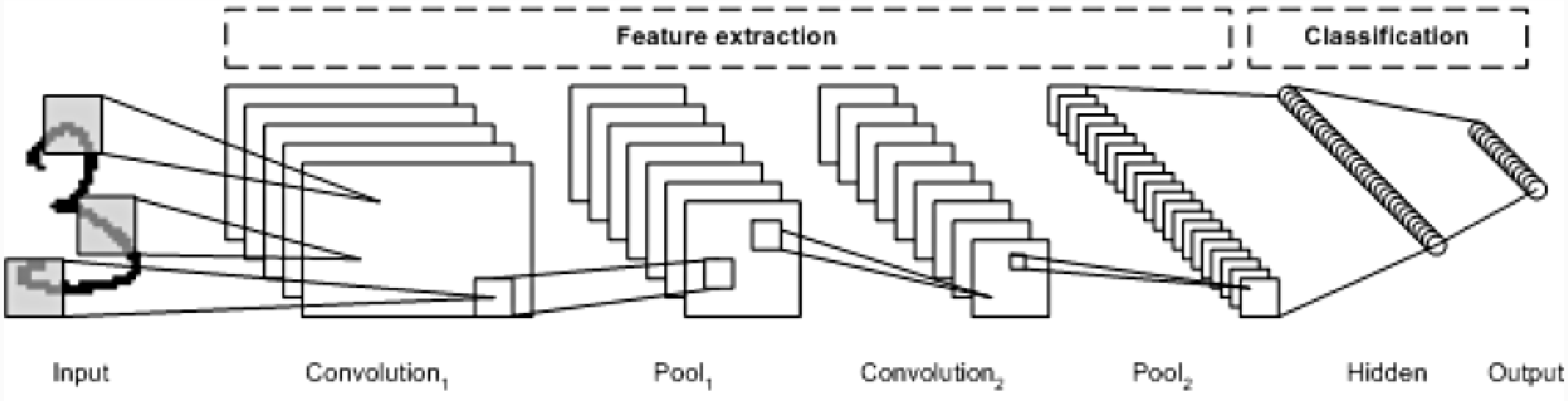
RNN

정인호

CNN 복습



CNN의 경우 convolutional technic을 이용해 Input을 효율적으로 튜닝

➡ Neural Network의 구조는 바꾸지 않고 Input만을 활용해 이미지(영상)을 학습

Q. 같은 방식으로 시계열 데이터를 학습할 수 있을까?

A Sequence Modeling Problem

Predict the Next Word

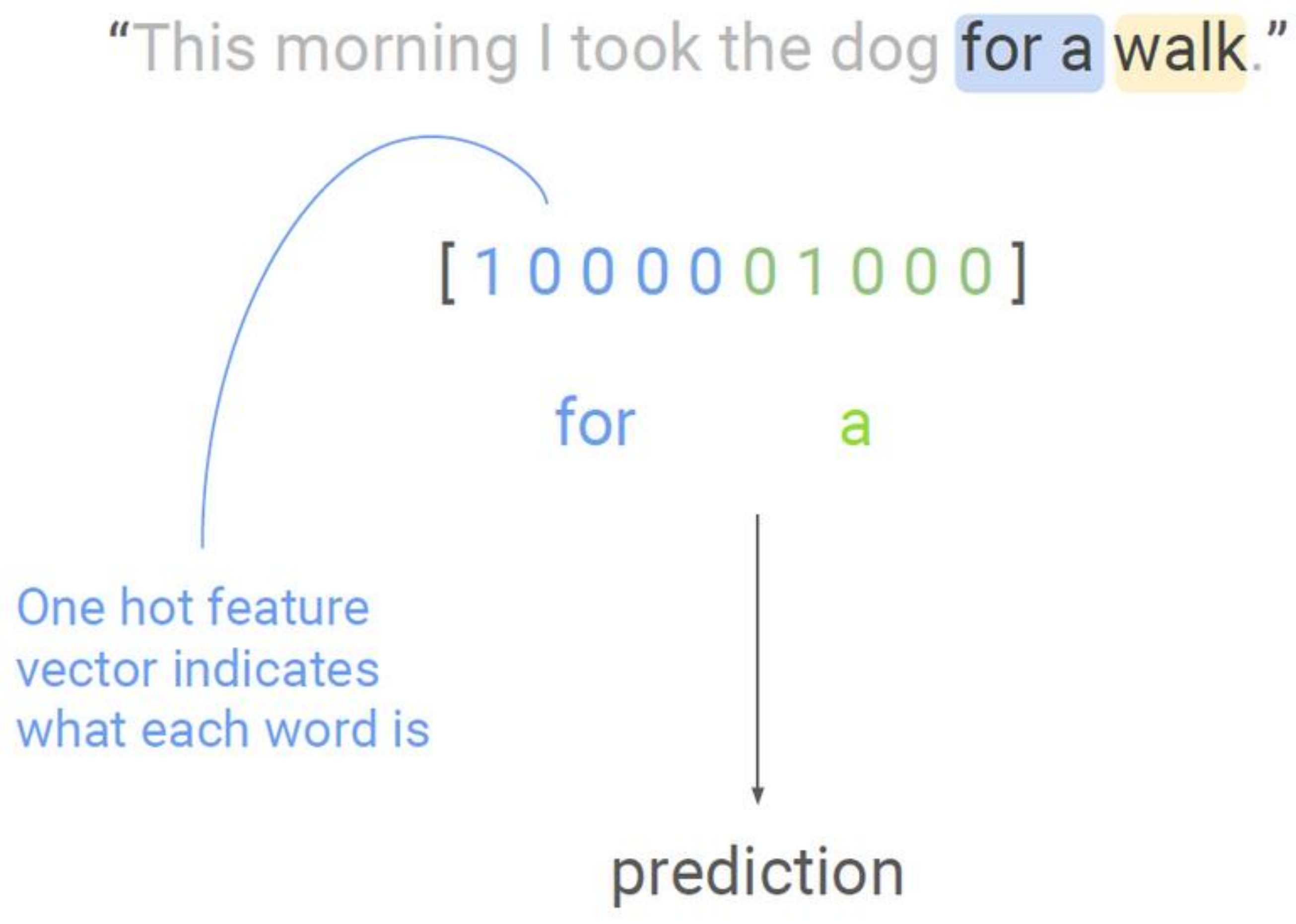
“This morning I took my dog for a walk.”

