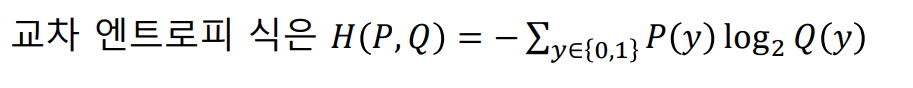
텍스트이(가) 표시된 사진

자동 생성된 설명

MSE의 단점: 오차가 큰 값이 오히려 벌점이 낮은 경우가 존재. 🡺 로지스틱 시그모이드 의 도함수는 wx+b 값이 커지면 gradient가 작아짐

Cross entropy: 레이블의 확률분포에 해당하는 P(y)와 신경망 출력의 확률분포에 해당하는 Q(y)의 확률 분포 비교. P(0) = 1-y, P(1) = y, Q(0) = 1-o, Q(1) = o



로그 우도: e = -log2(o\_y), o\_y는 샘플의 레이블에 해당하는 노드의 출력값. 로그우도는 학습 샘플이 알려주는 부류에 해당하는 노드만 본다. 따라서 softmax와 결합하여 사용.

전처리

* 규모 문제와 양수 문제를 해결해주는 정규화
* 명칭값을 원핫벡터로 변환
* 대칭적 가중치 문제를 해결하는 난수 초기화

여러가지 최적화 알고리즘

* Momentum

텍스트이(가) 표시된 사진

자동 생성된 설명

* Nesterov

텍스트이(가) 표시된 사진

자동 생성된 설명

* AdaGrad (적응적 학습률 적용)

텍스트이(가) 표시된 사진

자동 생성된 설명

* RMSProp

텍스트이(가) 표시된 사진

자동 생성된 설명

* Adam

RMSProp + Momentum

텍스트이(가) 표시된 사진

자동 생성된 설명

활성함수: 미분 가능하고, gradient vanishing을 막아야 하며, Non-linear해야 한다.

배치 정규화: 공변량 시프트 현상을 해결하기 위해 정규화를 모든 층에 적용한다. 활성함수를 적용하기 이전에 배치 정규화를 먼저 적용한다. 정규화 이후 linear transform이 적용된다.

규제의 필요성과 원리: 규제는 모델 용량에 비해 데이터가 부족한 경우의 불량 문제를 해결하는데 사용. 매끄러움 가정에 기반을 둔 식을 사용한다.

가중치 감쇠: L2 Norm. L2Norm일 때 w는 다음과 같다.

텍스트, 시계이(가) 표시된 사진

자동 생성된 설명

L1 Norm도 사용한다. L1Norm은 L2보다 더욱 sparse한 매개변수(0이 되는 매개변수가 존재). 둘 모두 최종 해를 원점 가까이 당기는 효과가 있다.

Early stopping: 기존 오류율 보다 더 낮은 오류율, 즉 새로운 최적을 발견한 상황에서 p번만큼 더 학습을 진행하여 새로운 최적이 발견되지 않으면 학습 중단.

데이터 확대: data augmentation (affine transform 등), Noise 추가

하이퍼 매개변수 탐색에는 격자 탐색보다 임의 탐색을 더 쓴다.

텍스트이(가) 표시된 사진

자동 생성된 설명

비지도 학습: 모든 데이터에 라벨이 있는 지도학습과, 일부분만 라벨이 있는 준지도 학습에 비해, 모든 데이터에 라벨이 없음.

k-means: 각 샘플별로 가까운 군집 중심으로 할당, 군집 중심의 위치는 군집의 할당된 샘플 평균으로 갱신. 위 과정을 더 이상 군집 중심이 업데이트 되지 않을 때까지 반복

EM알고리즘 기준으로, E(군집 중심 Z추정), M(군집 소속 정보A 추정) 반복한다.

k-medoids: 군집 중심의 위치는 평균이 아니라 대표값으로 정한다.

커널 밀도 추정: n차원 공간을 n’차원의 공간으로 사상하는 함수인 kernel을 갖고 밀도를 계산. 매끄러운 확률 밀도 함수를 갖는다. 커널의 대역폭 h가 작으면 뾰족한 모양이, h가 크면 뭉개진다.

커널 밀도 추정은 샘플을 모두 저장하고 있어야 하는 메모리 기반 방법이다.

가우시안 혼합: 데이터가 가우시안 분포를 따른다고 가정하고 평균과 공분산을 추정함.

