# Recurrent Neural Networks

## Language model and sequence generation

# What is language modelling?

Speech recognition

The apple and pair salad.

$\rightarrow$ The apple and pear salad.

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

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

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

$P\left(y^{<1>}, y^{<2>}, \ldots, y^{<T_y>}\right)$

# Language modelling with an RNN

Training set: large corpus of english text.

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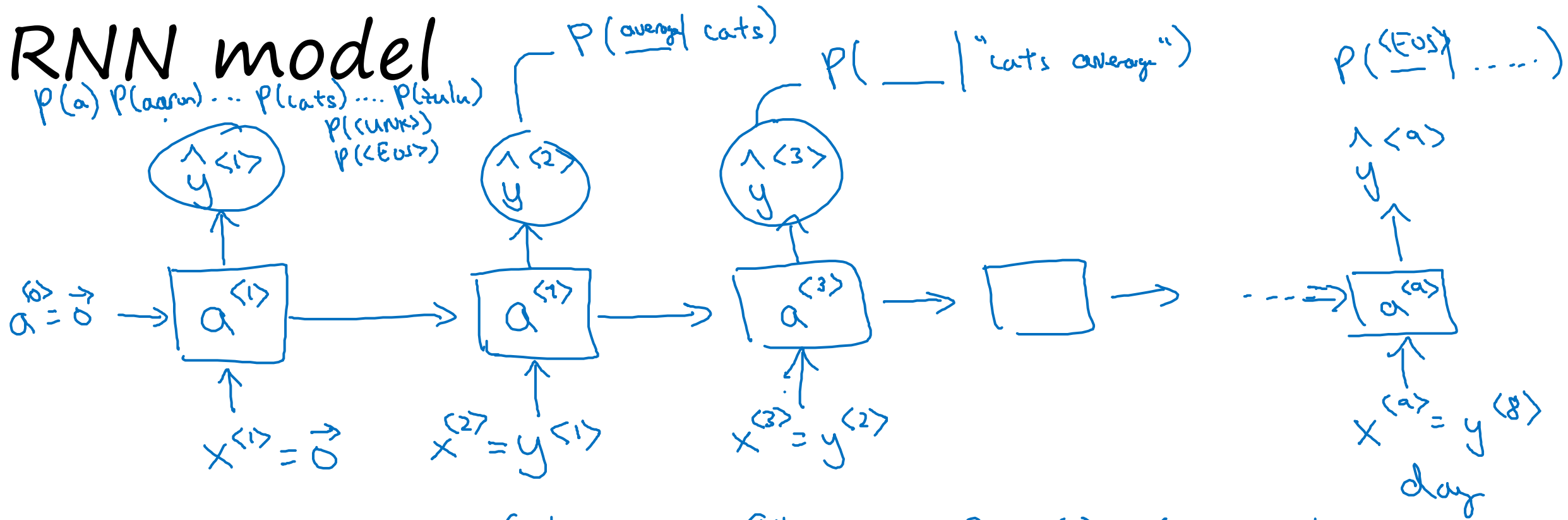
Cats average 15 hours of sleep a day. <EOS>

$y^{(1)}$     $y^{(2)}$     $y^{(3)}$     ...     $y^{(8)}$     $y^{(9)}$

$x^{(t)} = y^{(t-1)}$

The Egyptian Mau is a bread of cat. <EOS>

<UNK>

10,000

# RNN model



$P(a)\ P(aaron)\ \cdots\ P(cats)\ \cdots\ P(zulu)$
$P(<UNK>)$
$P(<EOS>)$

$P(average\ |\ cats)$

$P(\_\_\ |\ "cats\ average")$

$P(<EOS>\ |\ \cdots)$

$a^{<0>} = \vec{0}$

$x^{<1>} = \vec{0}$    $x^{<2>} = y^{<1>}$    $x^{<3>} = y^{<2>}$    $x^{<9>} = y^{<8>}$

Cats     average     day

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

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

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

$P(y^{<1>}, y^{<2>}, y^{<3>})$

$= P(y^{<1>})\ P(y^{<2>} | y^{<1>})$

$P(y^{<3>} | y^{<1>}, y^{<2>})$

Andrew Ng