Recurrent Neural Networks

Language model and sequence generation

What is language modelling?

Speech recognition

The apple and <u>pair</u> salad.

$\rightarrow$ The apple and <u>pear</u> salad.

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

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

$P(\text{Sentence}) = ?$ $\qquad P\left(y^{(1)}, y^{(2)}, \ldots, y^{(T_y)}\right)$

Language modelling with an RNN

Training set: large <u>corpus</u> of english text.

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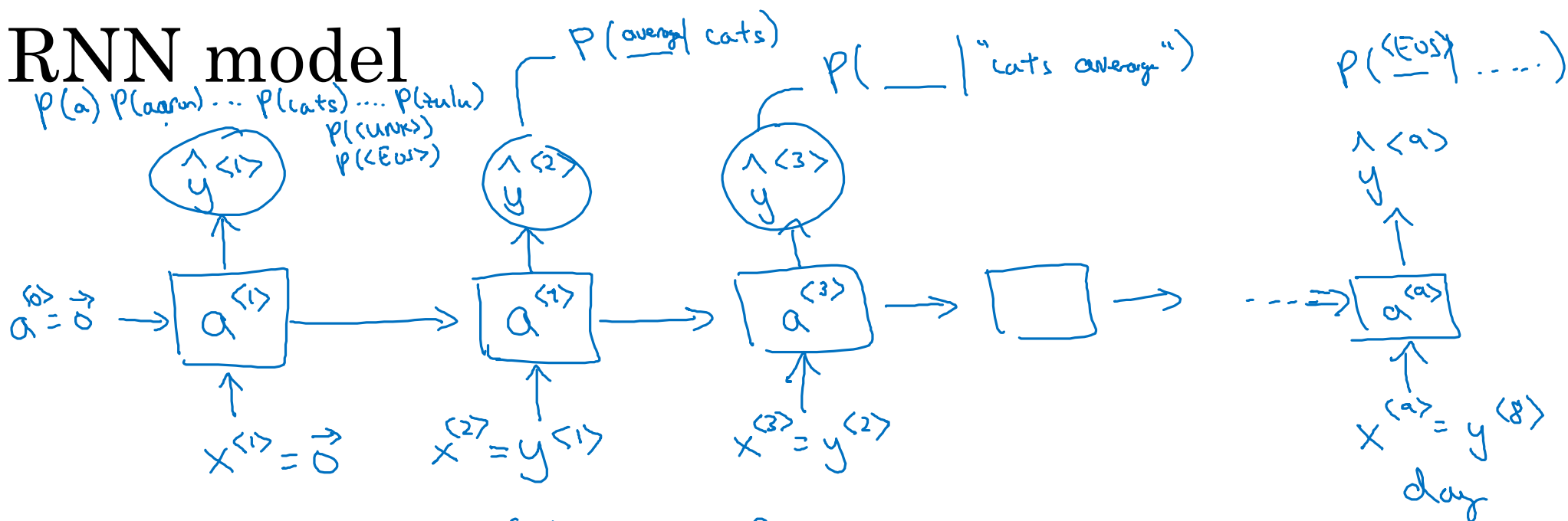
Cats average 15 hours of sleep a day. $\downarrow$ $<EOS>$

$y^{<1>}$ $y^{<2>}$ $y^{<3>}$ $\cdots$ $y^{<8>}$ $y^{<9>}$

$x^{<t>} = y^{<t-1>}$

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

$<UNK>$

$10,000$

Andrew Ng

RNN model

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

$P(average \mid cats)$

$P(\_ \mid \text{"cats average"})$

$P(<EOS> \mid \ldots)$

$\hat{y}^{<1>}$ $\hat{y}^{<2>}$ $\hat{y}^{<3>}$ $\hat{y}^{<9>}$

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

$a^{<1>}$ $a^{<2>}$ $a^{<3>}$ $a^{<9>}$

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

Cats   average   day

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

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

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

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

Andrew Ng