

삼성전기 AI전문가 양성과정 - 프로젝트 실습 (비영상)

# 자연어처리를 위한 Padding

현청천

2022.02.28

# What is Padding

나는	학생	입니다
----	----	-----

당신은	수학	선생님	입니다
-----	----	-----	-----

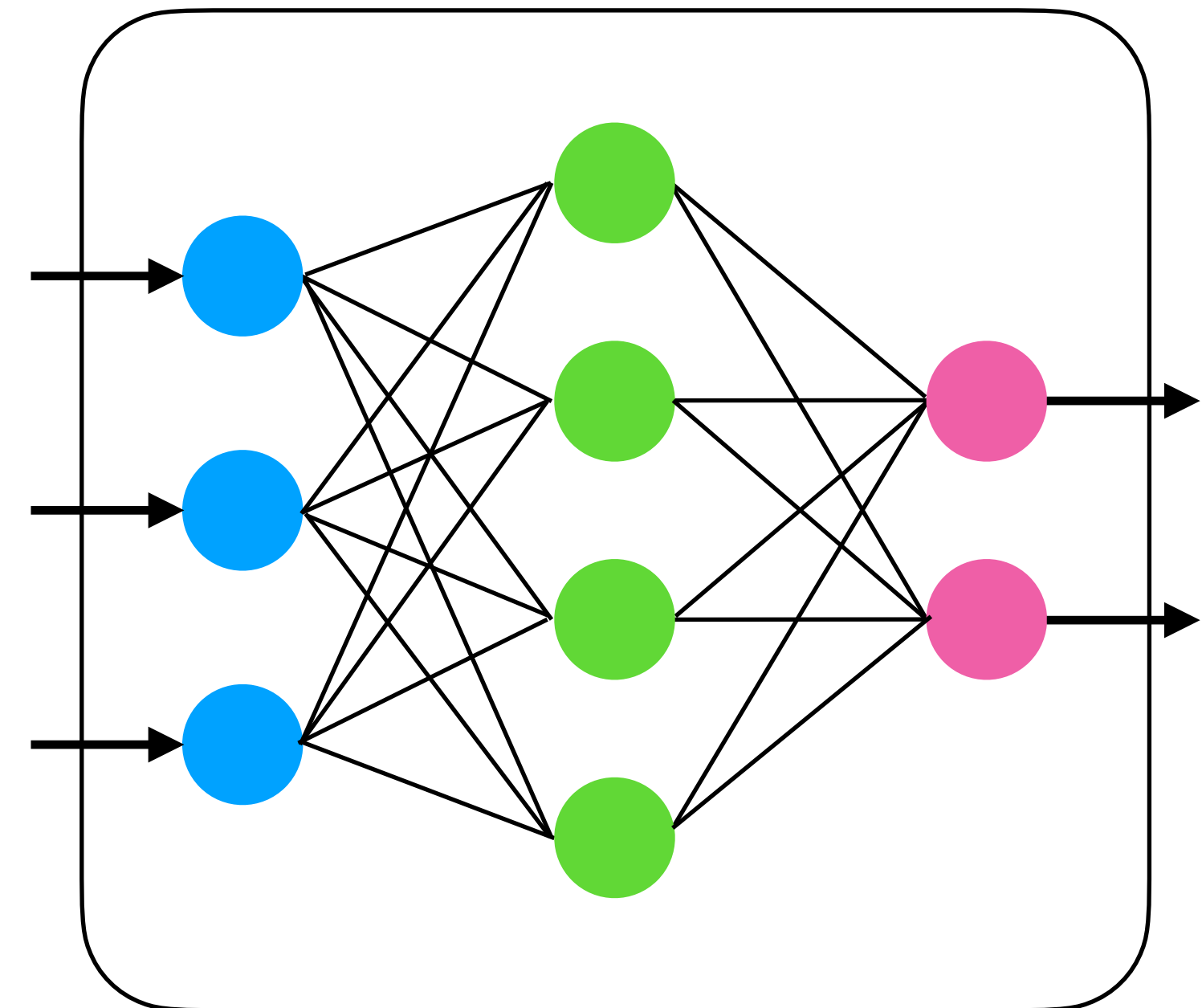
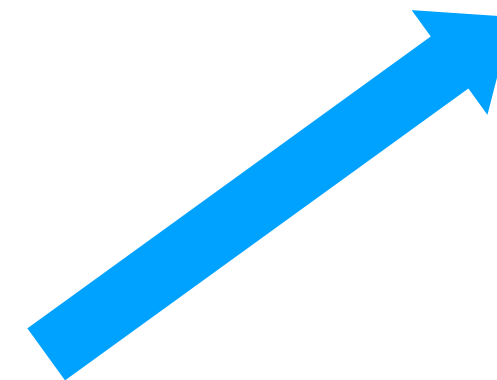


