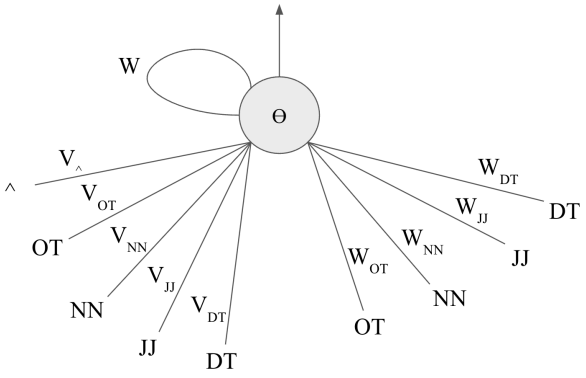
Assignment 2

Given a POS-tagged corpus, train a **single recurrent perceptron** to mark noun chunks in a sentence. In a noun chunk, only the noun is compulsory, determiners and adjectives are optional. Use the following single recurrent perceptron for training.



You are provided with two files:

train.jsonl - Contains training data

test.jsonl - Contains test data

For POS tagging, only 4 tags are used- Nouns (NN), Determiners (DT), Adjectives (JJ), and Others (OT) with the following mapping:

```
{
    “NN”:1,
    “DT”:2,
    “JJ” :3,
    “OT”:4
}
```

For noun chunks, the beginning of the chunk will be labeled 1 and the rest of the words in the chunk will be labeled 0. All other words are labeled 1.

Example format from midsem Q9:

Raw corpus: *Boys play joyfully with brown balls in the green big field*

POS tagged:

tokens - [*Boys, play, joyfully, with, brown, balls, in, the, green, big, field*]

pos_tags - [1,4,4,4,3,1,4,2,3,3,1]

Chunk:

tokens - [*Boys, play, joyfully, with, brown, balls, in, the, green, big, field*]

chunk_tags - [1,1,1,1,1,0,1,1,0,0,0]

Instructions:

1. Implement a single recurrent perceptron and BPTT (backpropogation through time) from scratch and train it using the training data provided.
2. Use 5-fold cross-validation and report results.
3. Report results on the test set with the model trained on the full training data provided.
4. Show error cases and perform analysis.
5. Report model weights and show that the following inequalities hold for the perceptron to satisfy the language constraints.

Prev (V) \ Current (W)	DT	JJ	NN	OT
^	$V_{\wedge}+W_{DT}>\Theta$	$V_{\wedge}+W_{JJ}>\Theta$	$V_{\wedge}+W_{NN}>\Theta$	$V_{\wedge}+W_{OT}>\Theta$
DT		$W+V_{DT}+W_{JJ}<\Theta$	$W+V_{DT}+W_{NN}<\Theta$	
JJ		$V_{JJ}+W_{JJ}<\Theta$	$V_{JJ}+W_{NN}<\Theta$	
		$W+V_{JJ}+W_{JJ}<\Theta$	$W+V_{JJ}+W_{NN}<\Theta$	

Prev (V) \ Current (W)	DT	JJ	NN	OT
NN				$V_{NN}+W_{OT}>\Theta$
				$W+V_{NN}+W_{OT}>\Theta$
OT	$W+V_{OT}+W_{DT}>\Theta$	$W+V_{OT}+W_{JJ}>\Theta$	$W+V_{OT}+W_{NN}>\Theta$	$W+V_{OT}+W_{OT}>\Theta$