

1. 불균형 데이터셋의 문제점과 함께 그것을 해결하는 방법에 대해 설명하시오.

- 문제점: 소수 데이터 샘플 클래스의 분포를 무시할 수 있음, 과적합 발생 가능성 있음

- 해결 방법.

: 데이터 추가 수집 - 가장 기본적인 방법으로, 데이터셋에 더 많은 샘플을 추가하여 클래스별 분포를 조정함.  
- 데이터 추가 수집이 어려울 수 있으며, 추가된 데이터가 기존 데이터와 다를 경우 overfitting 문제 발생

: 데이터 증강 - 기존 데이터를 회전, 이동, 크기 변경 등의 변형을 가해 새로운 데이터를 생성하는 기법.  
- 데이터셋의 크기와 다양성을 늘리고 모델이 클래스별 특징을 더 잘 학습하도록 함

: 손실 함수 계산시 클래스별 가중치 부여

- 일반적으로 학습시 모델은 데이터 수가 많은 클래스에 더 많은 비중을 둠

- 학습시 각 클래스의 손실(loss)을 계산할 때, 데이터 수가 적은 클래스에 더 큰 가중치를 부여

: 클래스 재샘플링

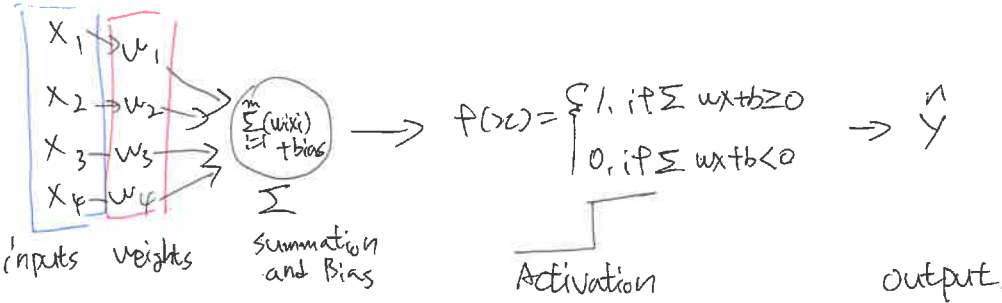
- 데이터셋에서 분포가 적은 클래스의 샘플을 복제하여 새로운 데이터셋을 생성하는 방법.

: 클래스 언더샘플링

- 분포가 많은 클래스의 샘플을 일부 제거하는 방법.

2. 퍼셉트론의 구조를 그림으로 설명하시오.

- 1959년, 로젠블랫이 제안한 학습 가능한 신경망 모델



3. 전이 학습에 대해 설명하시오.

- 사전 학습된 (pretrained) 모델을 이용하여 학습하는 것.  $\rightarrow$  data, time의 절감

- 사전 학습된 모델 중에서 어느층까지 고정하느냐에 따라 3종류로 구분.

1) pretrained network as a classifier

2) 전이 학습 : pretrained network as a feature extractor

3) 미세조정 : Fine-tuning - 학습률을 작게 설정

4. CNN 아키텍처의 특징과 주요 구성요소 및 구조를 서술하시오.

- 다층 퍼셉트론 (MLP)의 발전된 형태이며, 이미지 처리에 적합한 신경망.

: 인간의 시각 피질과 유사하고 2차원 (2D) 특징을 추출하고 학습.

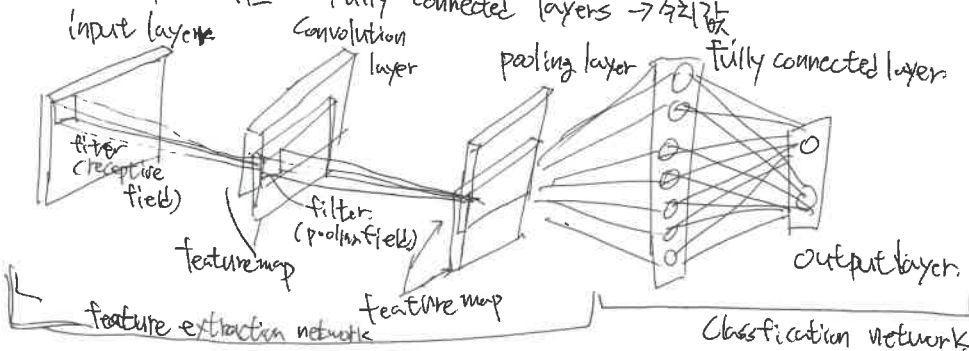
: 같은 크기의 DNN에 비해 파라미터 수가 매우 적다.

