Why Sequence Models ?

Examples:

1. Speech recognition → For a given audio clip we can output the sentence.

2. Music Generation → For a given input value(number, might be null) generates music(o/p is sequence)

3. Sentiment Classification → Given a sentence(sequence) predicts the star rating

4. DNA sequence analysis → Given a DNA sequence we find the required sequence(Protein sequence)

5. Machine Translation → For given sentence in one language we translate to another language

6. Video activity recognition → For a given video frames we predict the action happening in that frames.

7. Name Entity recognition → In a given sentence we identify the people in that sentence.

All the above are supervised learning, but the input and output length may vary also the type of input and output will differ.

Sequence data?

Notation :

Motivating example:

X → Harry Potter and Hermione Granger invented a new spell.

$X^{<1>}$ $X^{<2>}$ $X^{<3>}$………………… ……….. $X^{<7>}$…………………….. $X^{<9>}$

Y → 1 1 0 1 1 0 0 0 0

$y^{<1>}$ $y^{<2>}$ $y^{<3>}$………………… ……….. $y^{<7>}$…………………….. $y^{<9>}$

T_x=9 → length of input vector

T_y=9 → length of output vector

$X^{(i)<t>}$ → t element in i^{th} training example

$y^{(i)<t>}$ → t element in output sequence of i^{th} example

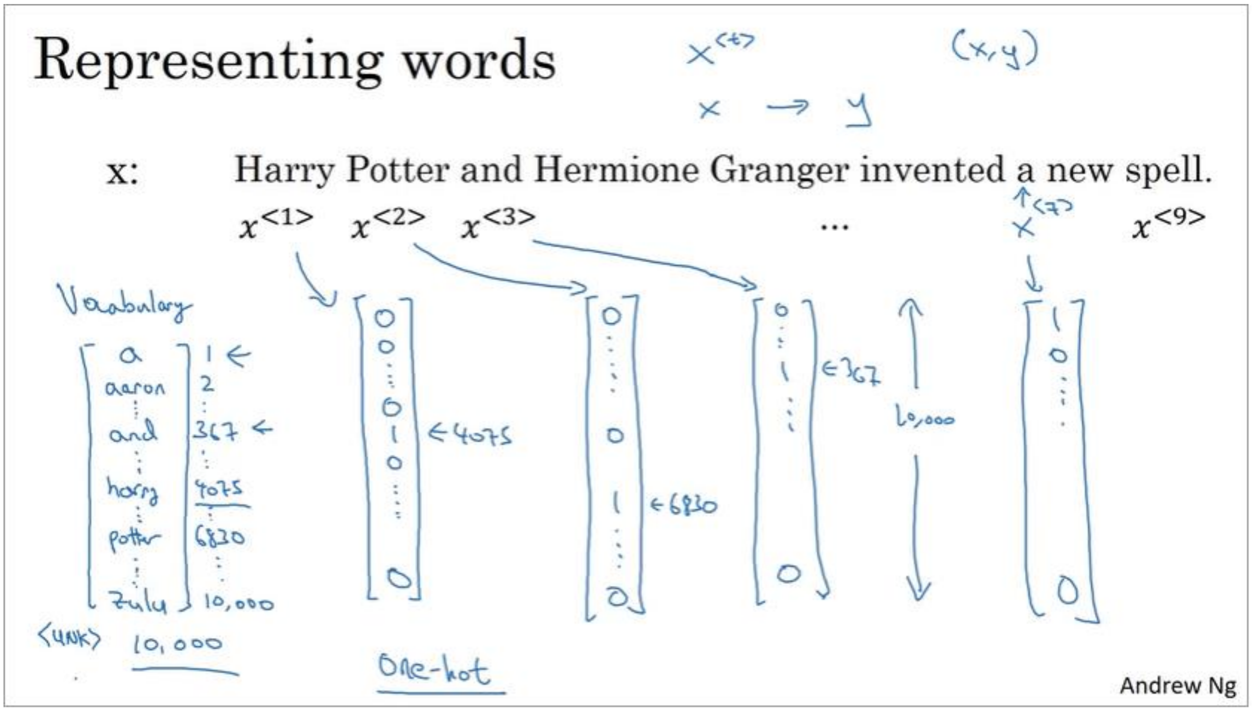
T^{i}_{x} → length of ith training examples.

T^{i}_{y} → length of output sample of ith training example.

Representing words :

We come up with a dictionary with words called vocabulary and we use one hot encoding for the words in the dictionary.

For example, if we have a 10000 1-D vector



Representing words

x: Harry Potter and Hermione Granger invented a new spell.

$x^{<1>}$ $x^{<2>}$ $x^{<3>}$ … $x^{<9>}$

Andrew Ng

What if we faced a new word which is not in our dictionary ?

➔ We marked it as unk (unknown) word.

Why not a standard network ?

1. Inputs and outputs can be of different lengths in different example, like music generation, sentiment classification for a given input sequence.
2. Doesn't share features learned previously,
   i.e. for name entity recognition if a word harry is learned as person then in the feature instance if harry appears then we can use previous learning to reduce cost.

Next chapter is RNN