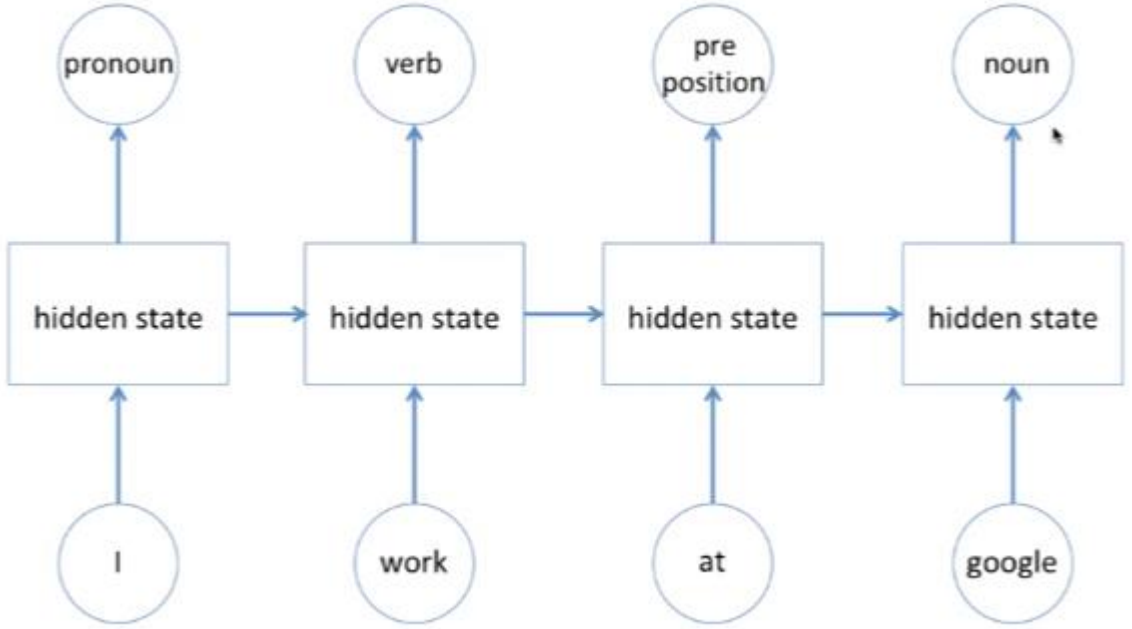
RNN


#nav_links {
width: 50%;
float: right;
}
#section_1{
width: 100%;
height: 500px;
box-sizing: border-box;
border: 1px solid red;
background-color: #f0f3f7;
position: relative;
}
#purple-box{
position: absolute;
width: 50%;
height: 1700px;
top: 50px;
right:0px;
border-radius:50px;
}
#other_links {
float: left;
list-style-type: none;
}
#other_links li {
display: inline;
padding: 20px;
}
#sign_in_up {
float: right;
list-style-type: none;
}
Line 27, Column 13 — 76 Lines
acer

What is RNN?

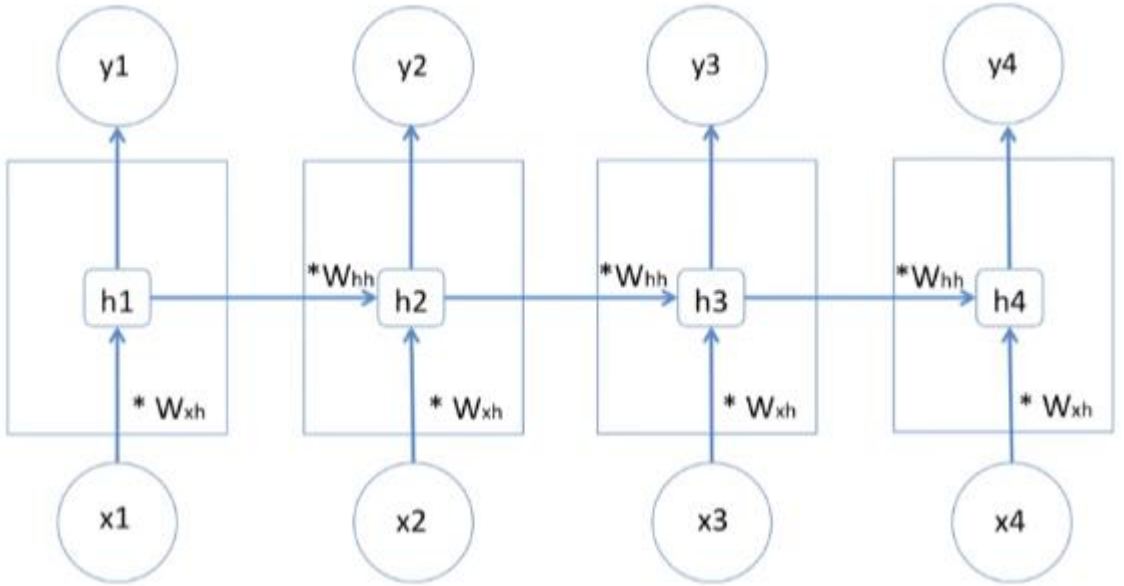
Recurrent Neural Network (RNN)