- CNN의 구조

: 특징 추출 네트워크 - Convolution layer (+ ReLU/BN) + pooling layer  $\rightarrow$  feature map 생성

: 분류 네트워크 - Fully connected layers + sigmoid/softmax  $\rightarrow$  단일/다중 class (별극형 데이터)

: 회귀 네트워크 - Fully connected layers  $\rightarrow$  수치값



5. Convolution 연산의 가중치 parameter의 총 수를 계산하는 과정을 그림으로 설명하시오.

- 필터(커널)는 입력 데이터와 동일한 채널을 가짐
- 출력 데이터의 채널은 필터의 수와 같다.
- 가중치 매개변수의 총수.

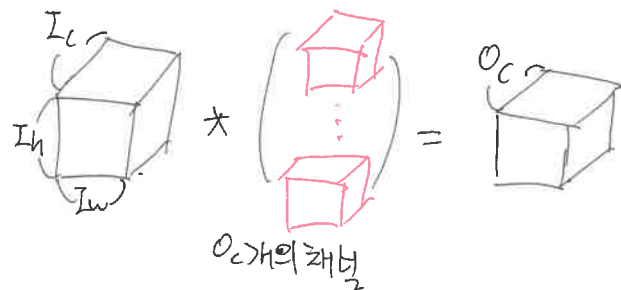
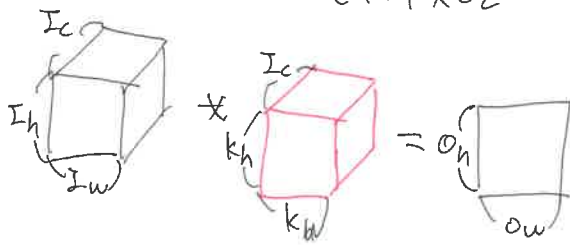
입력영상:  $I_h \times I_w \times I_c$

필터크기:  $K_h \times K_w$

출력영상:  $O_h \times O_w \times O_c$

$$(K_h \times K_w \times I_c + 1) \times O_c$$

$$= (K_h \times K_w) \times I_c + 1 \times O_c$$



6. 오차역전파의 개념과 기울기소실 문제를 설명하고, 이에 대한 해결 방법을 설명하시오  
 오차역전파 - 각 가중치 매개변수에 대한 손실함수의 기울기를 역방향으로 흐르는 chain rule 원리를 이용하여 이렇게 chain rule을 이용하여 미분을 계산하는 것을 오차역전파라고 한다.

기울기 소실 - 역전파 (back propagation) 과정에서 출력층에서 멀어질수록 Gradient 값이 매우 작아지는 현상

해결방법 - 활성화 함수를 Sigmoid에서 ReLU로 대체하여 완화.  
 - Batch Normalization: 공분산의 변동을 줄임

7. Confusion matrix를 그림으로 그리고, 정밀도, 재현율, F1-Score에 대해 식과 함께 설명하시오

	positive	Negative	
positive	True Positive (TP)	False Negative (FN)	TP
negative	False Positive (FP)	True Negative (TN)	TN
	Precision: $\frac{TP}{TP+FP}$		$\frac{TP+TN}{TP+FP+FN+TN}$
	정밀도		정확도

정밀도와 재현율의 조화평균

: F1-Score.

$$= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

정확도 (Accuracy)

$$= \frac{TP+TN}{P+TN}$$

$$\text{재현율 (Recall)} = \frac{TP}{TP+FN}$$

$$\text{정밀도 (Precision)} = \frac{TP}{TP+FP}$$

8. Gradient 기반 최적화 알고리즘의 종류와 그 특징에 대해 설명하시오

- 경사하강법: 기울어진 방향으로만 진행하기 때문에 비효율적.
- 모멘텀: 기울기 방향으로 힘을 받아 물체가 가속되는 물리법칙을 적용한 알고리즘
- AdaGrad: 각각의 매개변수에 맞춰 적응적으로 학습률을 조정하면서 학습을 진행
- RMSProp: 과거의 기울기를 계속 누적하는 것이 아니라 얼마전의 기울기는 서서히 잊고 새로운 기울기를 크게 반영.
- Adam: 모멘텀과 AdaGrad를 합친 것으로, 더 효율적으로 변이 탐색 함수 있으며, 하이퍼파라미터의 프러닝도 잘 진행된다.

9. 학습시 Overfitting을 방지하기 위한 규제 기법에 대해 설명하시오

- Dropout: 학습시 일정 확률로 뉴런과 뉴런의 연결을 무작위로 제거시키는 것.
- L1/L2 규제: cost function에 기존의 cost function에 파라미터 가중치를 L1/L2로 제곱한 형태로 만들어 값의 범위를 제한
- Batch Normalization: 그라디언트 소실 문제가 감소, 가중치 초기화 문제를 생략시킬 수 있도록, 오버피팅은 억제하지만 효과는 크지 않음