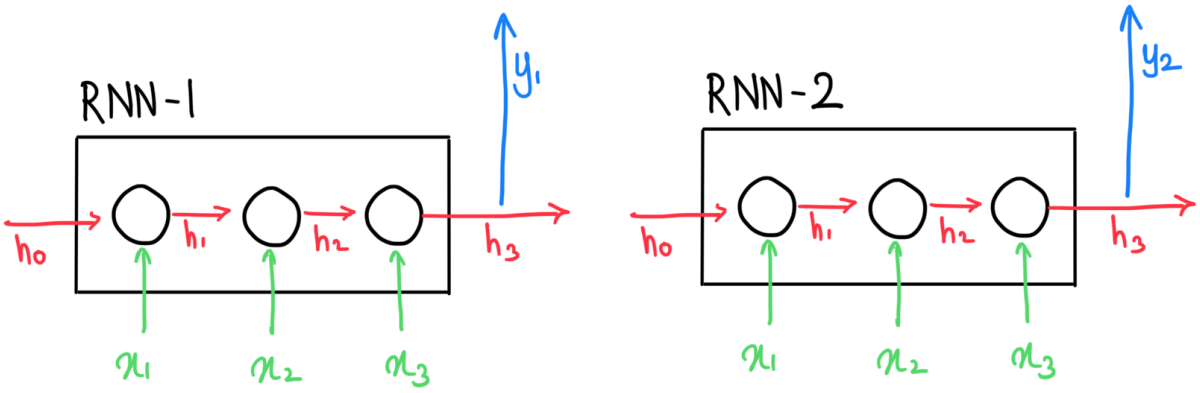
Assignment 5.1

Estimate the value of yhat for the below network,



$$W_x = [3, -4]$$

$$W_h = \begin{bmatrix} 4 & -5 \\ -3 & 2 \end{bmatrix}$$

$$W_y = \begin{bmatrix} -4 \\ 2 \end{bmatrix}$$

$$b_h = 0 \quad b_y = 10$$

$$h_0 = [0, 0]$$

$$x = [1, 2, 3]$$

Assignment 5.2

Problem Statement

To build an NLP model to combat fake news/contents menace using Embeddings and RNN

Data:

https://www.kaggle.com/datasets/ssismasterchief/machine-hack-fake-news-content-detection

Use only Train.csv - 10240 rows x 3 columns (includes Labels Columns as Target)

Variable Description:
Text - Raw content from social media/ new platforms
Text_Tag - Different types of content tags
Labels - Represents various classes of Labels

- Half-True - 2
- False - 1
- Mostly-True - 3
- True - 5
- Barely-True - 0
- Not-Known - 4

Publish your final work solution in this Kaggle dataset.

Task:

1. Read train.csv in pandas.
2. Calculate the distribution of labels.
3. Normalize the text by making it in lower case, and preprocess the text by removing punctuations, stopwords, repeated words, and words with length greater than 2.
4. Generate the word cloud for label 1 (False news).
5. Split the clean text and labels into a training and testing set with 80:20 ratio.
6. Tokenize the clean text on the training set using Tensorflow library. Generate the tokens for training and testing sets. Print total tokens.
7. Generate the sequences for the training and testing set.
8. Apply post padding on the sequences using Tensorflow with maxlen 20 on both sets.
9. Build the RNN to predict 6 possible labels with the help of Embeddings by setting the embedding dimension as 6.
    a. Add an embedding layer with input_length equal to padding maxlen.

    b. Add 3 RNN layers with units 64, 32, and 16 respectively.
    c. Add a dense layer with 24 units.
    d. Set metrics as F1 score.
10. Justify the total params of the designed network.
11. Train the model with 20 epochs, specifying the testing set.
12. Calculate the log loss, F1 score, and confusion matrix of the training and testing set.