{z}^{(2)}$, and $\mathbf{y}^{(3)}$.