



deeplearning.ai

Recurrent Neural Networks

Language model and
sequence generation

Language modeling is one of the most basic and important tasks in natural language processing. It's also one that RNNs do very well.

What is language modelling?

Speech recognition

The apple and pair salad.

And the way a speech recognition system picks the second sentence is by using a language model which tells it what the probability of either of these two sentence.

→ The apple and pear salad.

$$P(\text{The apple and pair salad}) = 3.2 \times 10^{-13}$$

$$P(\text{The apple and pear salad}) = 5.7 \times 10^{-10}$$

$$P(\text{Sentence}) = ?$$

$$P(y^{(1)}, y^{(2)}, \dots, y^{(T)})$$

Language modelling with an RNN

And the word corpus is an NLP terminology that just means a large body or a very large set of English text or English sentence.

Training set: large corpus of english text.

Tokenize

form a vocabulary and then map each of these words to one-hot vectors or to indexes in your vocabulary.

Cats average 15 hours of sleep a day. \downarrow $\langle \text{EOS} \rangle$

$y^{(1)}$ $y^{(2)}$ $y^{(3)}$... $y^{(8)}$ $y^{(9)}$
 $x^{(t)} = y^{(t-1)}$

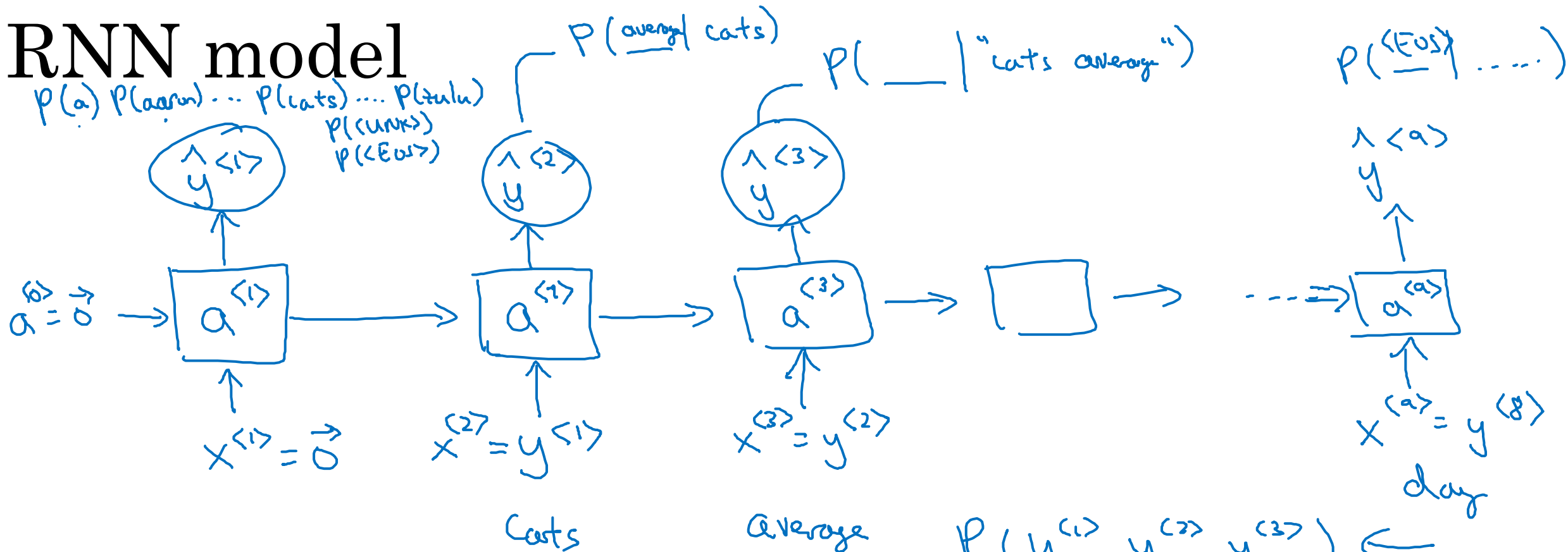
The Egyptian ~~Mau~~ is a breed of cat. $\langle \text{EOS} \rangle$

$\langle \text{UNK} \rangle$

we model the chance of unknown word instead of the specific word Mau.

10,000

RNN model



→ Cats average 15 hours of sleep a day. <EOS>

$$\mathcal{L}(\hat{y}^{<t>}, y^{<t>}) = - \sum_i y_i^{<t>} \log \hat{y}_i^{<t>}$$

← softmax loss function

$$\mathcal{L} = \sum_t \mathcal{L}^{<t>}(\hat{y}^{<t>}, y^{<t>})$$

$$p(y^{(1)}, y^{(2)}, y^{(3)}) \leftarrow$$

$$= \frac{p(y^{(1)}) p(y^{(2)} | y^{(1)})}{p(y^{(3)} | y^{(1)}, y^{(2)})}$$