I work at google

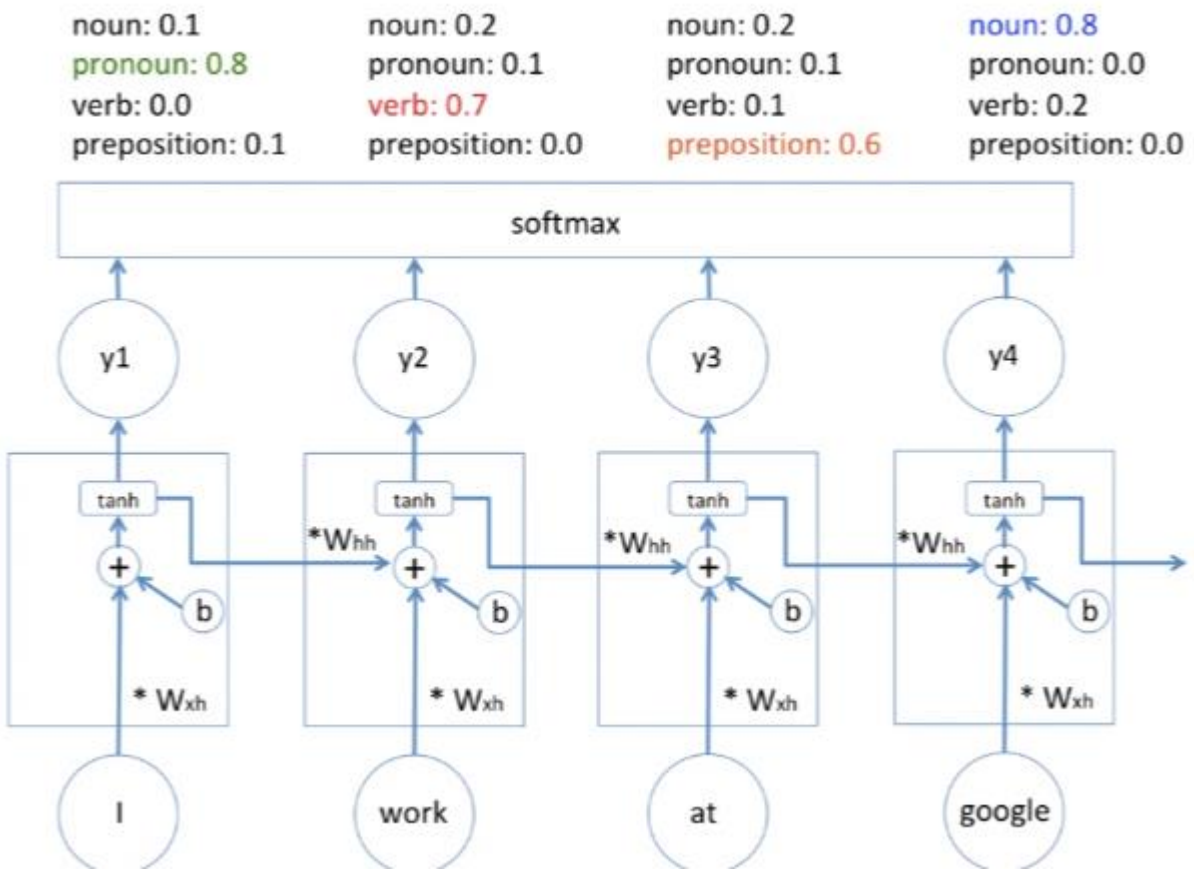
pronoun verb preposition noun

What is RNN?



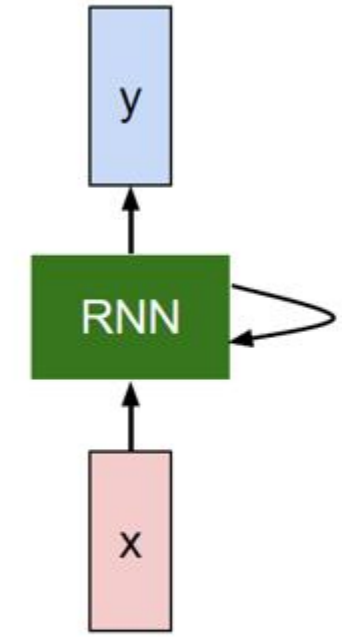
$$W_{hh} * (x_2) + W_{xh} * (x_1) + b$$

What is RNN?

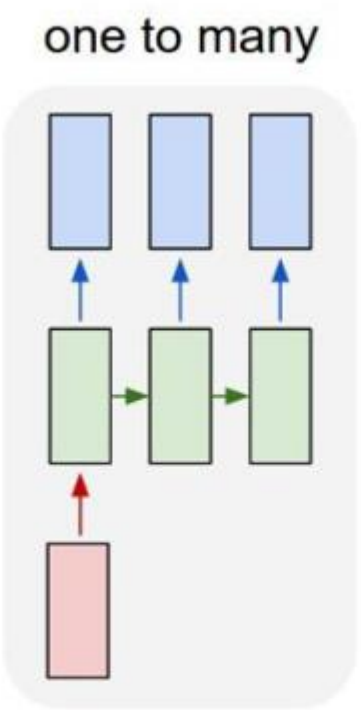

noun: 0.1
pronoun: 0.8
verb: 0.0
preposition: 0.1
noun: 0.2
pronoun: 0.1
verb: 0.7
preposition: 0.0
noun: 0.2
pronoun: 0.1
verb: 0.1
preposition: 0.6
noun: 0.8
pronoun: 0.0
verb: 0.2
preposition: 0.0
softmax
y1
y2
y3
y4
tanh
tanh
tanh
tanh
*Whh
*Whh
*Whh
+
+
+
+
b
b
b
b
* Wxh
* Wxh
* Wxh
* Wxh
I
work
at
google

What is RNN?