“This morning I took my dog for a ___?___.”

Check) neural network의 input-size는 dynamic하게 작동하지 않음. 고정되어 있어야 함.

Idea 1 : Use a Fixed Window

근처 몇 개의 data를 가지고 예측해보자


"This morning I took the dog for a walk."
[1 0 0 0 0 0 1 0 0 0]
for
a
One hot feature vector indicates what each word is
prediction

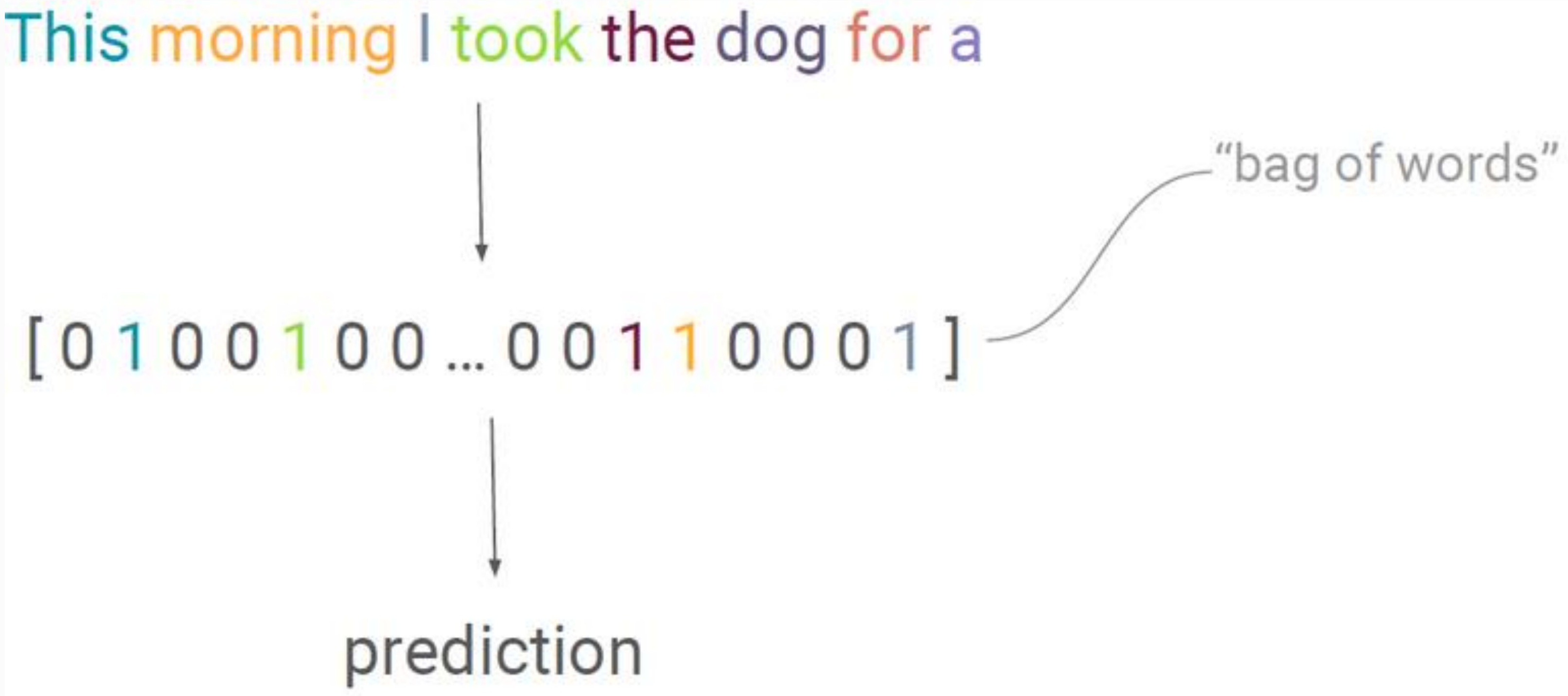
Problem : Can’t Model Long-Term Dependencies

“France is where I grew up, but I now live in Boston. I speak fluent ___ ? ___.”

