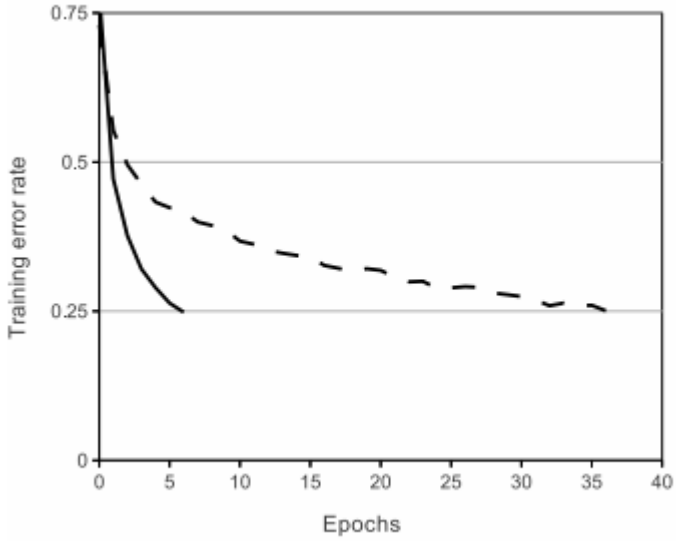
AlexNet (2012년) - ImageNet Classification with Deep Convolutional Neural Networks

CNN 모델

- ReLU 사용

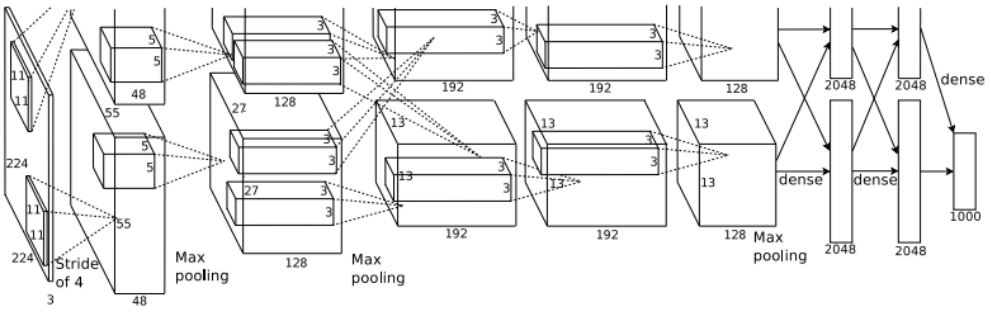


  - CNN모델의 전통방법인 f(x) = |tanh(x)| 이 아닌 ReLU를 사용
  - 훈련시간이 빨라졌을 뿐 아니라, 과대적합을 방지하는데 매우 효과적
  - ReLU는 saturating(기울기가 0에 수렴)을 막기위해 입력을 정규화할 필요가 없음

- Overlapping pooling
  - 풀링계층은 같은 커널맵내의 인접한 뉴런들을 압축하여 출력을 내보냄
  - 'Overlapping Pooling'은 z x z 크기의 풀링 윈도우를 사용하여 풀링을 수행
  - 그런데 여기서 s(풀링의 stride, 즉 풀링 윈도우가 얼마나 많이 움직이는지를 결정하는 값)가 z보다 작음

- 예를 들어, 3x3 크기의 풀링 윈도우(z=3)가 있을 때, 풀링 윈도우가 2 픽셀씩 이동(s=2)하면, 각 풀링 윈도우는 1 픽셀만큼 겹치게 됨
    - 특징 맵에서 더 많은 정보를 보존할 수 있고, 모델의 성능 향상으로 이어짐
    - 과대적합 방지 가능

- Overall Architecture
    - 신경망 구조
        - 5개의 합성곱계층과 3개의 완전연결계층을 포함해 총 8개의 계층으로 구성돼 있음
        - 마지막 softmax계층은 1000개의 클래스를 구별하기 위해 1000-way로 구성
        - 2, 4, 5번째 합성곱 계층의 커널은 오직 같은 GPU에 있는 이전 계층의 커널만 연결
        - 3번째 계층은 2번째 계층과 모두 연결
        - 완전연결게층의 뉴런들 또한 이전계층의 모든 뉴런들과 연결
        - response-normalization 계층은 1, 2번째 계층뒤에 따르고, Overlapping pooling에서 설명한 계층은 response-noramlization과 5번째 계층뒤에 따름
        - 비선형적 ReLU는 모든 합성곱, 완전연결계층의 출력에 적용