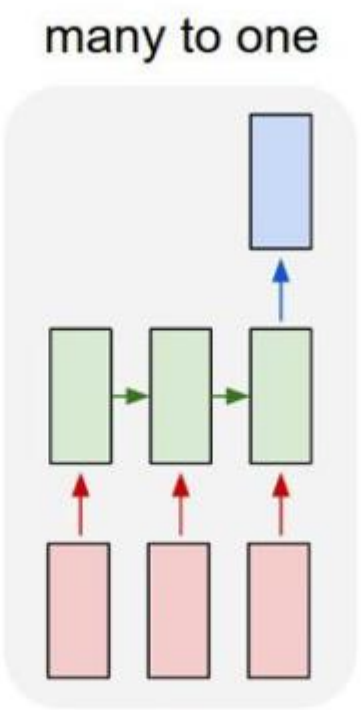
Recurrent Neural Network (RNN)



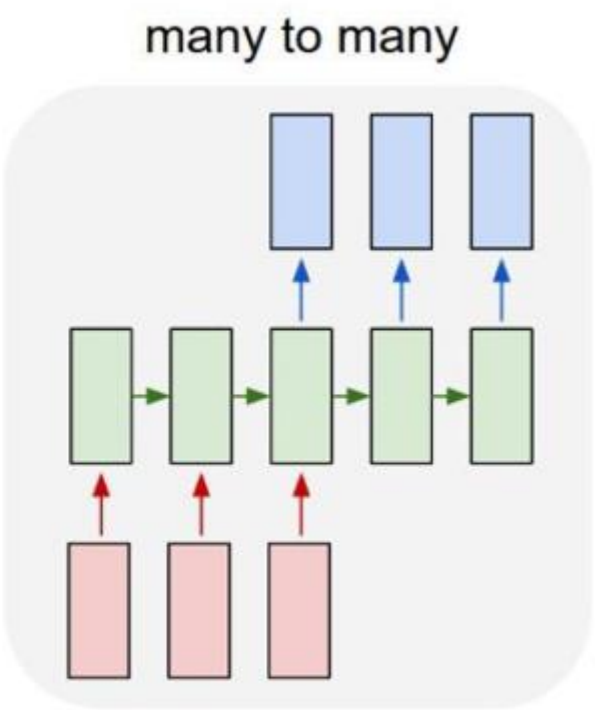
<기본구조>



<사진설명 붙이기>
사진 → 단어들

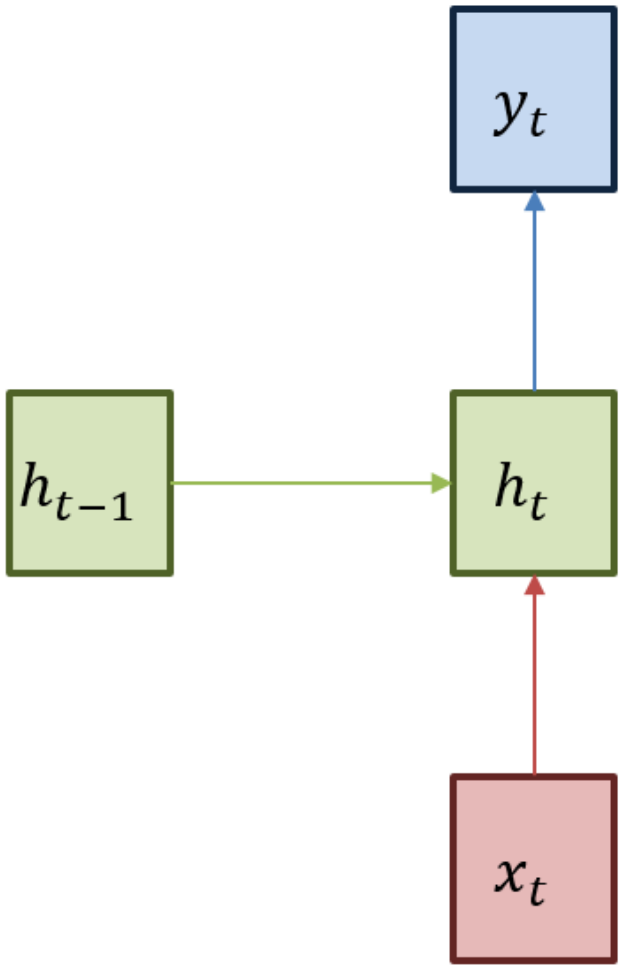


<감성분석>
단어들 → 감성점수



