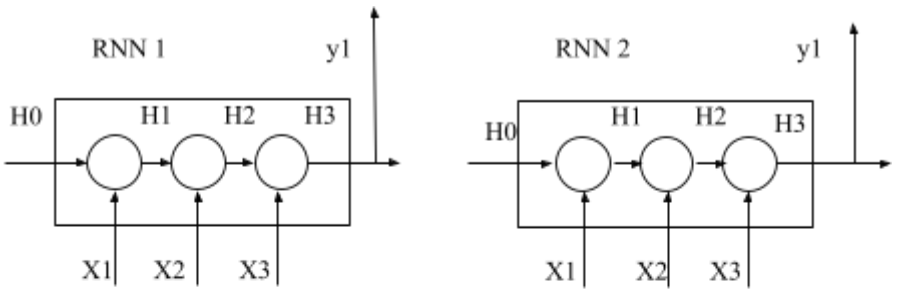
Assignment 5.1

Estimate the value of yhat for the below network,



$$Wx = \begin{pmatrix} -4 & 3 \end{pmatrix} \qquad H0 = \begin{pmatrix} 0 & 0 \end{pmatrix}$$

$$Wh = \begin{pmatrix} -5 & 3 \\ 2 & -3 \end{pmatrix} \qquad X = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix}$$

$$B0 = 0 \qquad By = 10$$

$$Wy = \begin{pmatrix} 2 \\ -4 \end{pmatrix}$$

Assignment 5.2

Problem Statement

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

Data:

https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset

Task:

1. Read fake.csv and true.csv and merge the contents in pandas dataframe.
2. Create a separate column with fake and true labels. Shuffle the data frame.
3. Calculate the distribution of labels.
4. Normalize the **text** column by making it in lower case, and preprocess the text by removing punctuations, stopwords, repeated words, and words with length greater than 2.
5. Generate two word clouds one for fake news and second for fake news in the subject of politics.
6. Split the clean text and labels into a training and testing set with 80:20 ratio.
7. Tokenize the clean text on the training set using Tensorflow library. Generate the tokens for training and testing sets. Print total tokens.
8. Generate the sequences for the training and testing set.
9. Apply post padding on the sequences using Tensorflow with maxlen ~20 on both sets.
10. Build the RNN with the help of Embeddings by setting the embedding dimension as 4.
    a. Add an embedding layer with input_length equal to padding maxlen.
    b. Add 3 RNN layers with optimal units.
    c. Add a dense layer with optimal units.
    d. Set metrics as ROC-AUC score.
11. Justify the total params of the designed network.
12. Train the model with 20 epochs, specifying the testing set.
13. Calculate the log loss, ROC-AUC score, and confusion matrix of the training and testing set.
14. Publish your final work solution in this Kaggle dataset.