

**서울대학교 4 차 산업혁명 아카데미**  
**빅데이터 플랫폼**  
**딥러닝 (강유 교수님) 숙제 2**

출제: 2017 년 11 월 9 일 목요일

제출: 2017 년 11 월 16 일 목요일

본 숙제의 목표는 실제 데이터에서 동작하는 Convolutional Neural Networks (CNNs) 모델을 만들어 보는 것입니다. 기본적으로 제공되는 스켈레톤 코드는 ipynb 파일로 작성이 되었고 사용 언어는 Python 3 이며 tensorflow framework 에서 제공하는 CNNs 함수들을 자유롭게 사용할 수 있습니다. 숙제에 사용할 데이터와 뼈대 코드는 제공됩니다.

## 1. 데이터 정보

숙제에 사용되는 데이터는 총 60,000 개의 이미지 데이터입니다.

- **이미지 크기:** 가로 세로의 길이가 32 이며 R, G, B 색상으로 표현하여 각 이미지는 32×32×3 3 차원 텐서 형태입니다.
- **데이터:** 총 60,000 개 중 50,000 개는 학습 데이터용이며, 나머지 10,000 개는 테스트 데이터용입니다.
- **클래스:** 각 이미지는 10 개의 클래스 중 하나에 속합니다. 클래스는 다음과 같습니다. {airplane, automobile, bird, cat, deer, dog, frog, horse, ship and truck}.

## 2. 문제 정의 및 평가 방법

본 숙제의 목표는 주어진 학습 데이터를 통해 CNN 모델을 학습하고 테스트 데이터의 클래스를 예측하는 것입니다. 예측의 정확도로 평가하며 정확도는 다음과 같이 계산합니다.

$$\text{정확도}(\%) = \frac{\text{클래스를 정확히 예측한 데이터 수}}{\text{테스트 데이터의 총 수}} \times 100 (\%)$$

예를 들어, 총 테스트 데이터의 수가 10,000 개라고 할 때, 그 중 6,000 개 데이터의 클래스를 모델이 올바르게 예측했을 때의 정확도는 60 %이다.

### 3. 제공 데이터

본 숙제에는 훈련, 테스트에 사용할 데이터 파일이 주어집니다.

- **학습 데이터(data\_batch):** 이미지 학습 데이터입니다.  $32 \times 32 \times 3$  이미지 데이터가 50,000 개 존재하며 각 이미지에 대한 클래스 정보가 담긴 데이터가 존재합니다.
- **테스트 데이터(test\_batch):** 이미지 테스트 데이터입니다. 학습 데이터는 10,000 개 존재하며 이미지 데이터의 크기는 학습 데이터와 같으며 클래스 정보가 담긴 데이터가 존재합니다.

### 4. 코드 구현

빠대 코드(hw2\_cnn.ipynb)가 제공되며 여러분은 CNN layer 를 구성하고 그에 대해 학습, 예측하는 내용을 채워 제출해야 합니다. 라이브러리를 사용하는 것은 자유입니다.

### 5. 보고서 작성

여러분은 1-2 페이지의 간단한 보고서를 작성하여 제출해야 합니다. 본 숙제는 모델의 객관적인 성능 뿐 아니라 여러분이 좋은 모델을 찾기 위해 노력한 과정을 기반으로 채점됩니다. 따라서 여러분의 보고서는 다음과 같은 내용을 필수적으로 포함해야 합니다.

- **사용한 모델의 구조:** CNN 모델의 Layer 를 어떻게 구성하신 것에 대해 서술을 해야 합니다.
- **인자 탐색 과정 및 결과:** 여러분이 사용한 모델의 최적 인자를 탐색하는 과정, 방법, 그로 인해 찾은 최적 인자를 서술해야 합니다.
- **모델 성능 향상 과정 서술:** 위의 두 가지 내용을 포함하며, 처음 모델의 예측 정확도에서 어떤 요인들 (최적 인자, Layer 의 구성요소 변화)에 의해 성능이 향상되었는지 서술합니다. 성능 향상에 대한 서술과 함께 정확도가 얼마나 나왔는지 명확하게 서술합니다. (예) 처음 모델을 완성하고 테스트했을 때 정확도 성능 20%. ~~~로 인해서 정확도 성능 50%로 향상, 그 이유는 ~~~~으로 추측이 된다.

## 6. 제출 방법

완성된 숙제는 압축하여 장준기 조교([elnino9158@gmail.com](mailto:elnino9158@gmail.com))에게 보내면 됩니다. 압축 파일의 이름에는 제출자의 이름이 반드시 포함되어 있어야 합니다. 압축 파일에 포함되어야 하는 파일의 목록은 다음과 같습니다. 이중 코드 파일(hw2\_cnn.ipynb)은 Python 3 환경에서 실행 가능해야 하며, 데이터 파일(test\_batch)가 주어졌을 때 CNN 모델로 클래스를 얼마나 잘 예측했는지 정확도를 출력합니다.

- **HW2\_{이름}.zip**

- **report.pdf:** 보고서 파일입니다. pdf 형식이어야 합니다.
- **hw\_cnn.ipynb:** 결과 코드입니다.
- **README.txt:** (선택사항) 추가적으로 언급할 내용을 적으면 됩니다.