<번역>
단어들 → 단어들

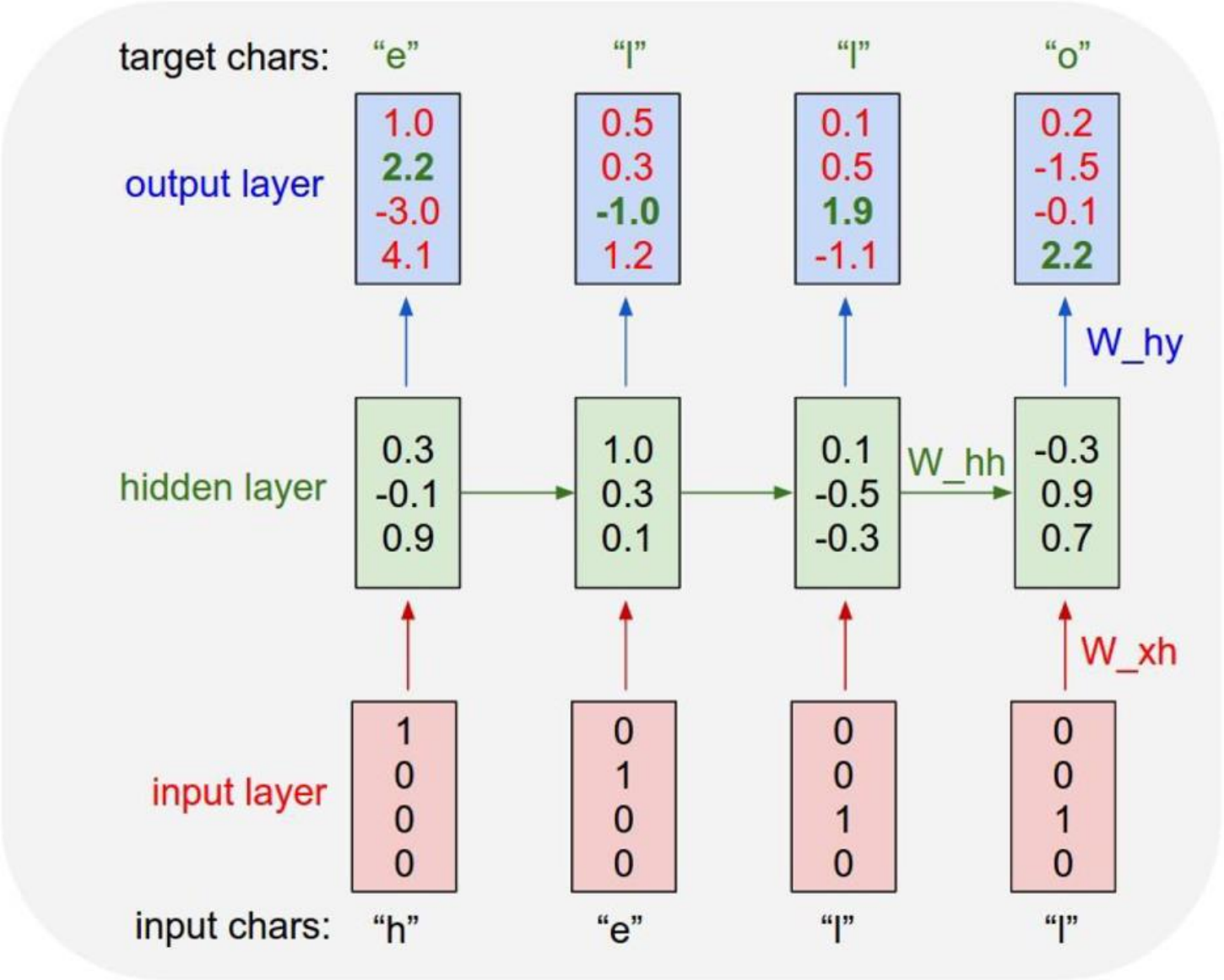
What is RNN?



$$y_t = W_{hy} h_t + b_y$$

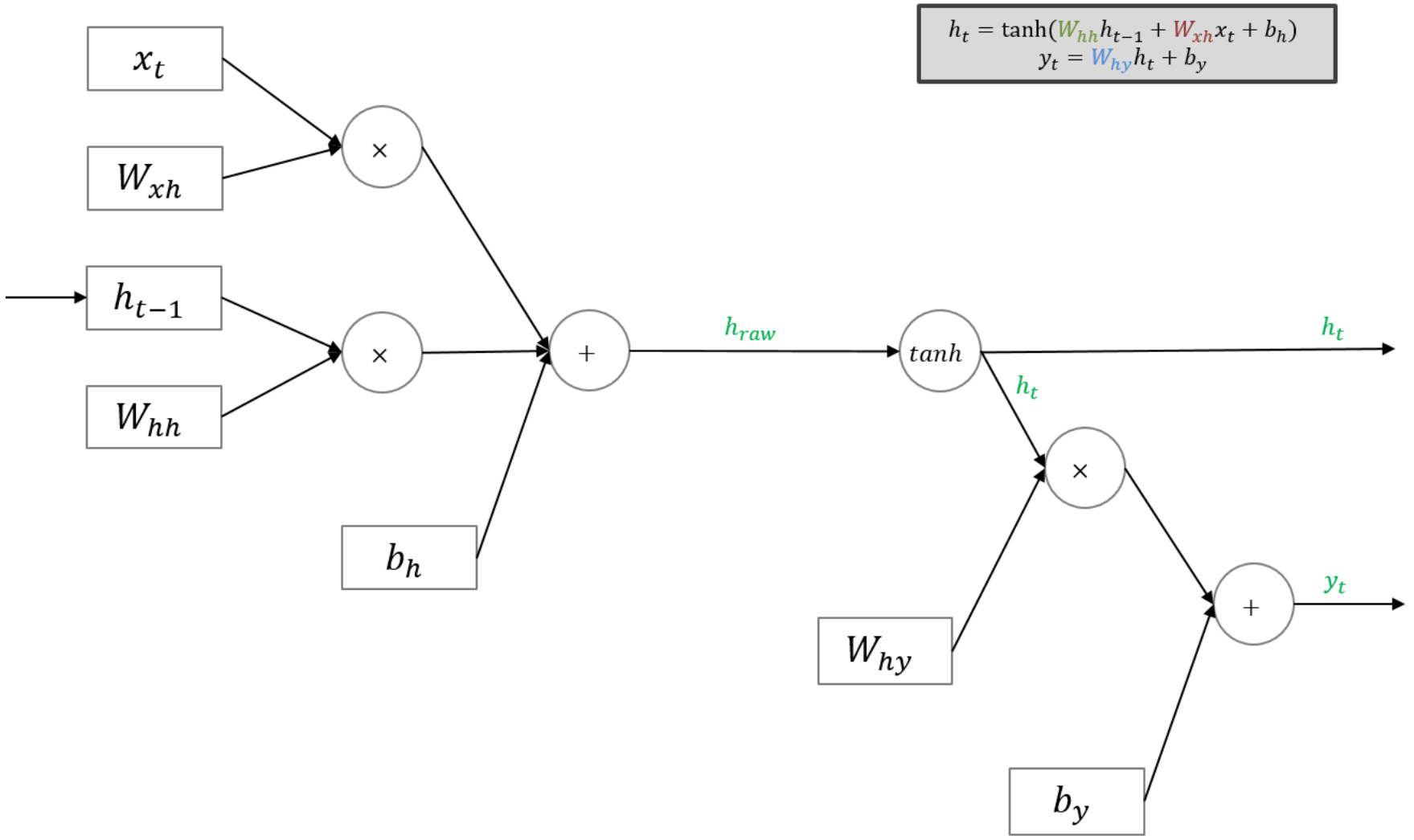
$$h_t = \tanh(W_{hh} h_{t-1} + W_{xh} x_t + b_h)$$

What is RNN?