나는	1
학생	2
입니다	3
당신은	4
수학	5
선생님	6



1	2	3
---	---	---

4	5	6	3
---	---	---	---



문장의 특성상 길이가 다양함

# What is Padding

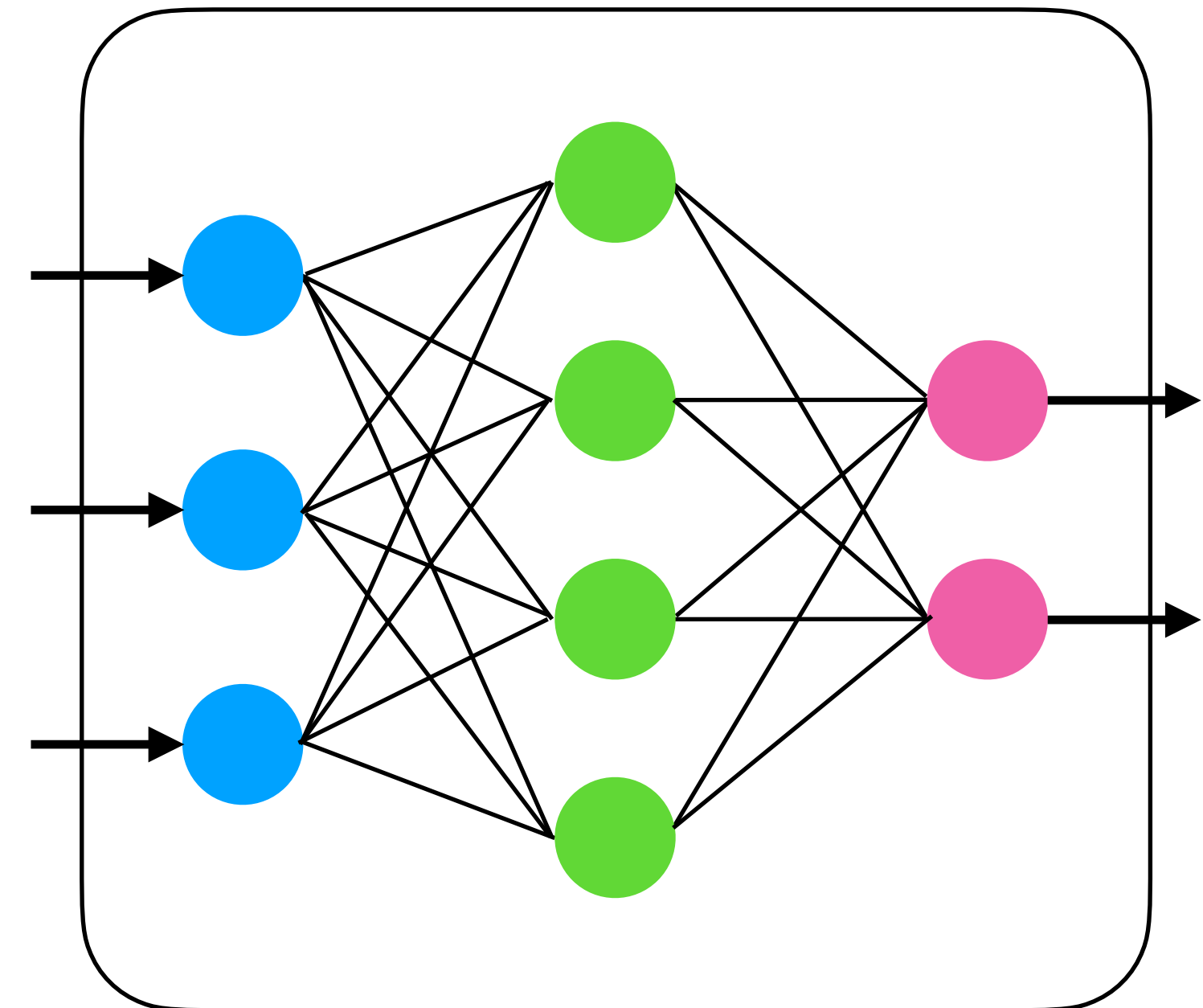
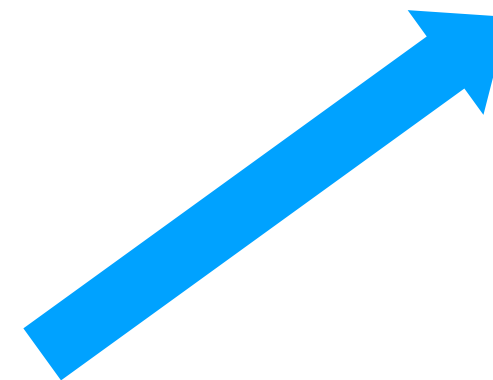
나는	학생	입니다	
당신은	수학	선생님	입니다