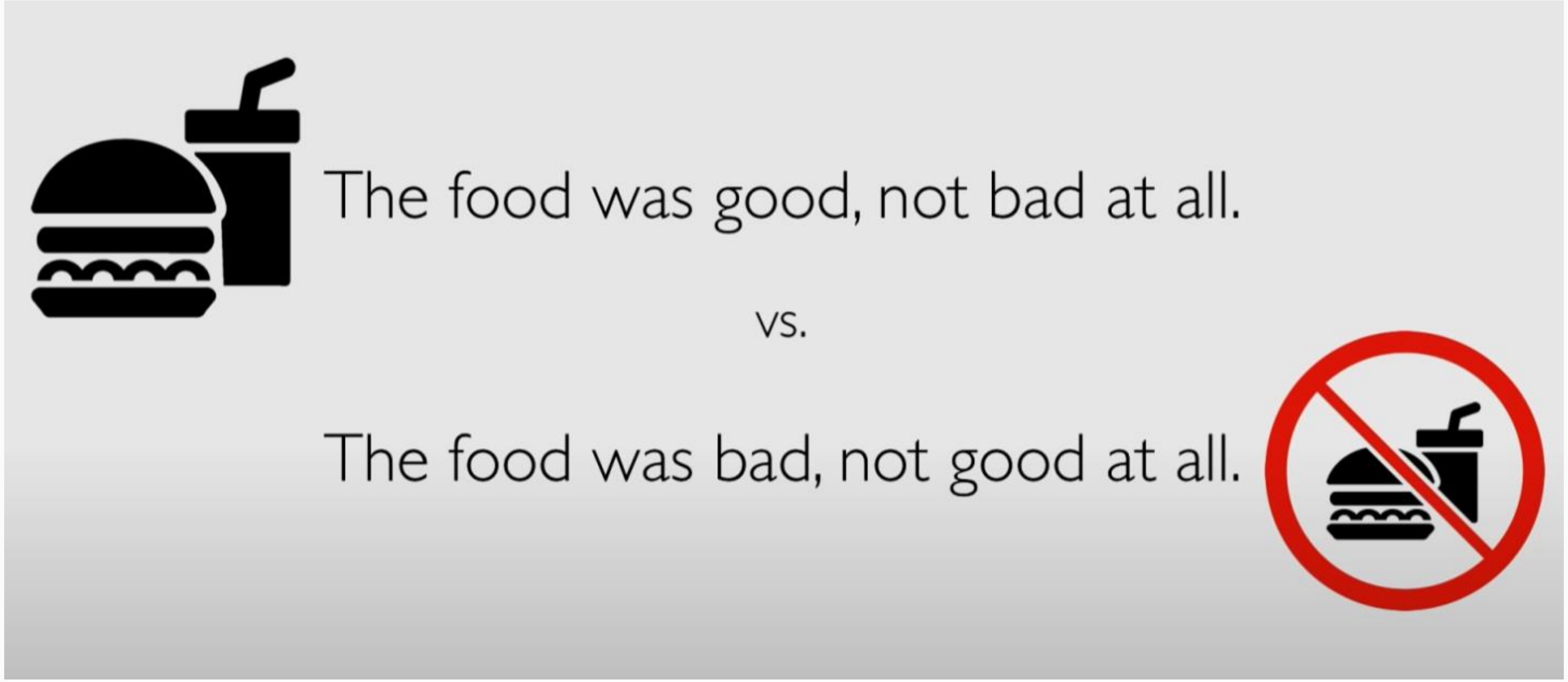
→ We need information from the distant past to accurately predict the correct word.

Idea 2 : Use Entire Sequence as Set of Counts

Sequence의 전체 정보를 얻기 위해 bag of words 방법을 이용하자



Problem : Counts Don’t Preserve Order



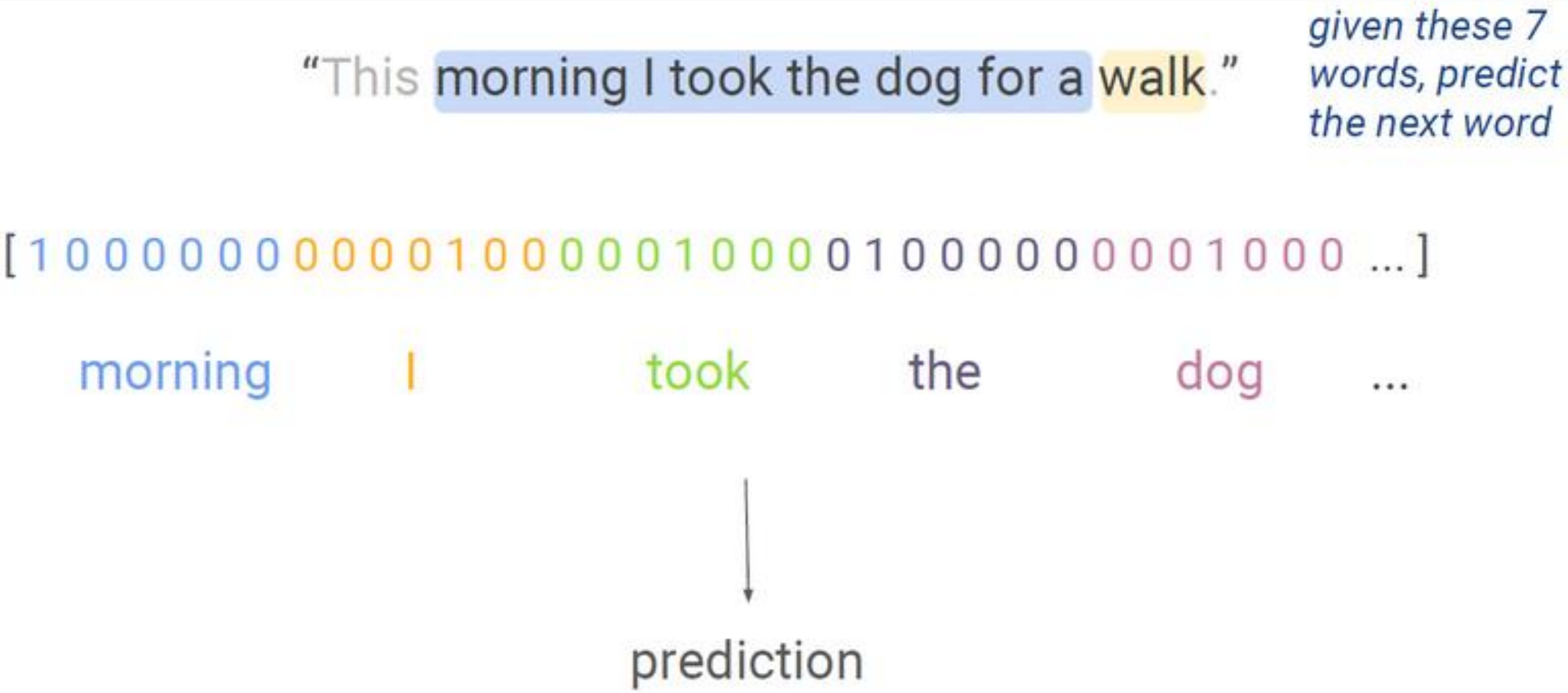
The food was good, not bad at all.

vs.

