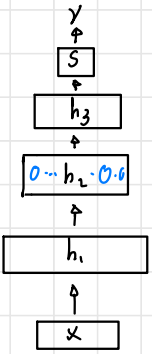
Lecture 17

- Deep Learning
  - Slogan 1 : approximator
  - Slogan 2 : learned computation / algorithm
  - Slogan 3 : team of vectors

- MLP :



More abstract

group of neurons / thought vector

- Math language / Representation :

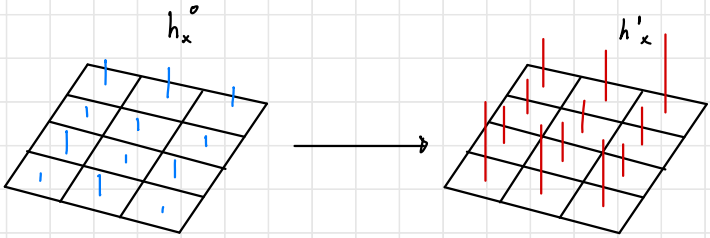
$$h_\ell = \sigma(w_\ell h_{\ell-1} + b_\ell)$$

verb noun

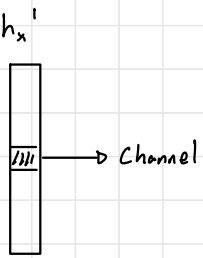
- Embedding / Distributed Representation :

- Convolutional Neural Network (CNN / ConvNet)

x : image

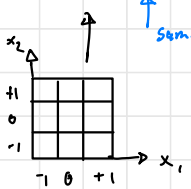


vector : $\begin{pmatrix} R \\ G \\ B \end{pmatrix}$ → channel



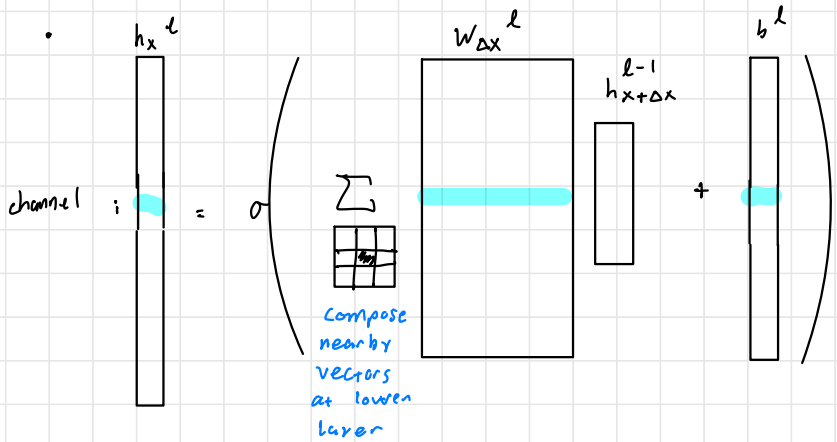
$$h_x^l = \sigma\left(\sum_{\Delta x \in \Delta} W_{\Delta x}^l \, h_{x+\Delta x}^{l-1} + b^l \right)$$

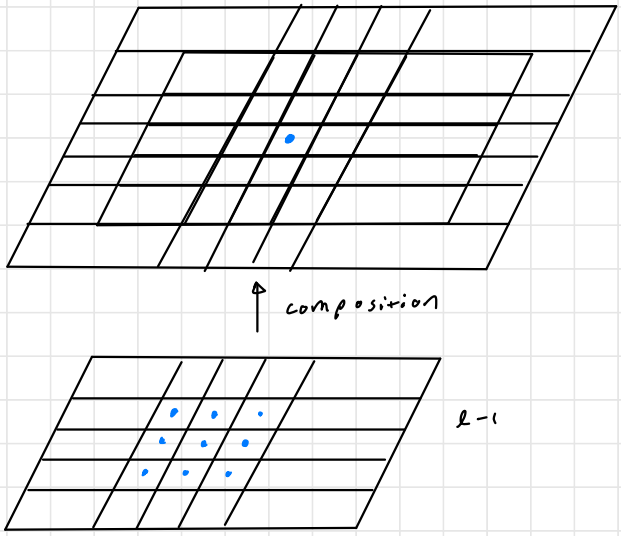
same $\forall x$



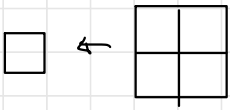
- $W_{\Delta x}^l$: kernel 3×3

- channel :