나는	1
학생	2
입니다	3
당신은	4
수학	5
선생님	6



1	2	3	
4	5	6	3



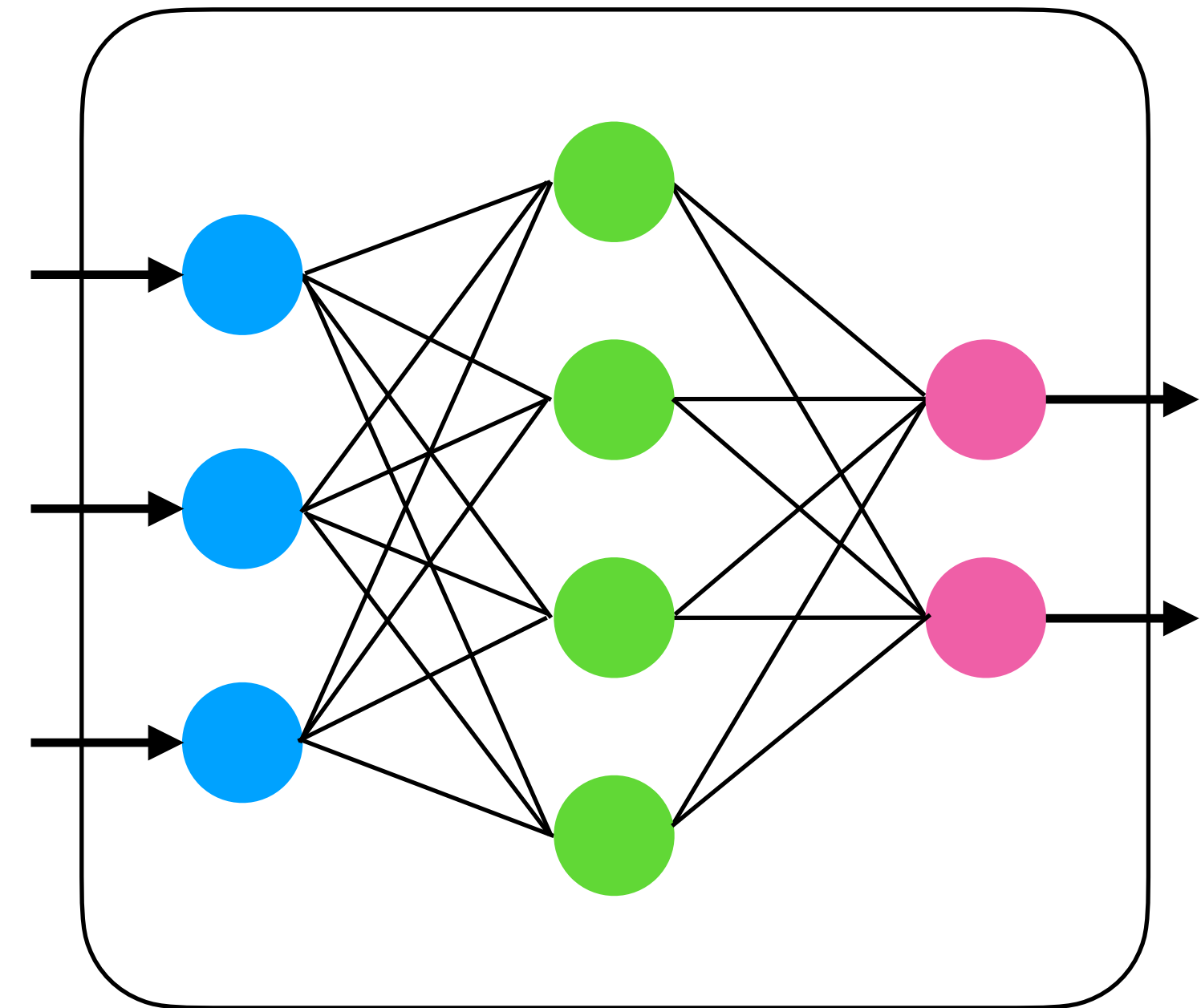
다양한 길이의 문장을 하나의 행렬로 만들어서 입력 해야 함

# What is Padding

나는	학생	입니다	[PAD]