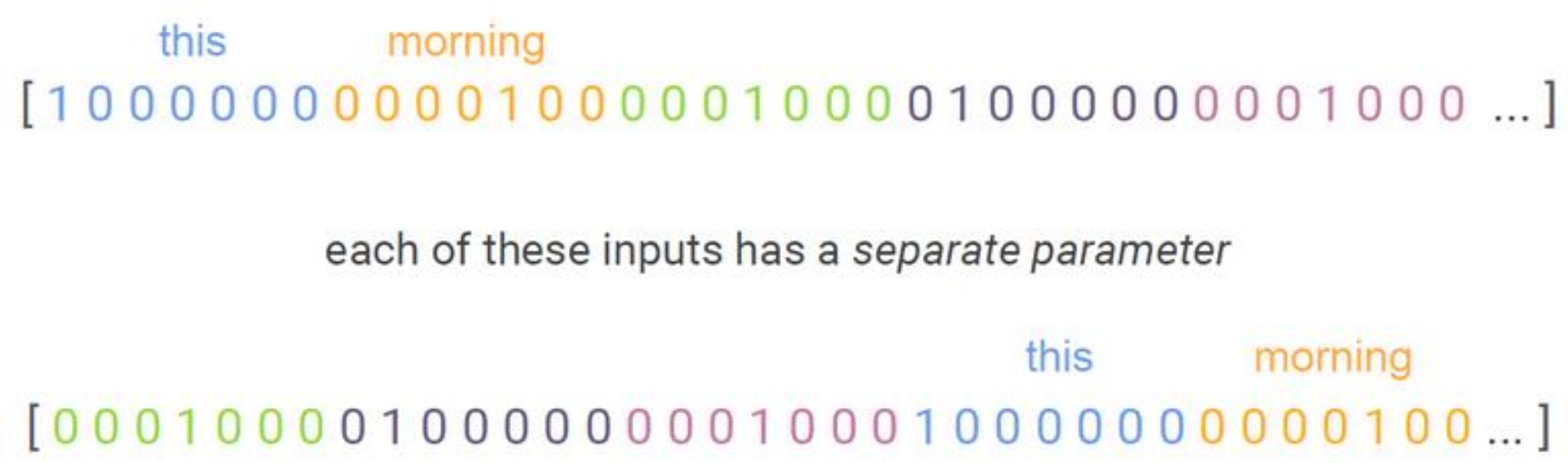
The food was bad, not good at all.

Idea 3 : Use a Really Big Fixed Window

(비효율적이라도) order와 counts를 모두 가진 매우 큰 window를 사용해보자.



Problem : No parameter Sharing



each of these inputs has a *separate parameter*

This morning의 경우 앞에 있나 뒤에 있나 문장의 의미는 바뀌지 않음.

하지만 network의 경우 완전히 새로운 input으로 인식.

Why? 각 vector에 연결된 뉴런이 주변의 parameter를 모름.

Sequence Modeling : Design Criteria

To Model Sequence, We need to

1. Handle variable-length sequence.

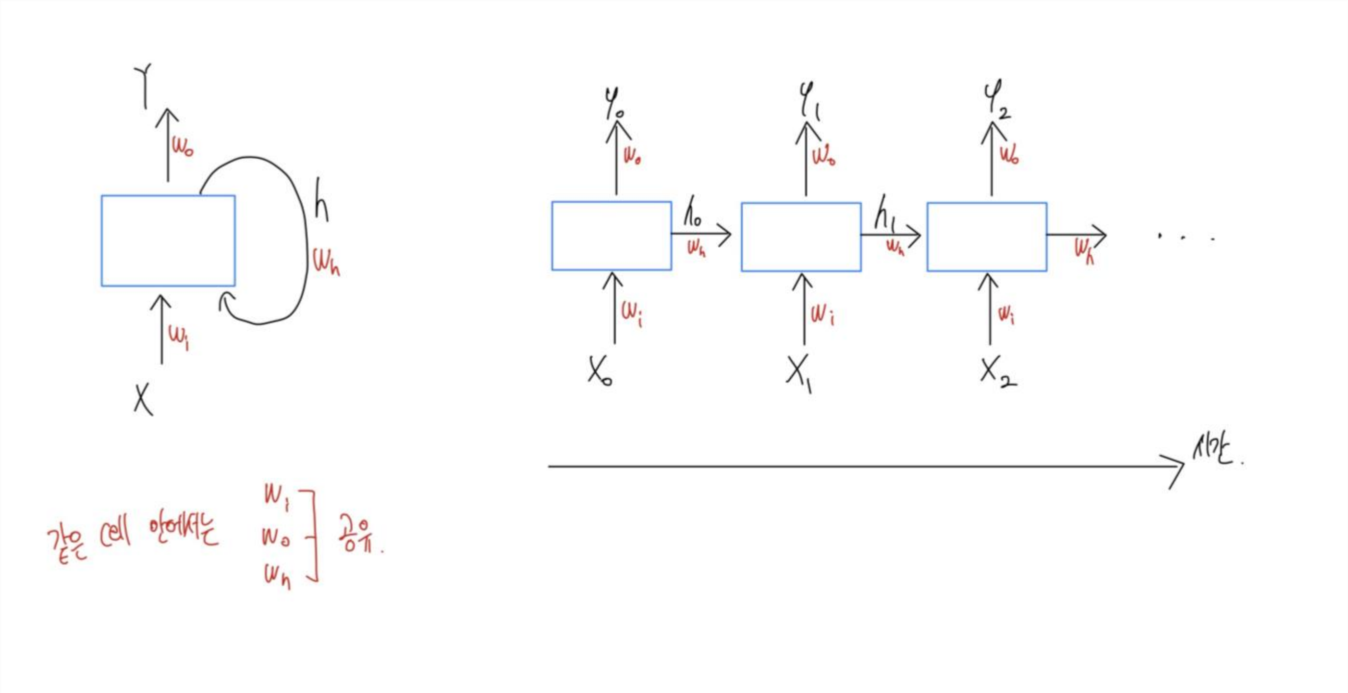
2. Track long-term dependencies.