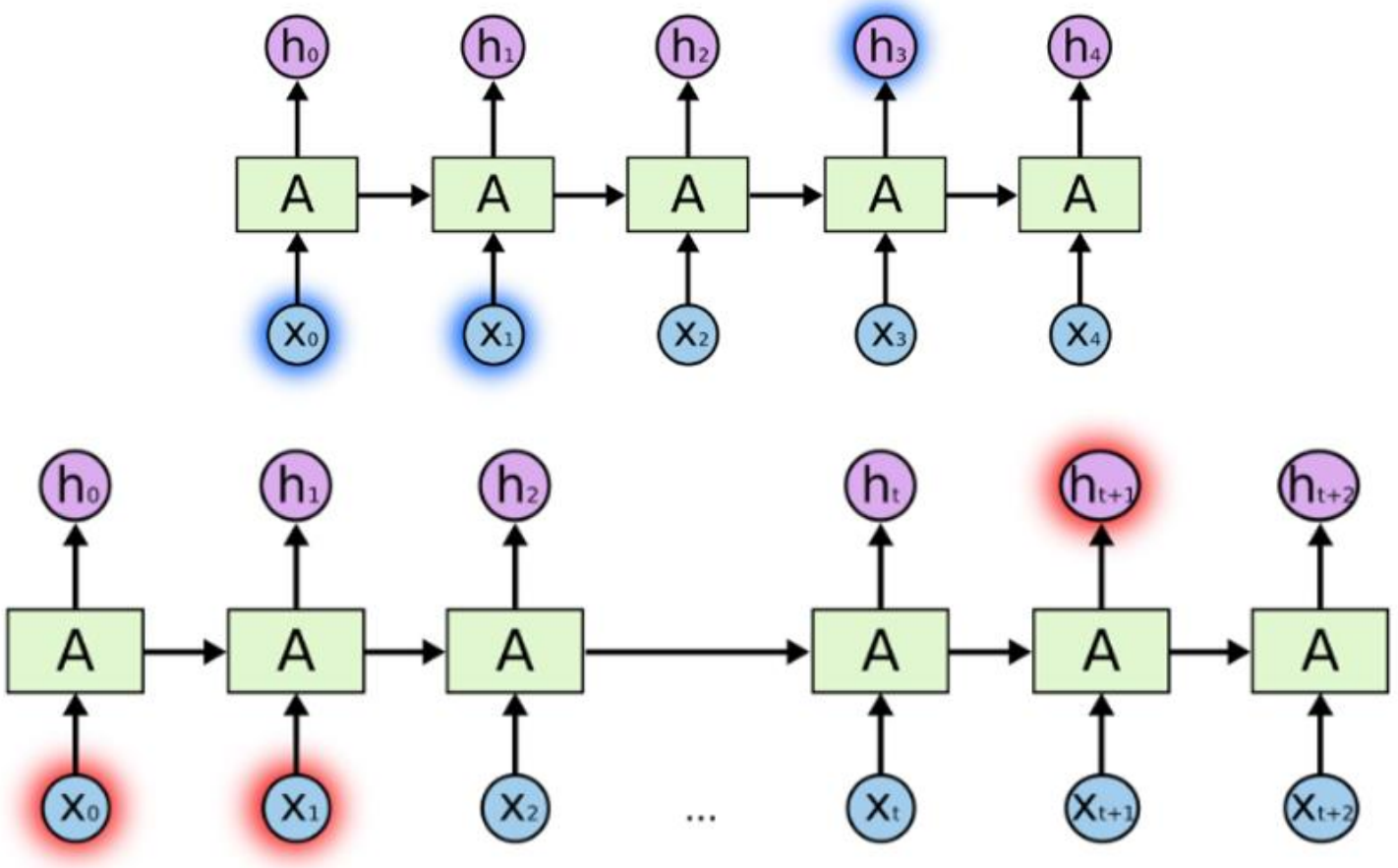


다음 글자를 예측하는 모델 만들기

What is RNN?


$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t + b_h)$
$y_t = W_{hy}h_t + b_y$
x_t
W_{xh}
h_{t-1}
W_{hh}
b_h
×
×
+
h_{raw}
tanh
h_t
h_t
×
W_{hy}
+
y_t
b_y

Limitation of RNN

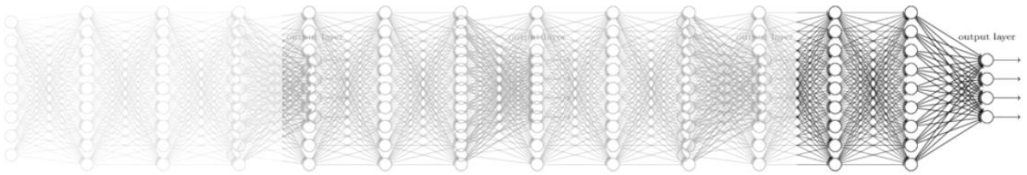


<관련 정보와 그 정보를 사용하는 지점 사이 거리가 멀 경우 RNN 학습능력 저하>

Vanishing gradient problem.