당신은	수학	선생님	입니다

[PAD]	0
나는	1
학생	2
입니다	3
당신은	4
수학	5
선생님	6

1	2	3	0
4	5	6	3



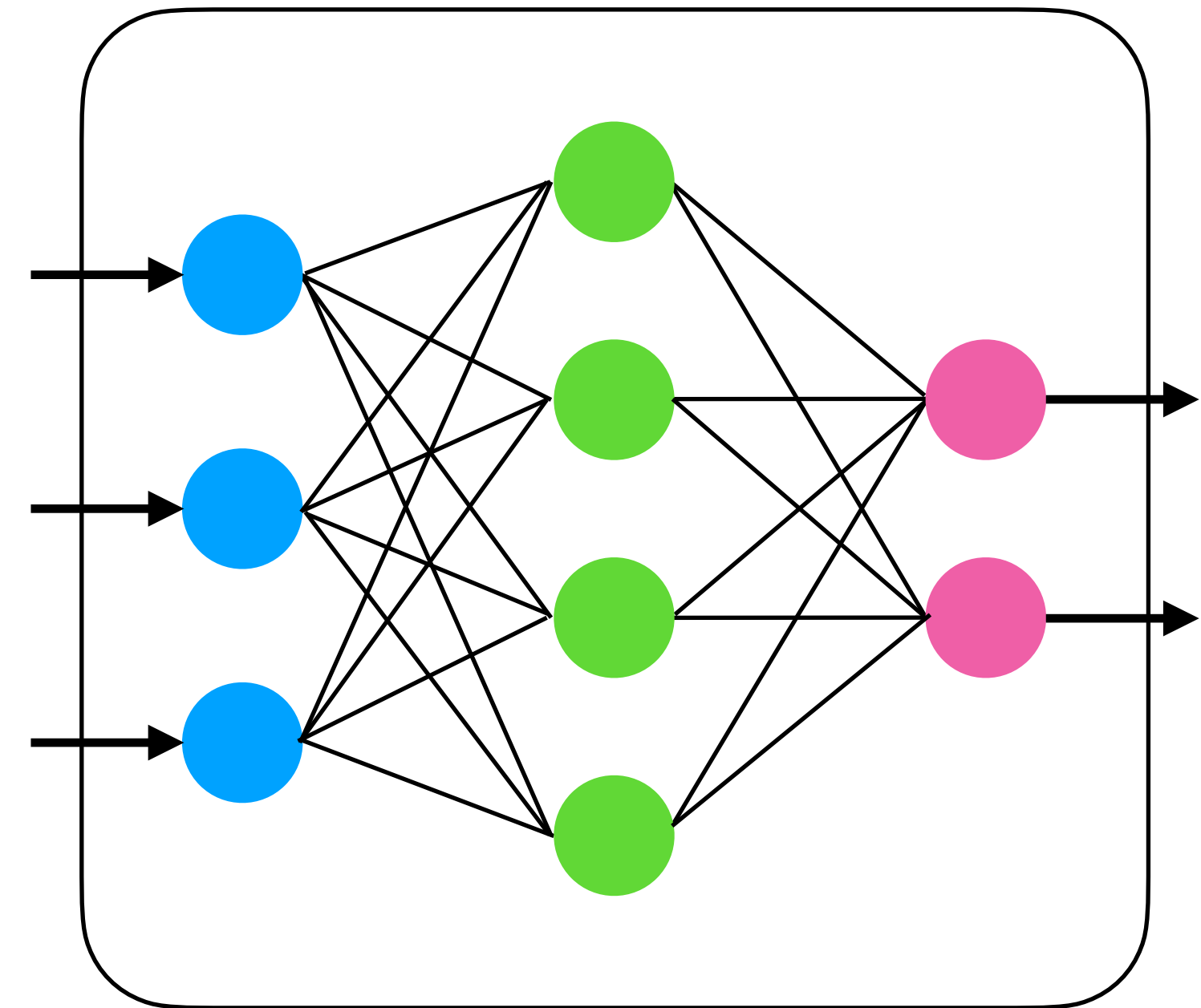
**[PAD] 라는 특수토큰을 짧은 문장의 뒤에 추가해서 길이를 맞춤**