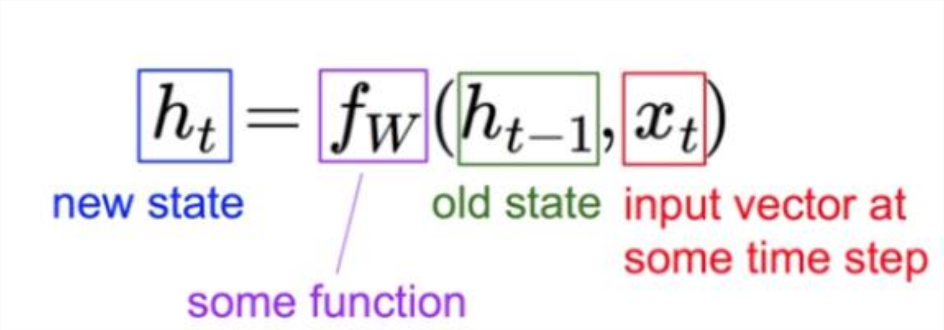
3. Maintain information about order.

4. Share parameters across the sequence.

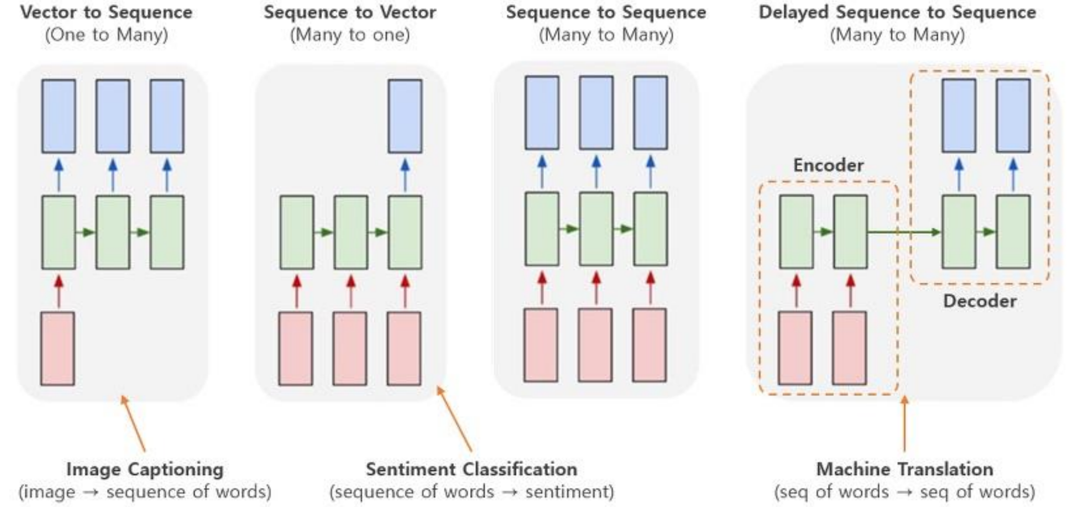
➡ RECURRENT NEURAL NETWORKS (RNNS)

