

241101 3강

pdf 보면서 다시 공부하자...

▼ Basic Gradient Descent

회귀에서의 경사하강법

- 최소값 찾는 방법 : $f'(x)$ (기울기?)가 0인 지점을 찾는 것
 - 즉, 머신러닝의 목적은 손실함수의 도함수($f'(x)$)가 0이 되는 지점을 찾는 것
 - 전체 손실함수의 형태는 알 수 없지만 기울기는 계산 가능. 경사를 타고 내려가자
- 경사하강법 수행에 필요한 하이퍼 파라미터:
 - 학습률 (Learning rate) : 가중치 업데이트에 기울기를 얼마나 반영할 것인가
 - 에포크 횟수 : 업데이트를 얼마나 반복할 것인가
 - 가중치 변화량의 threshold를 설정할 수도 있다
- update:
 - 회귀식: $y^{\wedge} = wx + b$
 - 각 가중치(w, b)에 대한 gradient 계산
 - $w = w - \text{learning rate} * \text{gradient}_w$
 - $b = b - \text{learning rate} * \text{gradient}_b$
 - update until convergence or num of epochs

```
for i in range(num_epoch):
    optimizer.zero_grad()
    output = model(x)

    loss = loss_func(output, y)
    loss.backward()
    optimizer.step()
```

- 최적점 찾기
 - 3차원에서 보면 토기 모양, 2차원에서 보면 원 모양

- 가운데, 즉 z 값이 0인 값을 찾는 것이 목적

경사하강법의 문제

- 자잘한 문제
 - 적절한 학습률 값은?
 - data scale의 문제
 - 손실함수의 gradient가 0이 되는 지점이 여러 군데
 - 학습 과정에 거치는 모든 함수가 미분 가능해야 한다
- 큰 문제
 - global minimum과 local minimum
 - 기울기가 0인 여러 좌표 중, 최소값을 가지는 global minimum이 아닌 local minimum에 가중치가 수렴될 위험
 - global minimum을 정확하게 계산할 수 있는 공식/알고리즘 없음
 - 방법론: momentum, Adagrad 등
 - 가중치 정규화, 앙상블 기법...
- 또다른 문제?
 - Feature Scaling 중요함
 - Feature들이 같은 scale을 가지면 등치선이 고르게 분포
 - → 모든 feature를 동일한 중요도로 다룬다고 표현됨
 - Unnormalized: 한쪽이 너무 크거나 작아서 등치선이 고르지 않은 상태
- 적절한 learning rate 설정이 중요한 이유
 - 너무 작다 → 수렴 속도가 느리다
 - 너무 크다 → 최적점을 지나쳐 수렴하지 못할 수도 있다
 - 학습률을 변경하며 학습을 반복하거나, learning rate scheduling을 사용한다.

적용:

- 다양한 문제에 적용 가능한 알고리즘
- Class의 확률값을 예측하는 분류 문제에도 사용 가능
 - Cross Entropy Loss는 미분이 가능하기 때문

- MLP(Multi Layer Perception)와 함께 딥러닝의 기반이 되는 알고리즘
 - 모델의 가중치를 학습하는 방법으로서 기능한다.

▼ 비지도 학습과 준지도 학습

비지도 학습의 정의와 종류

- 정의:
 - 모델이 스스로 학습하며 데이터에 숨겨진 구조를 찾으려 하는 머신러닝 기법
 - Given x_1, x_2, \dots, x_i , find a hidden structures
 - ex) Clustering, Anomaly detection, density estimation
- 군집화 정의:
 - 데이터 포인트의 특징, 분포를 파악해 n개의 군집으로 묶는다.
 - 주로 데이터 포인트 간의 거리, 유사도 등 데이터에서 뽑아낼 수 있는 특성을 사용해 수행한다.
 - 통계 분석에 주로 사용. insight를 찾을 때 좋음
 - ex) K-means, Mean-Shift, DBSCAN...
- 특이치 탐지:
 - 데이터셋에서 일반적인 데이터 포인트들과 다른 특이한 데이터 포인트를 검출하는 알고리즘
 - 사기 탐지, 제조 공정 이상 탐지, 게임 핵 감지 등에 이용
 - 다양한 특이치 탐지 알고리즘이 있음
 - 그러나 여기에선 Isolation Forest를 다룬다
- 생성형 작업:
 - 이미지나 텍스트 등을 생성하는 작업 (비지도 학습에 속함)
 - 정답 라벨이 없음. 훈련 데이터셋의 분포를 학습함
 - GAN, VAE 등의 생성형 모델이 이에 해당함
 - LLM도 비지도학습?
 - 정답이 따로 없으나, 학습 과정에서 대부분 지도학습 방법론 사용
 - → 자기 지도 학습 이라는 용어를 사용
 - 이미지 생성 모델도 마찬가지