# What is Padding

[PAD]	나는	학생	입니다
당신은	수학	선생님	입니다

[PAD]	0
나는	1
학생	2
입니다	3
당신은	4
수학	5
선생님	6

0	1	2	3
4	5	6	3



필요한 경우는 [PAD] 앞에 추가해 주기도 함

# Padding Strategy

수학은	수식이	복잡해서	어렵다										
수학은	공식이	어렵다											
수학은	수식이	이해되면	쉽다										
숙학은	공식이	능통하면	쉽다	영어는	단어가	어렵다							
영어는	듣기가	복잡해서	어렵다	영어는	단어가	이해되면	쉽다						
영어는	듣기가	능통하면	쉽다	국어는	지문이	어렵다							
국어는	한문이	많아서	어렵다	국어는	한문이	능통하면	쉽다						
국어는	지문이	이해되면	쉽다	수학은	공식이	쉽다							
수학은	수식이	복잡해서	어렵다	영어는	듣기가	능통하면	쉽다	국어는	지문이	이해되면	쉽다	한문이	어렵다

매우 긴 문장으로 인해 많은 [PAD]가 필요할 수 있음

# Padding Strategy

수학은	수식이	복잡해서	어렵다				
수학은	공식이	어렵다					
수학은	수식이	이해되면	쉽다				
숙학은	공식이	능통하면	쉽다	영어는	단어가	어렵다	
영어는	듣기가	복잡해서	어렵다	영어는	단어가	이해되면	쉽다
영어는	듣기가	능통하면	쉽다	국어는	지문이	어렵다	
국어는	한문이	많아서	어렵다	국어는	한문이	능통하면	쉽다
국어는	지문이	이해되면	쉽다	수학은	공식이	쉽다	
수학은	수식이	복잡해서	어렵다	영어는	듣기가	능통하면	쉽다

국어는	지문이	이해되면	쉽다	한문이	어렵다

매우 긴 문장을 일부 버려서 [PAD] 수를 줄일 수 있음

# Padding Strategy

수학은	수식이	복잡해서	어렵다
수학은	공식이	어렵다	
수학은	수식이	이해되면	쉽다

숙학은	공식이	능통하면	쉽다	영어는	단어가	어렵다	
영어는	듣기가	복잡해서	어렵다	영어는	단어가	이해되면	쉽다
영어는	듣기가	능통하면	쉽다	국어는	지문이	어렵다	

국어는	한문이	많아서	어렵다	국어는	한문이	능통하면	쉽다						
국어는	지문이	이해되면	쉽다	수학은	공식이	쉽다							
수학은	수식이	복잡해서	어렵다	영어는	듣기가	능통하면	쉽다	국어는	지문이	이해되면	쉽다	한문이	어렵다

배치 단위로 [PAD]를 추가해서 [PAD] 수를 줄일 수 있음



**감사합니다.**