RNN

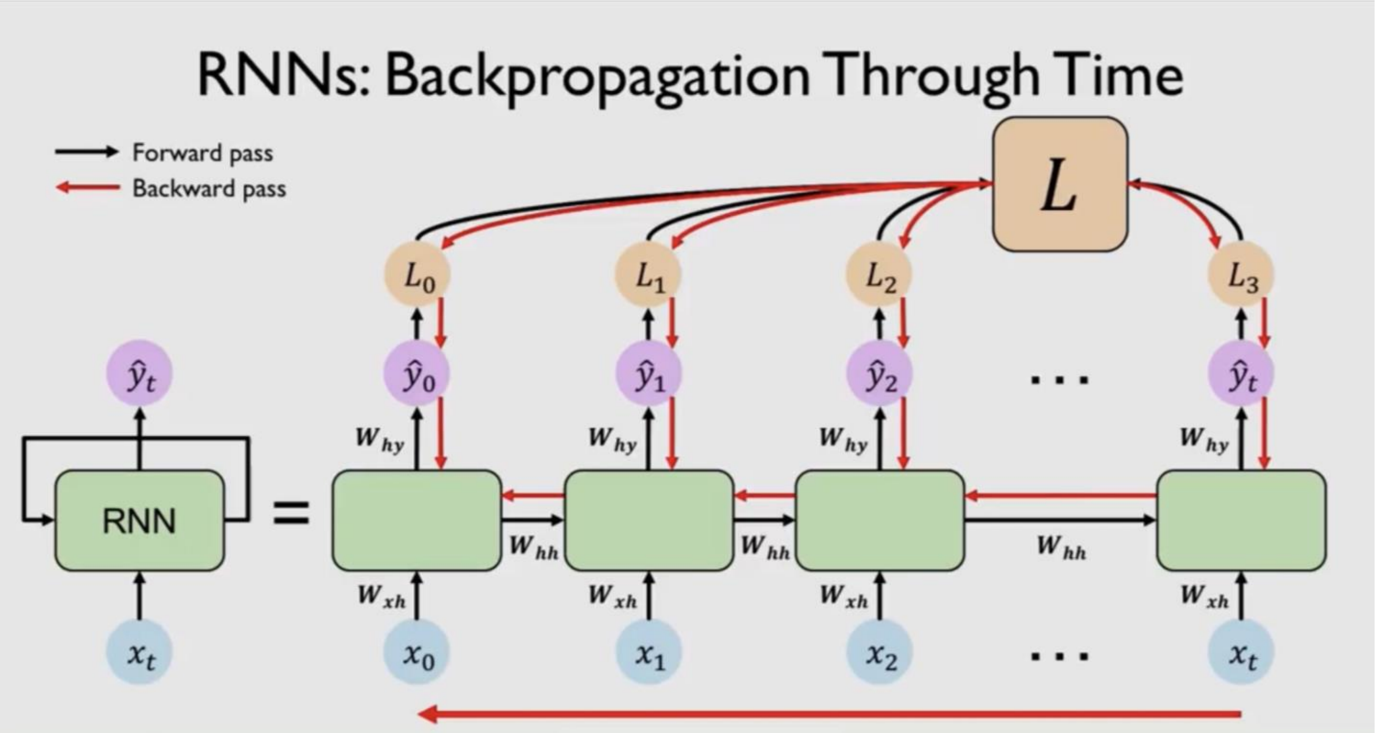




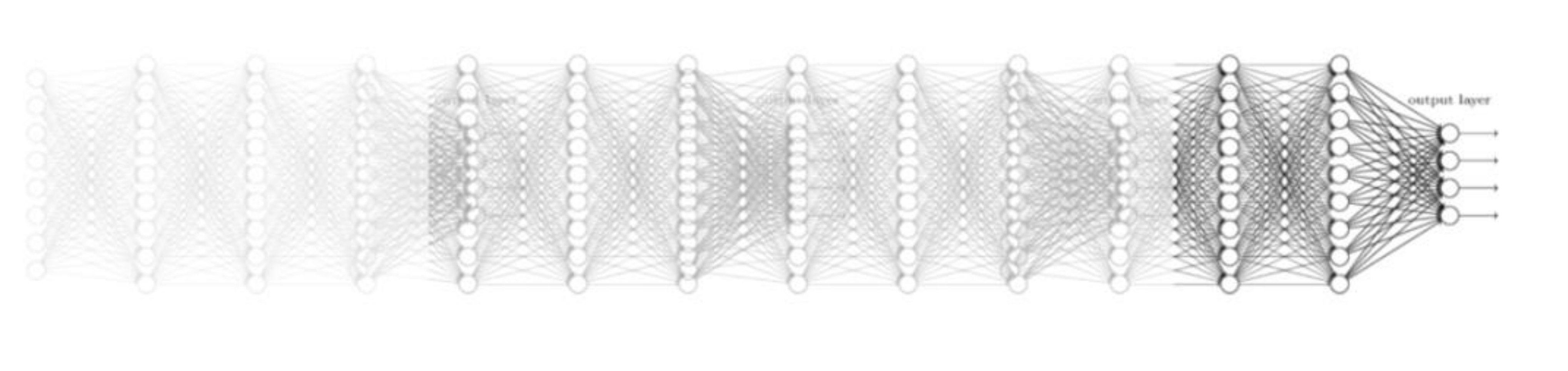
$$h_t = f_W(h_{t-1}, x_t)$$

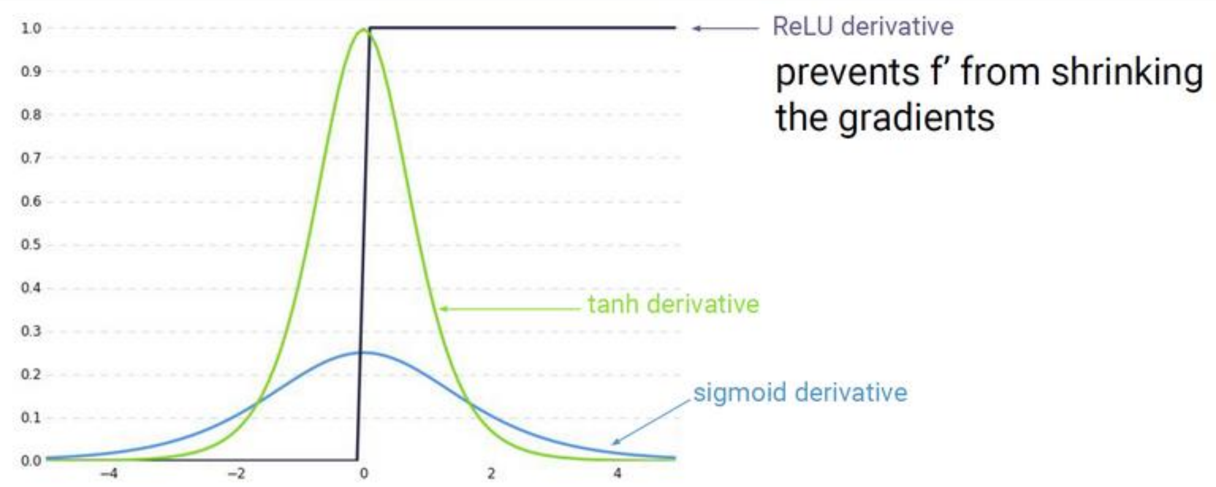


Backpropagation Through Time (BPTT)



Gradient Vanishing Problem in NN



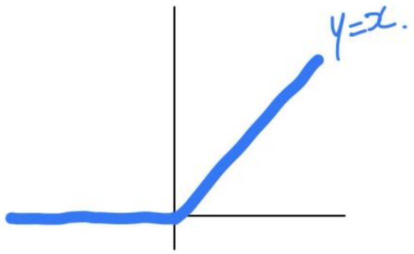


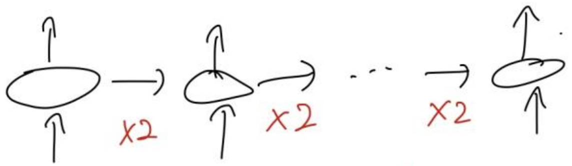
prevents f’ from shrinking the gradients

Gradient Vanishing(Exploding) in RNNs

If timestep =100 , $W_h = 2$?

in ReLu.





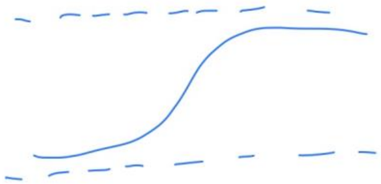
→ $\times 2^{100}$

"Value 폭주"