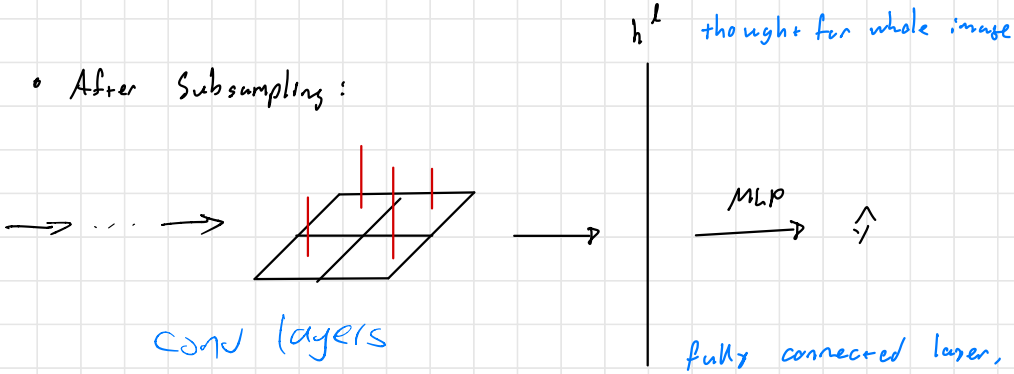


- Subsampling : $h_x^{l,sub} = h_{2x}^{l}$

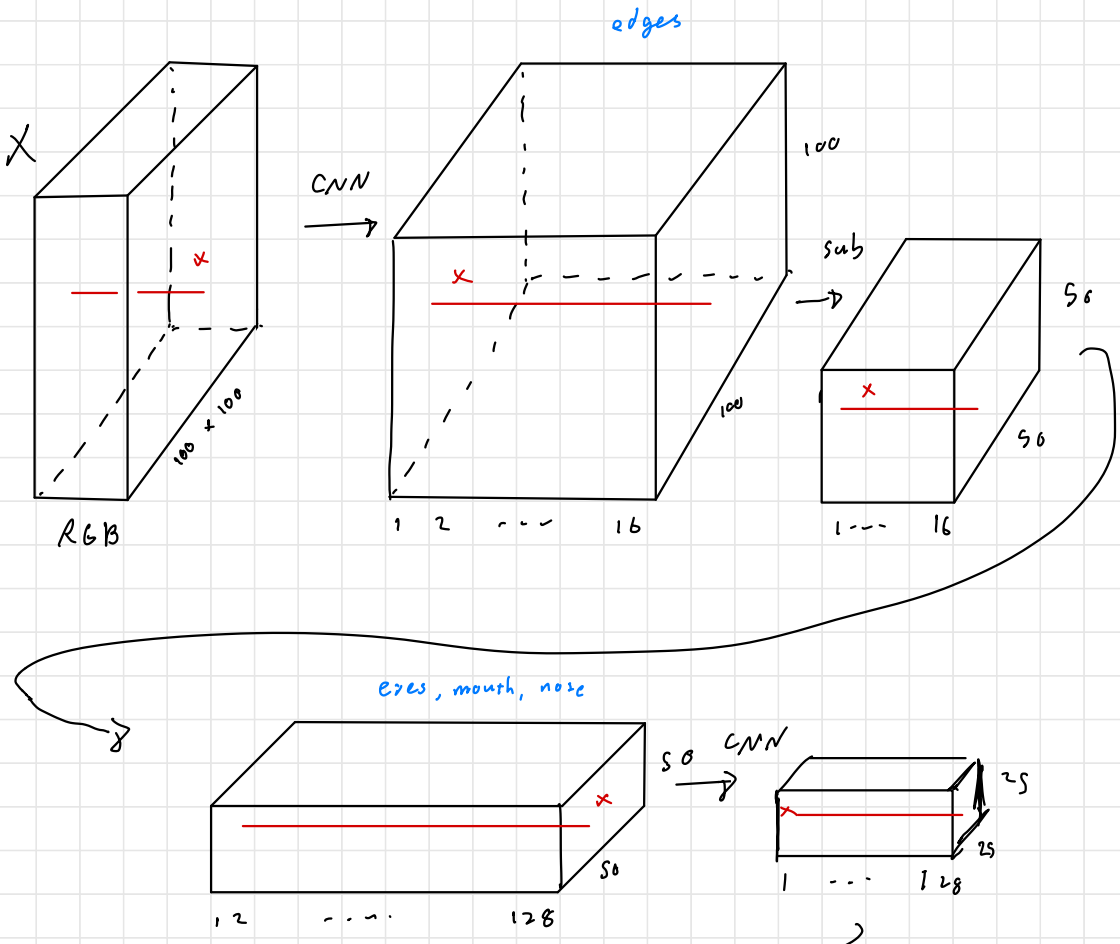


- Can also do Averaging / Max Pooling

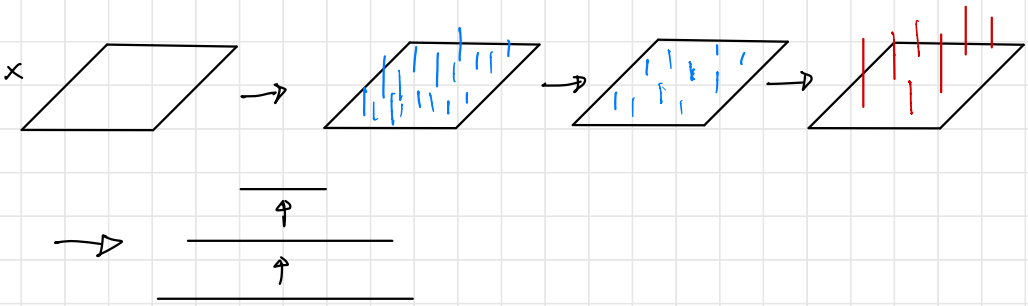
- After Subsampling :



$$h^l = \sigma\left(\sum_x w_x h_x^{l-1} + b^l\right)$$

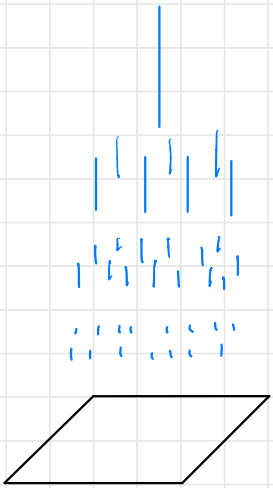

edges
X
CNN
100
sub
56
100 x 100
RGB
1 2 ... 16
1 ... 16
eyes, mouth, nose
50
CNN
25
25
128
sub
2
MLP
Softmax
4056
4056
1000
