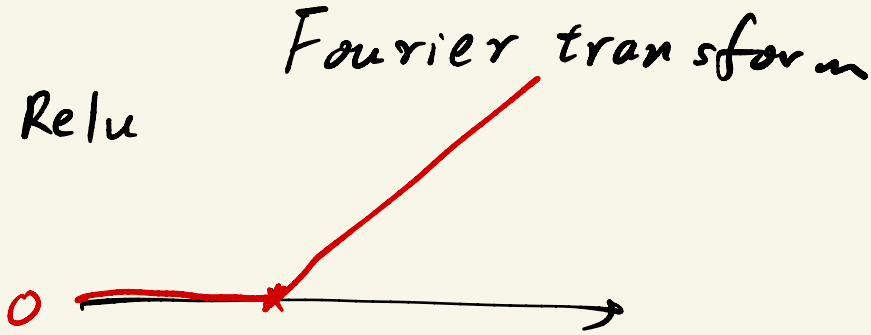
Lecture 19 Neural Arch



$\langle (1,1,1), (-1,0,1) \rangle = 0$

$(1, 1, 1)$

$(1, 1, 2)$

n, m

r

kernel filter

$C_o \times C_i$

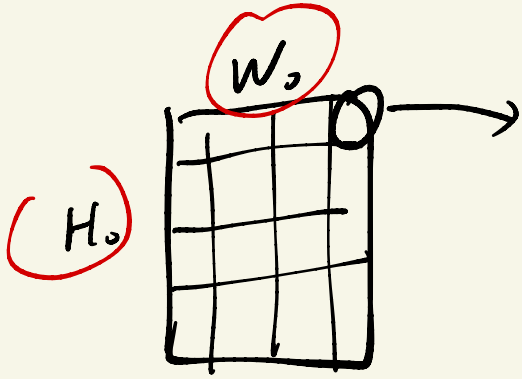
$wx + b$

$3 \times 64 \times 64$

(R, G, B)