($\times (\frac{1}{2})^{100}$

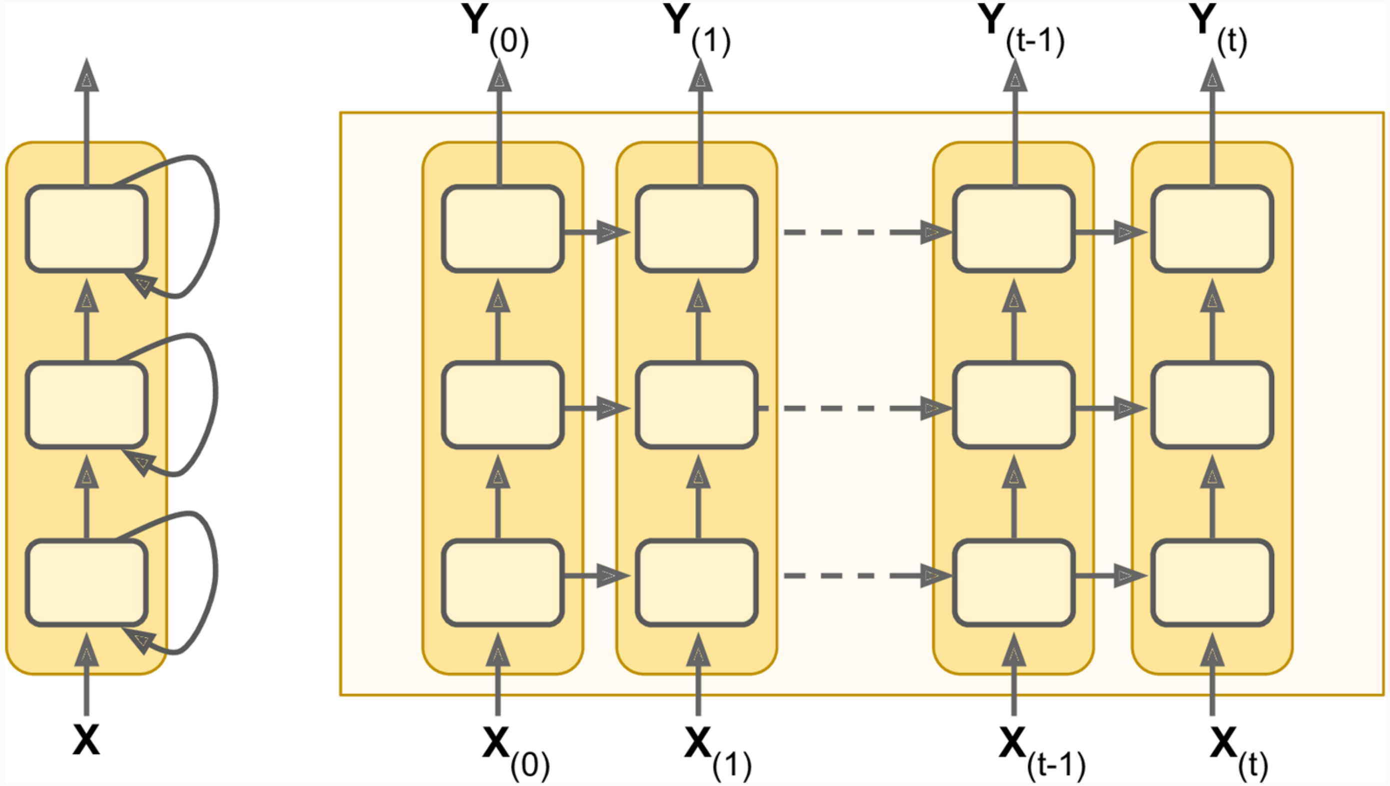
"Value 소실"

⇒ 폭주하지 않게 눌러주는 활성화 함수



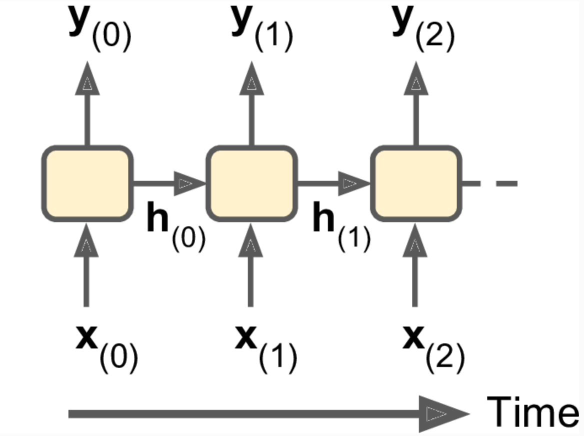
tanh 사용.