p
y one hot
face
abstract face
Loss function h classification
-log p(y|x)

- Plot in terms of vectors:

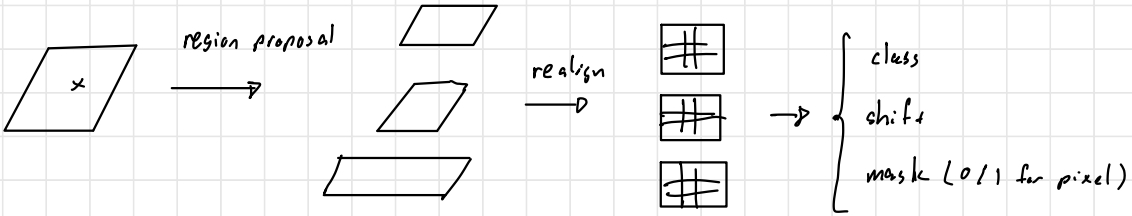


compose & abstract

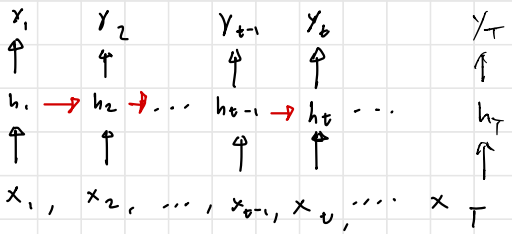
A team of vectors formation

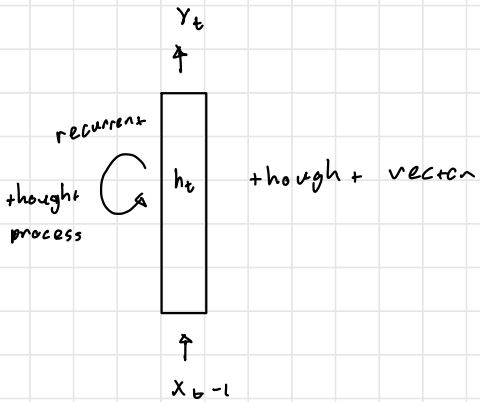


- detection:

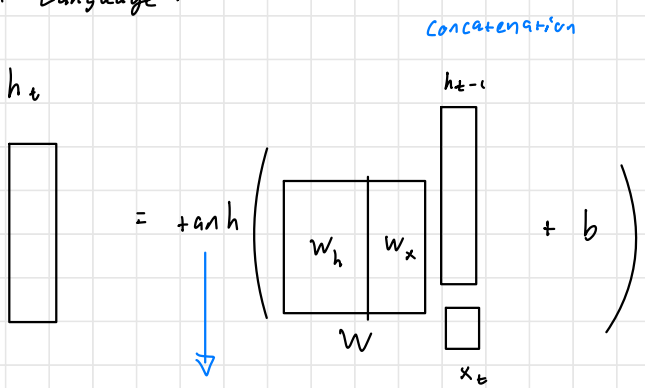


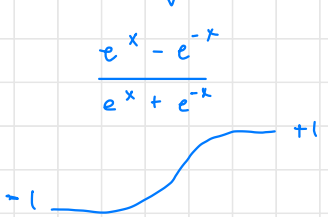
- Image : Spatial
- What about temporal?
  - Recurrent Neural Network (RNN)