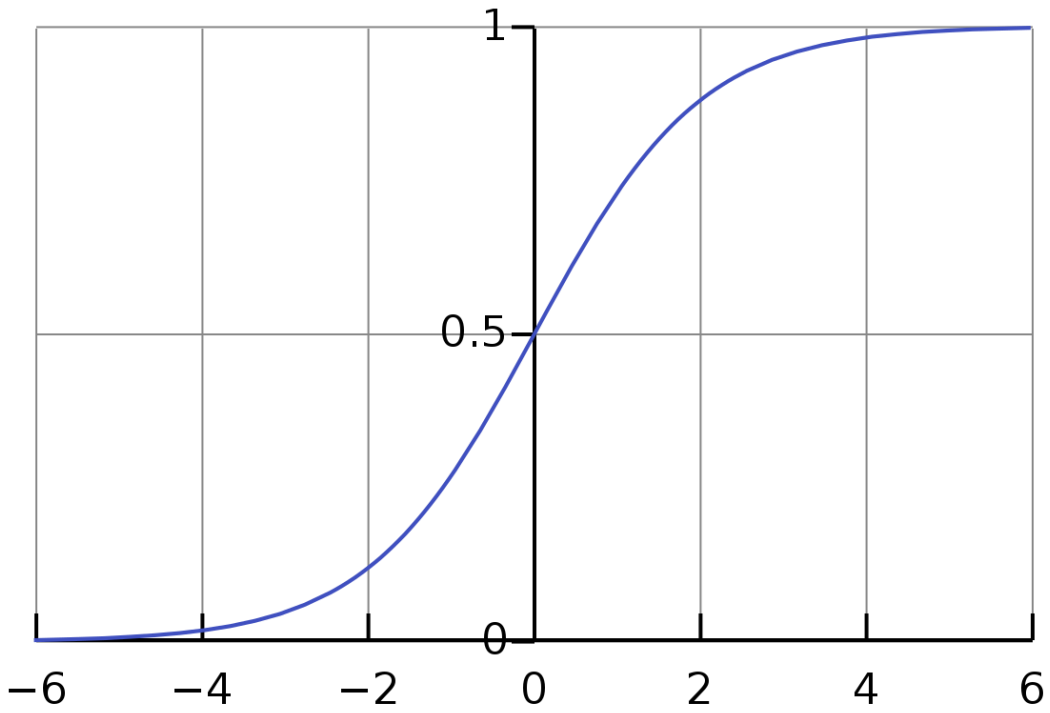
Vanishing gradient problem

Vanishing gradient (NN winter2: 1986-2006)