        - 첫번째 합성곱 계층은 224 X 224 X 3 의 이미지 데이터를 입력으로 받아 stride=4, 11 X 11 X 3 크기의 96개의 커널을 출력
        - 두번째 합성곱 계층은 (response-normalization과 풀링계층을 통과한) 첫번째 합성곱 계층을 입력으로 받아 5 X 5 X 48 크기의 256개의 커널을 출력
        - 3, 4, 5번째 합성곱 계층은 풀링계층이나 정규화 없이 서로 연결

        - 3번째 합성곱 계층은 (정규화, 풀링을 통과한) 2번째 계층을 입력으로 받아 3 X 3 X 256 크기의 384개의 커널을 출력
        - 4번째 계층은 3 X 3 X 192 크기의 384개 커널, 5번째 계층은 3 X 3 X 192 크기의 256개의 커널로 이루어져있음
        - 각각의 완전연결계층은 4096개의 뉴런들로 구성되어있음

    - 합성곱계층과 완전연결계층을 나눠둔 이유
        - 특징추출
            - 합성곱 계층은 이미지의 특징을 추출하는데 주로 사용
            - 해당 계층은 필터(또는 커널)를 사용하여 이미지에서 유용한 특징을 학습하고 이 특징들이 다음 계층으로 전달됨
            - 합성곱 연산은 공간적 게층 구조를 유지하면서 이미지의 로컬 패턴을 인식하는데 효과적
        - 차원축소
            - 완전 연결 계층은 합성곱 계층을 통과한 후 얻은 특징 맵을 사용하여 클래스 확률과 같은 출력을 생성
            - 해당 과정에서 차원이 축소되고 공간적 정보는 더이상 중요하지 않게 되며, 대신 각 클래스와 연관된 특징들이 중요해짐
        - 매개변수 감소
            - 합성곱 계층만 사용하면 모델의 매개변수 수가 매우 많아질 수 있음
            - 완전 연결 계층을 사용하면 학습해야할 매개변수 수를 줄이고 계산 효율성을 높일 수 있음
        - 고수준 추론
            - 합성곱 계층에서 추출한 특징들은 완전 연결 계층에서 결합되어 보다 복잡하고 추상적인 표현을 형성
            - 예를 들어, 합성곱 계층에서는 엣지나 질감과 같은 저수준 특징을 학습하지만, 완전 연결 계층에서는 이러한 특징들을 결합하여 '고양이', '자동차' 등의 고수준 개념을 인식
        - 최종 결정
            - 마지막으로 완전 연결 계층은 네트워크의 마지막에 위치하여 최종적인 분류 결정을 내리는 역할을 함

        - Softmax 함수와 같은 활성화 함수를 통해 각 클래스에 속할 확률을 출력
    - 이러한 차이 때문에, 대부분의 CNN 아키텍처는 합성곱 계층을 통해 특징을 추출하고 완전 연결 계층을 통해 최종 분류 결정을 내림
    - 최근에는 완전 연결 계층을 사용하지 않고 전역 평균 풀링(Global Average Pooling)을 사용하는 등 다양한 변형이 연구되는 중

- 과적합 예방
    - Data Augmentation
        - 데이터증강
            - 256 × 256 크기의 이미지에서 랜덤하게 224 x 224 크기의 데이터를 추출하고 학습
            - 훈련 세트가 2024배 증가함
            - 테스트 시에는 각 모서리 4곳과 가장 중앙의 1곳, horizontal-reflections까지 합한 총 10개로 증강된 이미지의 예측을 softmax계층에서 평균함으로 예측
        - RGB채널 변경
            - 각 훈련 이미지에 평균이 0이고 표준 편차가 0.1인 가우시안값에 비례하는 크기의 랜덤 변수와 이미지에서 발견된 주요구성요소의 배수를 더함
            - 각 RGB 픽셀$[I_xy^R, I_xy^G, I_xy^B]^T$에 대하여 $[P_1, P_2, P_3][\alpha_1\lambda_\iota, \alpha_2\lambda_2, \alpha_3\lambda_3]^T$를 더함
            - P_i와λ_i는 i번째 고유벡터와 RGB픽셀의 3x3 공분산 행렬의 고유값이고 각 a_i는 특정 훈련 이미지가 다시 훈련될 때 모든 픽셀에 대해 한번만 그려짐
            - 해당 방법으로 원본이미지에 대해 대략적으로 중요한 특성을 뽑아내고 조명의 세기와 색상의 변화에 대해 변하지 않음
    - DropOut
        - '드롭아웃'이라고 불리는 이 최신 기술은 각 은닉층의 뉴런들을 0.5의 확률로 0을 출력하게함
        - '드롭아웃' 된 뉴런들은 순전파와 역전파에 모두 관여하지 않음