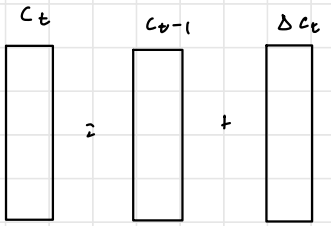
- Math Language :





- long term memory : in $\theta = (w, b, \forall \text{ layers})$
- short term memory : in h_t
- long short-term memory : (LSTM)

memory cells



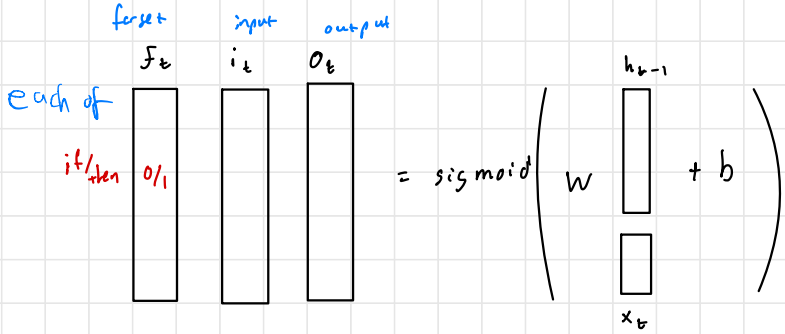
(residual)

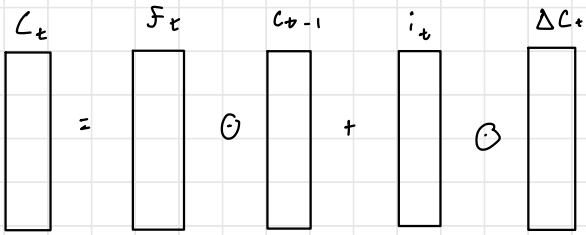
Δc_t

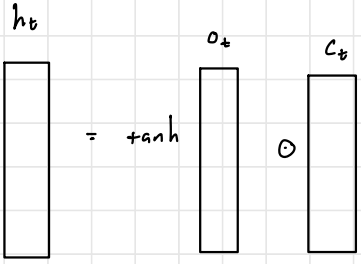
$$\square = \tanh\left(W \begin{pmatrix} h_{t-1} \\ x_t \end{pmatrix} + b \right)$$

$$h_t = \tanh\left(\square_{c_t} \right)$$

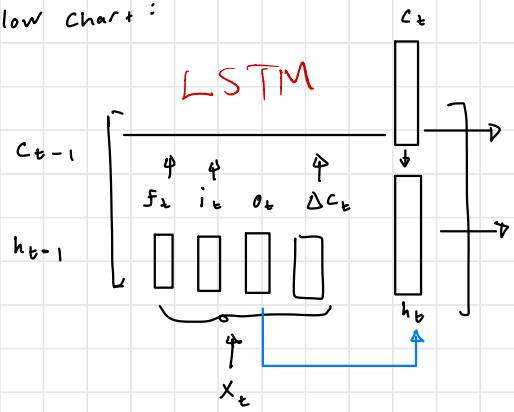
- Gates:

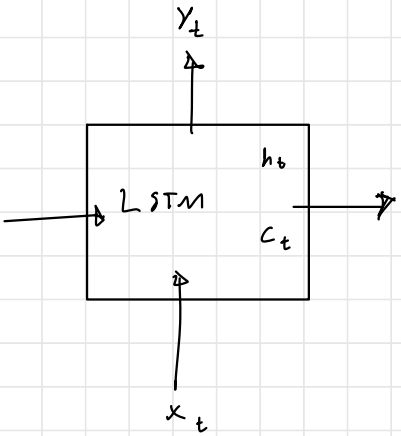

forget
input
output
f_t
i_t
o_t
each of
if/then 0/1
= sigmoid
W
+ b
h_{t-1}
x_t


C_t
f_t
C_{t-1}
i_t
ΔC_t
=
⊙
+
⊙


h_t
o_t
C_t
= tanh
⊙

- Flow chart :





loss function

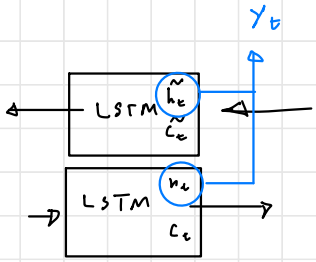
h_t abstract memory

$$-\sum_t \log p_\theta(y_t \mid x_{\le t})$$

regression

classification

- Bidirectional LSTM :



Sequential