텍스트이(가) 표시된 사진

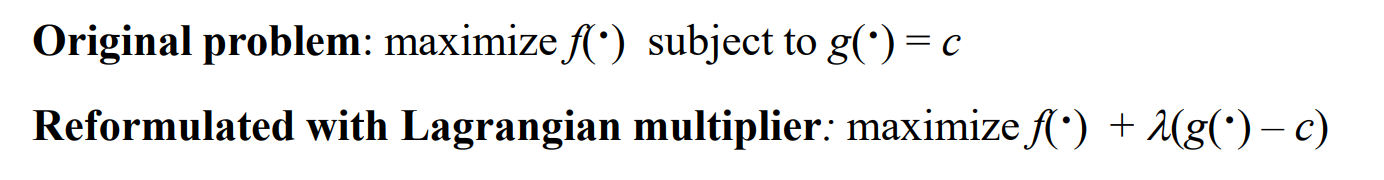
자동 생성된 설명

EM알고리즘의 관점에서: 샘플의 소속 확률 계산(E)단계, 소속 확률로부터 가우시안 개선(M) 단계.

PCA: 주성분 분석. 데이터를 원점 중심으로 옮기는 전처리(정규화) 수행 후 차원 축소.

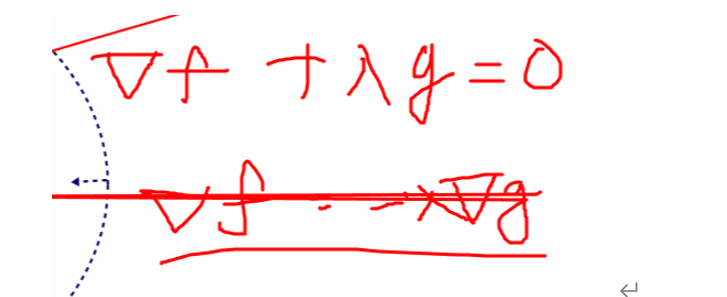
목적은 손실을 최소화하면서 저차원으로 변환하는 것. 변환된 Z의 분산이 클수록 정보손실이 적다.

라그랑주 승수: 조건이 있는 최적화 문제에 사용.



최적의 f(x, y)는 g(x, y)=c위에 존재한다.

F랑 g는 gradient방향이 같거나, 180도 반대 방향(부호가 반대)이다.



텍스트이(가) 표시된 사진

자동 생성된 설명

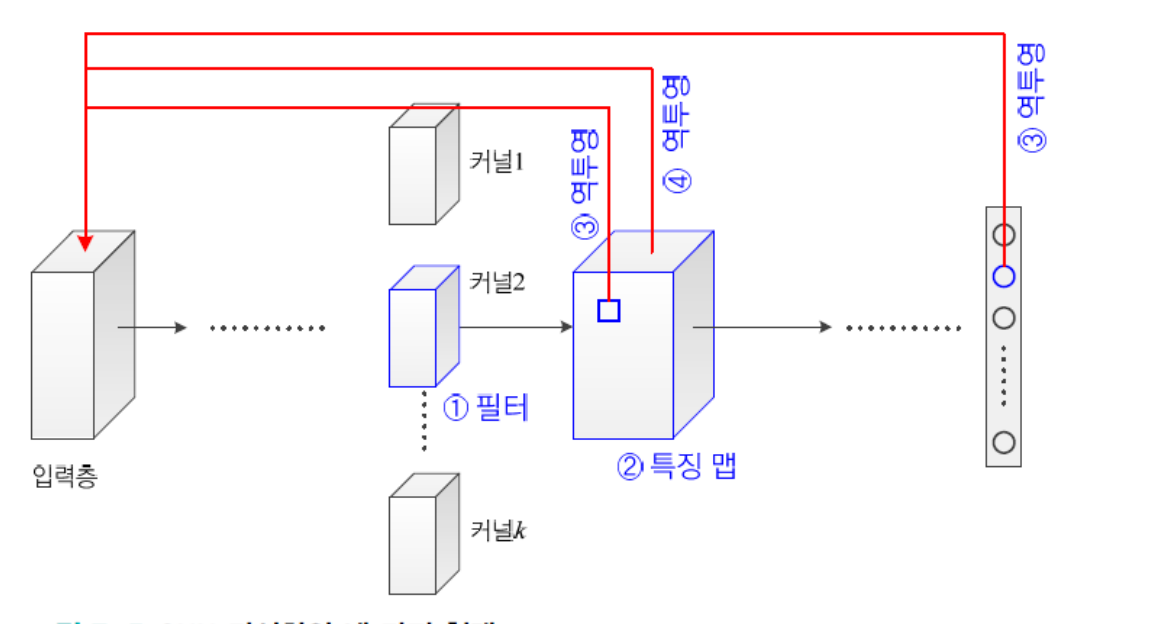
텍스트이(가) 표시된 사진

자동 생성된 설명

표현 학습의 중요성: 좋은 표현 공간을 출력하는 함수를 학습하면, 다양한 응용에 이 함수를 적용 가능.