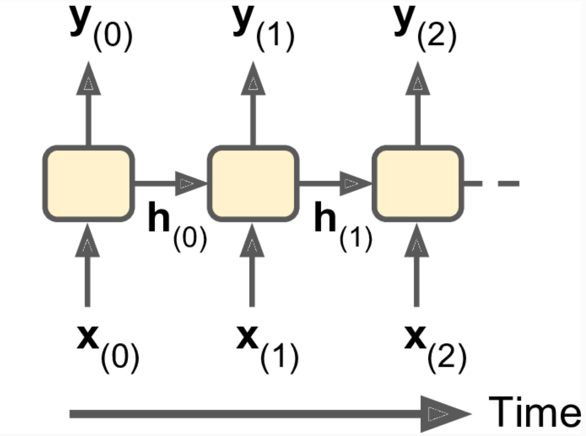
Deep RNN

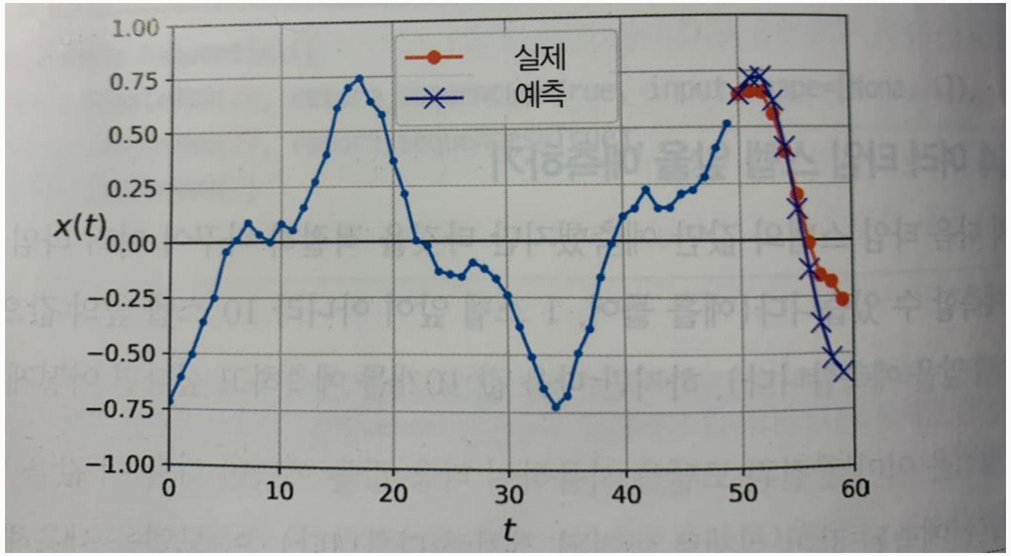


Predict Multiple Timesteps Ahead

Ex) 10 time-step

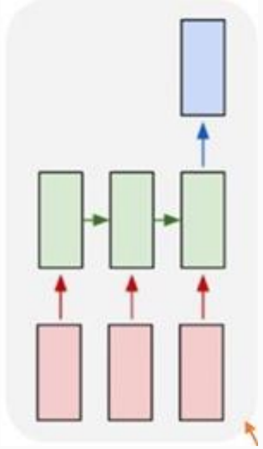


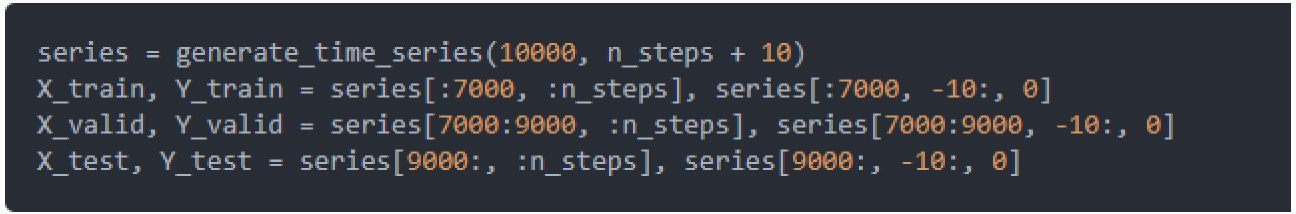




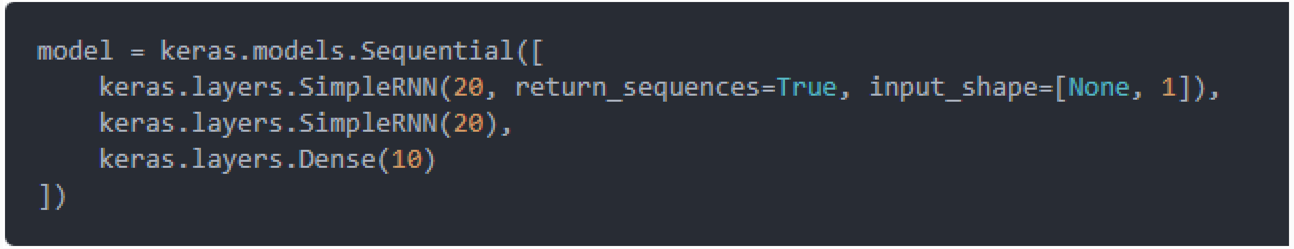
Predict Multiple Timesteps Ahead

Ex) 10 time-step

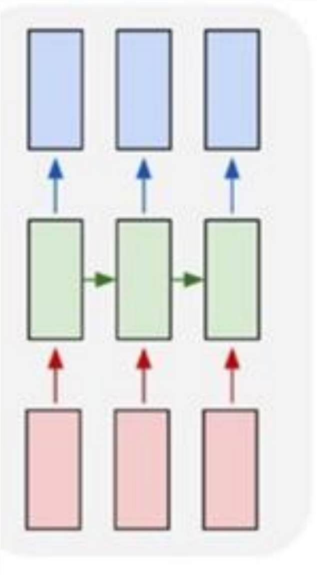


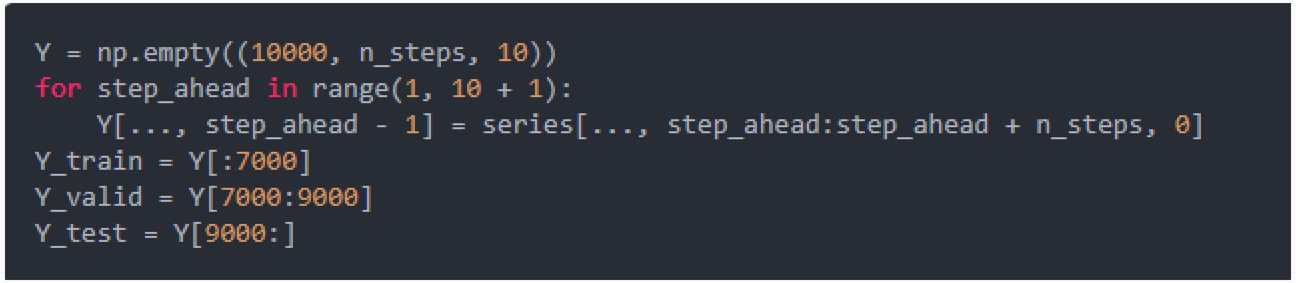


```
series = generate_time_series(10000, n_steps + 10)
X_train, Y_train = series[:7000, :n_steps], series[:7000, -10:, 0]
X_valid, Y_valid = series[7000:9000, :n_steps], series[7000:9000, -10:, 0]
X_test, Y_test = series[9000:, :n_steps], series[9000:, -10:, 0]
```



```
model = keras.models.Sequential([
    keras.layers.SimpleRNN(20, return_sequences=True, input_shape=[None, 1]),
    keras.layers.SimpleRNN(20),
    keras.layers.Dense(10)
])
```





```
Y = np.empty((10000, n_steps, 10))
for step_ahead in range(1, 10 + 1):
    Y[..., step_ahead - 1] = series[..., step_ahead:step_ahead + n_steps, 0]
Y_train = Y[:7000]
Y_valid = Y[7000:9000]
Y_test = Y[9000:]
```

Q & A

참고자료
-https://hive.blog/deeplearning/@yskoh/mit-6-s191-2
-MIT 6.S191 (2020) : RNN