- 차원 축소 알고리즘
 - 고차원 데이터에 데이터의 특성 반영, 저차원에 투영하는 알고리즘
 - 장점: 고차원 데이터가 가지는 '차원의 저주'를 일부 해결, 학습이 효율적
 - 단점: 데이터가 가지고 있는 특성 또는 정보 손실 위험
- DBSCAN
 - **Density-based spatial clustering of applications with noise**
 - 밀도 기반 클러스터링 방법론
 - 단순명료한 군집화 방식을 사용
 - 복잡한 형태의 군집도 구분 가능
 - 알면 좋은 용어
 - core point : 어떤 data point의 epsilon 반경 안에 minPts 이상의 point
 - epsilon : data point의 반경
 - noise : core point의 epsilon 반경 바깥 데이터

Isolation Forest

- Decision Tree 변용
- 모든 데이터 포인트를 고립시킬 때, 빨리 고립되는 포인트 → 특이치
- 모든 포인트 고립될 때까지 tree 생성, root와 가까운 leaf 데이터 포인트가 특이치
- 특징:
 - 큰 data set에 대해서 잘 작동
 - 시각적으로 확인하기 힘든 고차원 데이터의 특이치를 효과적으로 검출
 - 서로 연관없고 비선형적인 feature로 이루어진 데이터가 잘 작동, 실업에서 많이 사용됨

준지도 학습

- 지도학습에 필요한 데이터는 만들기 어려움
- 작은 숫자의 라벨링 데이터, 큰 숫자의 언라벨링 데이터 이용
 - 지도학습 task 수행
- 비슷한 데이터는 비슷한 라벨을 가짐
 - 결정경계 생성에 도움이 됨

- 회귀보다 분류에 많이 쓰임
 - 일반적으로 Classification labeling 작업이 비용, 시간 많이 듦
 - 준지도학습 방법론은 대부분 분류 작업을 전제로 함
- 준지도 학습의 가정:
 - Smoothness Assumption : 두 데이터 포인트가 가깝다면 출력도 가까울 것이다
 - Cluster Assumption : 두 데이터 포인트가 같은 군집에 속한다면 같은 class 일 것이다

준지도 학습 방법론

- Proxy Label Method
 - labeled data로 미리 학습된 모델을 이용해 unlabeled data를 분류한다.
 - labeled data + unlabeled data를 labeling한 pseudo-labeled data를 함께 사용해 모델을 다시 훈련한다
 - 문제점: unlabeled 데이터에 대한 분류가 잘못되었을 가능성을 생각하지 않음
 - → labeled data로 학습된 모델도 추정 함수이기 때문에, labeling이 정확하지 않다
 - 해결법:
 - pseudo-label의 신뢰성이 높은 데이터만 사용한다.
 - 2번 과정을 순차적으로 여러 번 반복한다.
- Consistency training
 - data에 작은 변화를 주어도 예측의 결과는 같을 것이다.
 - ex) image augmentation
 - unlabeled data를 증강(augmentation)하여 원본과 같은 label이 나오도록 추가 학습을 진행

과제:

- 오늘 배운 내용 정리
- DBSCAN 하이퍼파라미터 변경해가며 결과값 어떻게 달라지는지 실험 (주피터노트북 파일에 작성)
 - 추천) Gradient Descent 학습, 관련 내용 서치