$C \times H \times W$

3 64 64



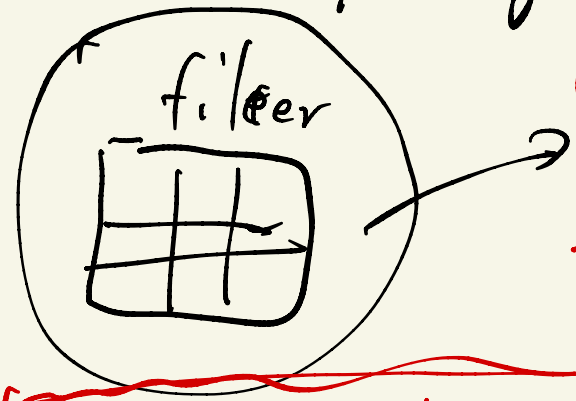
operation

$k_h * k_w * C_o * C_i$

sliding window

convolution is linear

mean pooling



LSTM

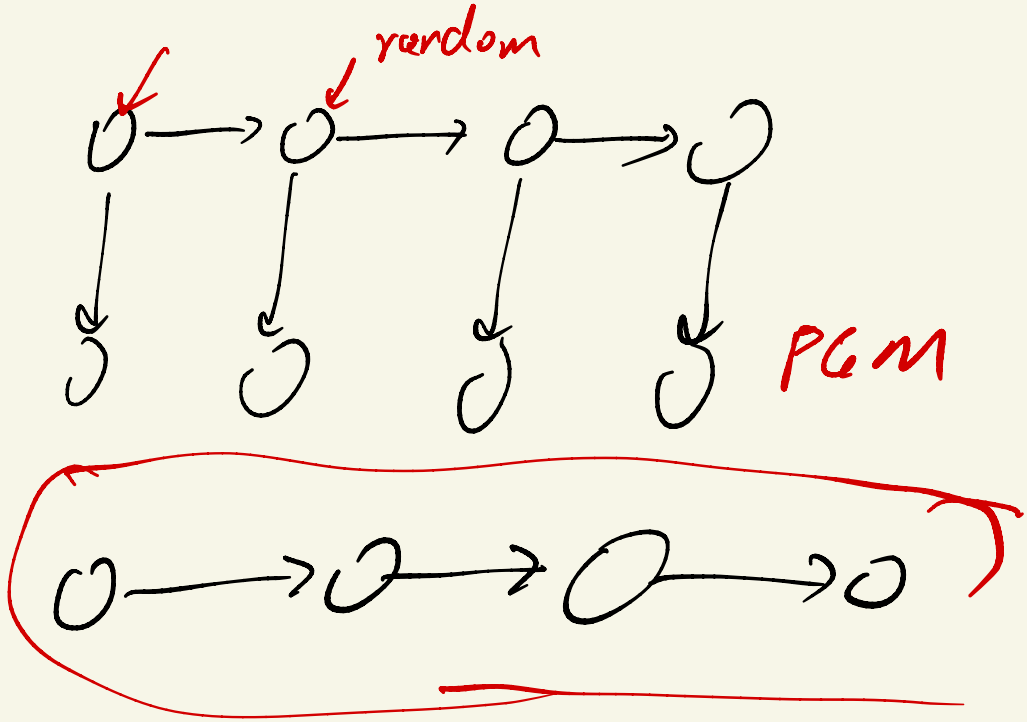
inductive bias

feature engineering

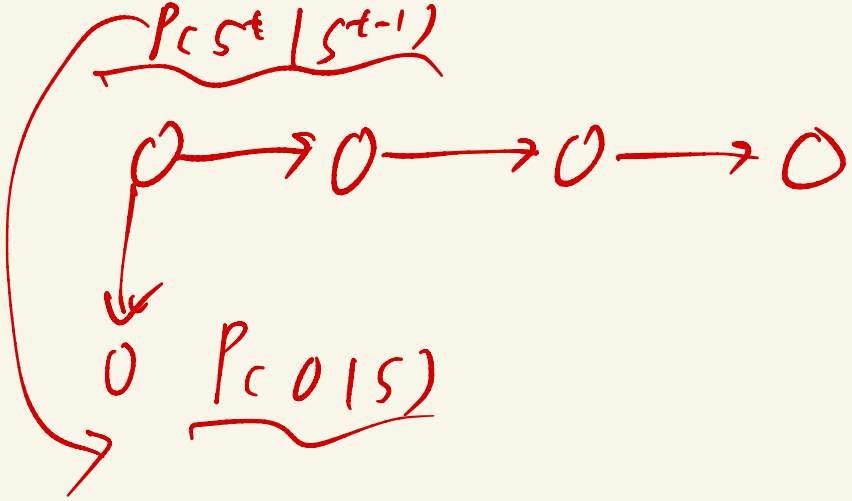
IBM translation

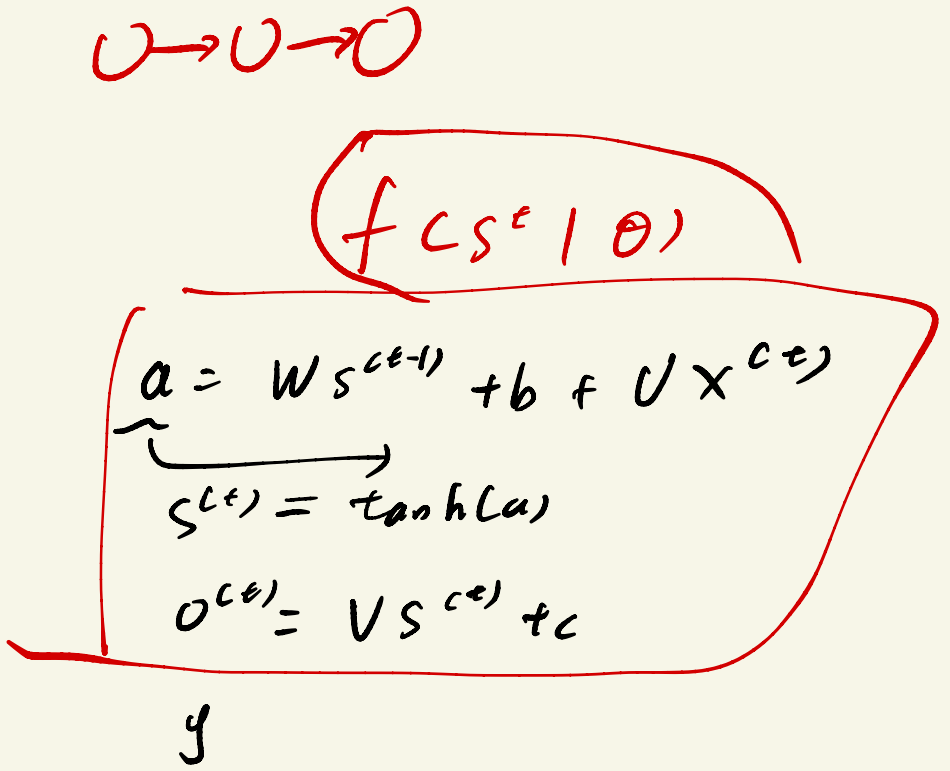
Every time I fire a linguist, my machine translation

