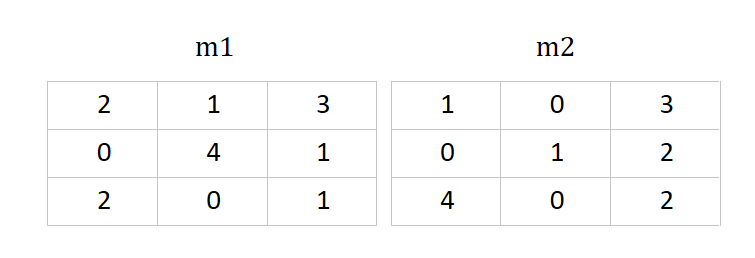
**Machine Learning & Data Mining, Spring 2020**

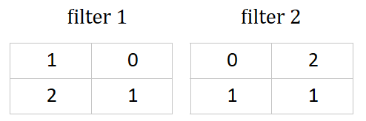
**Homework 8**

Due May 28

1. In Convolutional Neural Network (CNN), the -th output feature map at layer is given as follows.

Where is the input feature map and is the convolution filter (or kernel). Suppose a CNN layer has the following two feature maps at a certain layer:



1. Apply the convolution operation to the above feature maps using the following filters(2x2x2). Show the output feature maps (Assume stride = 1 and with zero-padding size 1) [5]

|  |  |  |  |
| --- | --- | --- | --- |
| 2 | 5 | 5 | 6 |
| 0 | 6 | 10 | 5 |
| 2 | 4 | 5 | 3 |
| 0 | 2 | 0 | 1 |

|  |  |  |  |
| --- | --- | --- | --- |
| 1 | 1 | 3 | 3 |
| 2 | 1 | 9 | 2 |
| 4 | 6 | 6 | 2 |
| 8 | 0 | 4 | 0 |

m1 \* filter1 m2\*filter2

Output feature maps (m1\*filter1 + m2\*filter2)

|  |  |  |  |
| --- | --- | --- | --- |
| 3 | 6 | 8 | 9 |
| 2 | 7 | 19 | 7 |
| 6 | 10 | 11 | 5 |
| 8 | 2 | 4 | 1 |

(b) Apply 2x2 max pooling operation to the output maps obtained in (a). Show the resulting layer. [5]

|  |  |  |  |
| --- | --- | --- | --- |
| 3 | 6 | 8 | 9 |
| 2 | 7 | 19 | 7 |
| 6 | 10 | 11 | 5 |
| 8 | 2 | 4 | 1 |

|  |  |
| --- | --- |
| 7 | 19 |
| 10 | 11 |

->

1. Unlike VGG-net or GoogleNet, ResNet can consist of more than 100 layers. Briefly explain what technology makes it possible and how. [10]

두 가지 기술이 100 layer보다 더 깊게 쌓을 수 있도록 해주었는데,

이것들은 skip connection과 residual connection입니다.

residual connection은 타겟값을 직접 학습하는 것이 아닌, 타겟이 되는 값에서 input feature값을 뺀 값을 학습시켜서 output을 표출 시, input feature값을 더해서 최종적인 결과를 내는 방법입니다. skip connection은 residual connection 시에, input을 output에 더하는 과정에서 중간 layer를 건너뛰고 input에서 output으로 바로 가기 때문에 중간과정을 skip하는 것을 일컫습니다. residual connection이 기존 방식에 비해 더 좋은 점은 기존에는 학습한 정보(hidden layer input)을 보존하지 않고, 이 정보를 토대로 새로운 정보를 만들어서 학습량이 많았습니다. 반면, residual은 기존의 정보를 보존하고, 거기에 추가적인 학습을 하기때문에 기존의 방식에 비해서 학습량이 줄어들게 됩니다. 그래서 만약 배워야할 feature가 많아지게 되면 기존의 NN은 한 번에 학습해야 할 것이 많아지게 되어 깊게 쌓지 못하고, 반면 Residual은 이전의 정보가 넘어오니 거기에 추가적으로 필요한 부분만 학습하게 되는데, 이미지 분류라는게 사실상 input과 output이 같은 쪽으로 가게끔 매핑되어야 하기때문에 결과물이 입력을 대변하는 것이라 할 수 있겠습니다. 예로 강아지 그림이 들어오면 아웃풋 이미지도 강아지 이미지와 유사해야 합니다. 때문에 입력을 유지하고, 출력과 입력의 차이를 학습시켜서 이 차이가 0으로 가게끔 만들어 주면 입력이미지와 출력이미지가 유사해지는 것을 의미합니다. 무에서 유를 창조하는 기존의 방식과 있는 것을 모방하여 학습을 덜 시키게끔 하는 방식의 차이라고 할 수 있겠습니다.

1. Among deep learning-based methods, the CNN-based model in the computer vision field and the RNN-based model in the NLP field have good performance. For each field, describe at least one reason why the model has a good performance based on the characteristics of data. [10]

CNN 모델이 컴퓨터 비젼 분야에서 좋은 성능을 보이는 이유는 CNN 모델이 기존의 NN을 이미지에 최적화 시킨 모델이기 때문인데, 기존의 NN은 단순히 input feature 벡터에 weight 벡터를 내적시키는 연산을 하였다면, CNN은 이 연산을 convolution 연산으로 대체하여, 이미지라는 특징이 있는 데이터가 잘 학습될 수 있게끔 만든 것이기 때문입니다. 필터라는 개념이 도입되어서 pixel level에서의 가로, 세로, 대각 등등 이런 특징을 잡기 때문에 이미지에서 좋은 성능을 보이는 것입니다.

RNN 모델이 NLP(자연어 처리)분야에서 좋은 성능을 보이는 이유는 RNN모델이 기존 NN에서 sequential한 데이터를 처리하게끔 변형한 모델이기 때문입니다.

RNN은 이전의 hidden layer의 정보가 다음 hidden layer에 넘겨주고 input이나 output이 가변적인 길이를 가질 수 있다는 특징이 있는데, 이는 기존의 NN에서는 없던 방식이고, 이전의 정보를 넘겨받기 때문에 syntex, sementic라는 것이 앞 뒤 단어나 문장에 종속성에 의해 좌우되기 때문에 이전의 정보를 넘겨받아서 다음 생성정보에 영향을 주어야합니다. 때문에 RNN모델이 NLP나 번역 등등 문자 처리나 서로 앞 뒤로 종속관계에 있는 Sequential한 데이터에서 좋은 성능을 보이는 것입니다.