매니폴드: 광활한 특징 공간에서 실제로 데이터가 생성되는 부분은 매우 좁음. 가상의 매니폴드에 대해, p,q축은 변화 인자에 해당 (획의 기울음, 두께, 끊김, 잡음 정도)

딥러닝 내부 가시화 기법



* 필터 가시화
* 특징 맵 가시화
* 역투영 가시화
  + 최적화를 이용한 역투영: 영상 x가 입력되었을 때, 노드 i의 활성값을 최대로 하는 x를 찾는 문제로 생각할 수 있다.
  + Deconv를 이용한 역투영

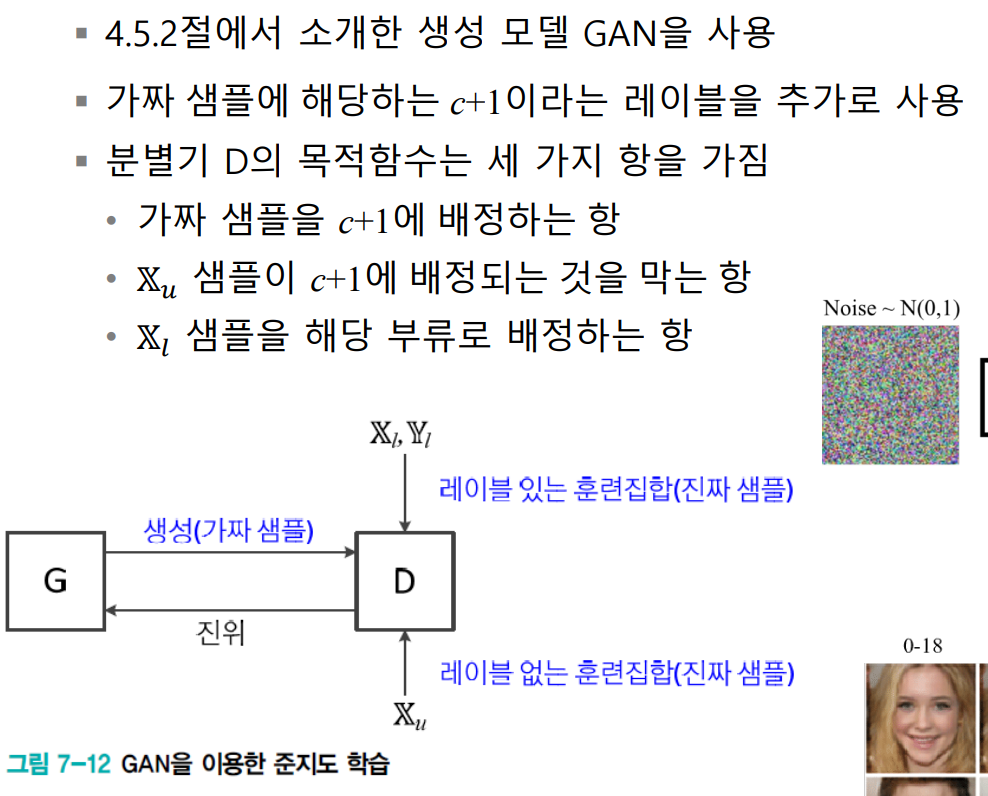
준지도 학습

* 동기와 원리: 통제된 시스템에서 연구와 개발을 하지만, 실제상황에서는 통제되지 않은 상황이 더 많다. (레이블이 없는 데이터가 존재할 확률이 매우 높다)
* 생성 모델

텍스트이(가) 표시된 사진

자동 생성된 설명

* 현대적 생성 모델



* Self-learning

텍스트이(가) 표시된 사진

자동 생성된 설명

* Co-training

텍스트이(가) 표시된 사진

자동 생성된 설명

두 분류기가 협동하는 경우와, 서로를 가르치는 경우도 있다.

* 그래프 방법, 표현 변환(PCA, Auto Encoder)
* 밀집 지역 회피: In-class variation은 줄이고, between-class distance(variation)은 늘이는 것이 좋은 특징 공간이다.

전이 학습

* 과업 전이: 이미 학습된 모델을 다른 응용분야에 적용(영상인식->음성인식)
  + Freezing: x로부터 바로 y를 학습하는 것이 아닌, x에서 x’을 추출하여 x’을 갖고 y를 학습한다.
  + Fine-tuning: FC부분을 떼어낸 후 새로운 구조를 덧붙여 다시 학습
* 도메인 전이: 특징 공간이 다른 경우(한글:불어 번역 -> 한글:영어 번역), 특징 공간은 같은데 확률 분포가 다른 상황(따낸 나뭇잎 -> 나무에 붙어있는 나뭇잎)

텍스트이(가) 표시된 사진

자동 생성된 설명