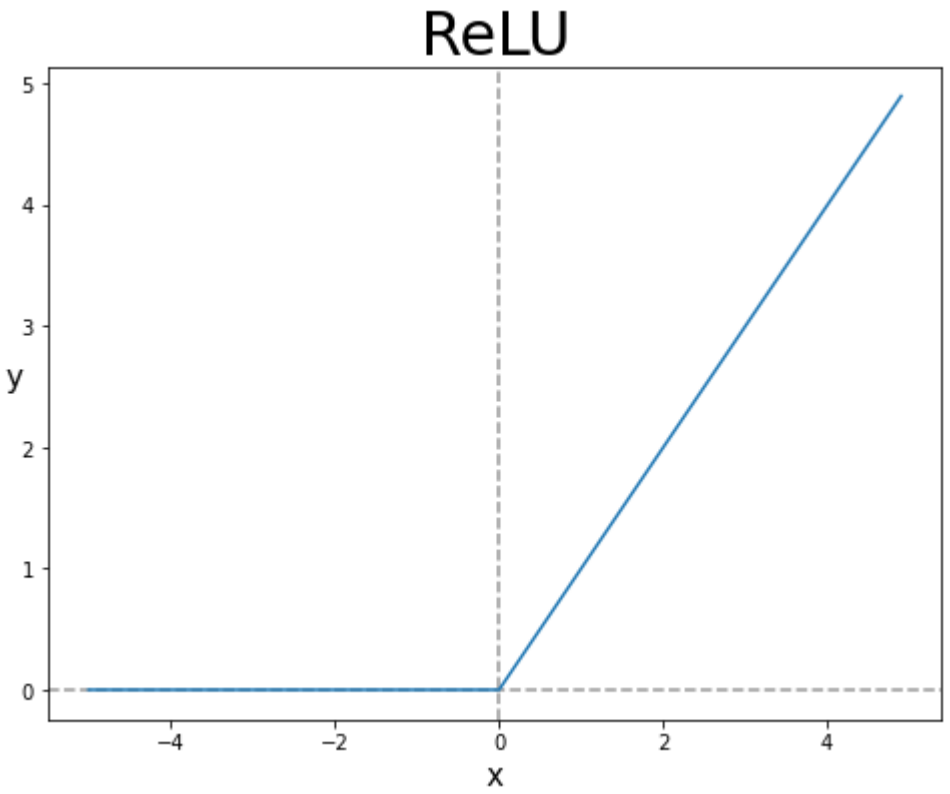


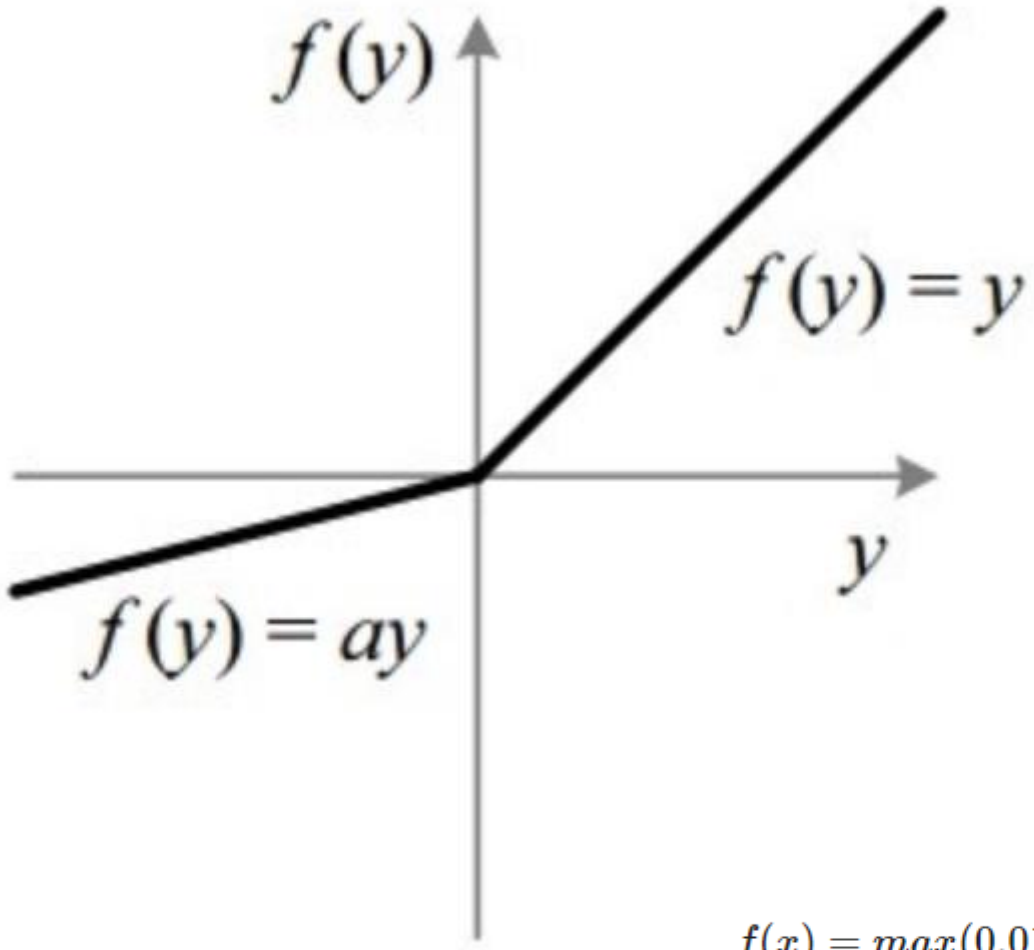
$$sigmoid(x) = \frac{1}{1 + e^{-x}}$$

$$W \leftarrow W - \eta \frac{\partial L}{\partial W}$$



Solving

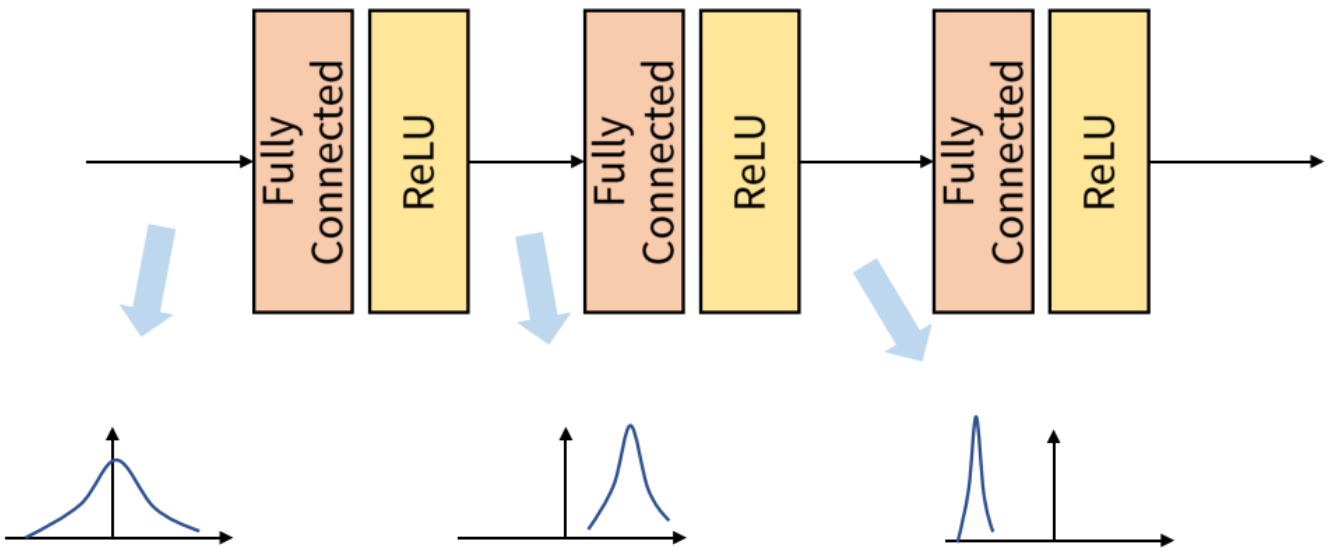




$$f(x) = max(0.01x, x)$$

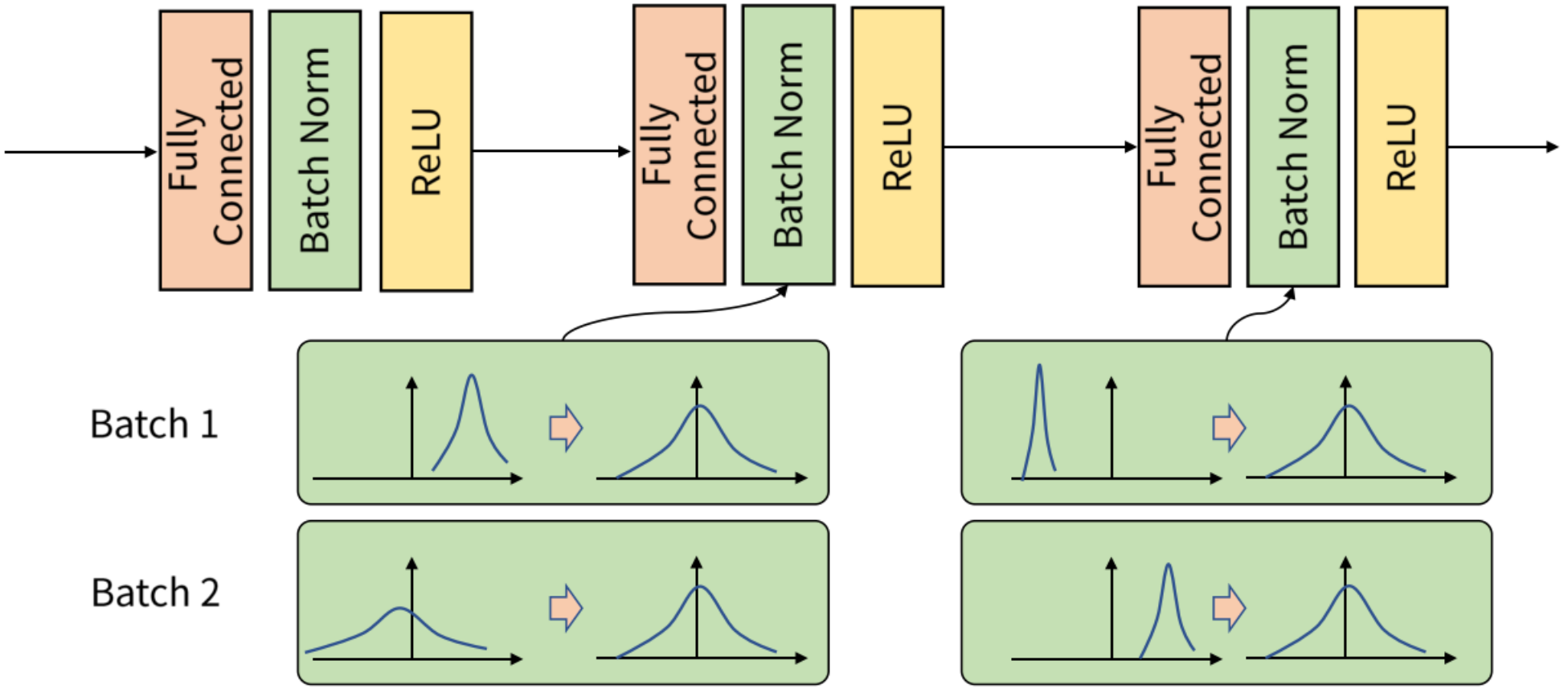
Leaky relu

Batch normalization



Internal covariate shift 문제 발생. 즉 배치마다 데이터의 분포가 달라지기 때문에, 학습이 불안정하게 진행된다.

Batch normalization



Batch normalization

Input: Values of x over a mini-batch: $\mathcal{B} = \{x_{1...m}\}$;
Parameters to be learned: γ, β
Output: $\{y_i = \mathrm{BN}_{\gamma,\beta}(x_i)\}$

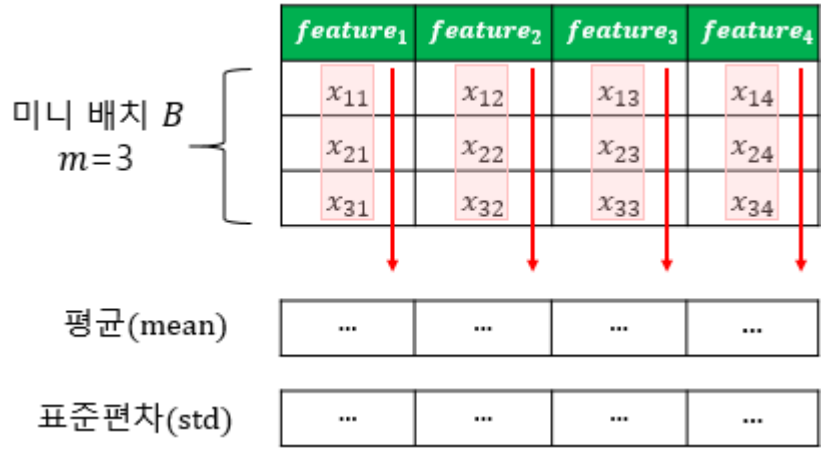
$$\mu_{\mathcal{B}} \leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i \qquad \text{// mini-batch mean}$$

$$\sigma^2_{\mathcal{B}} \leftarrow \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_{\mathcal{B}})^2 \qquad \text{// mini-batch variance}$$

$$\widehat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma^2_{\mathcal{B}} + \epsilon}} \qquad \text{// normalize}$$

$$y_i \leftarrow \gamma\widehat{x}_i + \beta \equiv \mathrm{BN}_{\gamma,\beta}(x_i) \qquad \text{// scale and shift}$$

Algorithm 1: Batch Normalizing Transform, applied to activation x over a mini-batch.



Batch normalization

Input: Values of x over a mini-batch: $\mathcal{B} = \{x_{1...m}\}$;
Parameters to be learned: γ, β
Output: $\{y_i = \mathrm{BN}_{\gamma,\beta}(x_i)\}$

$$\mu_{\mathcal{B}} \leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i \qquad \text{// mini-batch mean}$$

$$\sigma^2_{\mathcal{B}} \leftarrow \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_{\mathcal{B}})^2 \qquad \text{// mini-batch variance}$$

$$\widehat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma^2_{\mathcal{B}} + \epsilon}} \qquad \text{// normalize}$$

$$y_i \leftarrow \gamma\widehat{x}_i + \beta \equiv \mathrm{BN}_{\gamma,\beta}(x_i) \qquad \text{// scale and shift}$$

Algorithm 1: Batch Normalizing Transform, applied to activation x over a mini-batch.