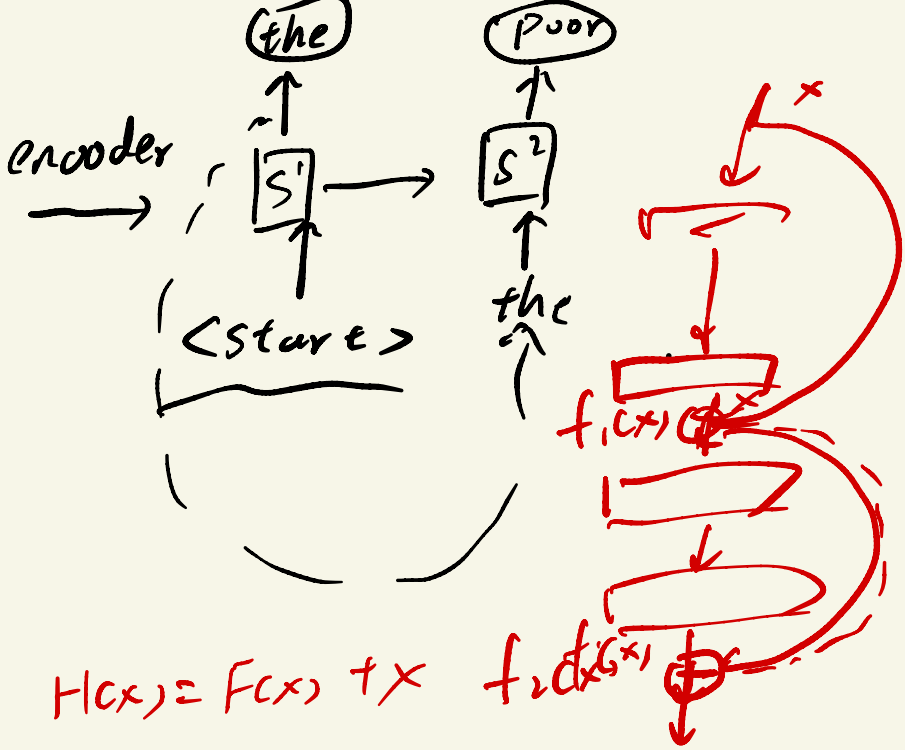
data-driven


random
PGM

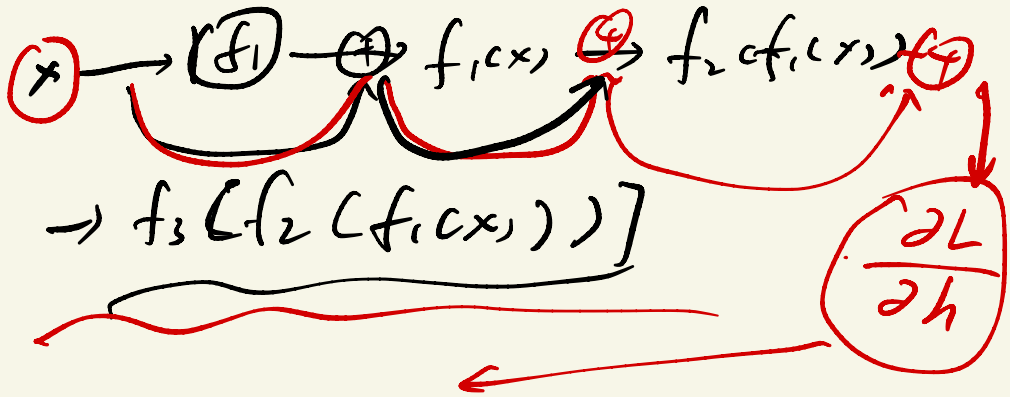
hidden states




f(s^(t-1), θ)
a = W s^(t-1) + b + U x^(t)
s^(t) = tanh(a)
o^(t) = V s^(t) + c
y



H(x) = F(x) + x



Layer1: $x + f_1(x)$

layer2: $x + f_2(x + f_1(x)) + f_1(x)$

layer3: $x + f_3(x + f_2(x + f_1(x))) + f_2(x + f_1(x)) + f_1(x)$

$\frac{\partial L}{\partial x} + \frac{\partial f_3}{\partial} - - -$