

1. 선형 회귀의 중요성: 선형 회귀 모델의 훈련 과정이 매우 단순하여 딥러닝 심층 신경망 모델 등 대다수의 머신러닝 모델이 훈련 과정에서 선형 회귀 모델의 훈련 방식을 활용하여 보다 복잡한 문제를 해결한다.

2. 머신러닝 모델 훈련의 목표: 타겟에 최대한 가까운 예측 값 계산을 목표로 하며, 모델의 예측 성능을 최대화해야 하고, 모델 성능은 일반적으로 모델의 비용 함수를 이용하여 계산한다. 비용함수는 일반적으로 MSE를 사용하는데 모델 훈련의 최종 목표는 MSE가 최소가 되도록 하는 값을 찾는 것이다.

3. 선형 회귀 모델 훈련과 경사하강법

MSE를 비용함수로 사용하는 경우) 세타를 임의의 값으로 지정한 후 훈련을 시작하며, MSE가 허용 오차보다 적게 작아질 때까지 "배치 크기만큼의 훈련 샘플을 이용하여 예측 값 생성 후 MSE를 계산하고, 세타를 업데이트하는 과정을 반복한다.

4. 학습률의 중요성

학습률이 너무 적은 경우) 비용 함수가 전역 최솟값을 갖도록 하는 세타에 너무 느리게 수렴한다.

학습률이 너무 큰 경우) 비용 함수가 전역 최솟값을 갖도록 하는 세타에 수렴하지 않고 발산한다.

비선형 모델 훈련의 어려움) 비선형 모델은 학습률과 상관없이 파라미터를 초기화하는 방식에 따라 지역 최솟값에 수렴하거나 수렴하지 못하고 정체할 수도 있다.

특성 스케일링의 중요성) 특성들의 스케일을 통일시키면 보다 빠른 학습이 이루어진다.

5. 경사하강법 종류

배치 경사하강법: 에포크마다 한 번 그래디언트를 계산하여 파라미터 조정, 사이킷런 모델을 포함하여 일반적으로 사용되지 않는다./훈련 세트가 크면 그래디언트를 계산하는데 많은 시간이 필요하고 아주 많은 데이터를 저장해야 하는 메모리 문제도 발생할 가능성이 있다.

확률적 경사하강법: 매우 큰 훈련 세트를 다룰 수 있음. 학습 과정이 매우 빠르며 파라미터 조정이 불안정 할 수 있기 때문에 지역 최솟값에 상대적으로 덜 민감하다./학습 과정에서 파라미터의 동요가 심해서 경우에 따라 전역 최솟값에 수렴하지 못하고 계속해서 발산할 가능성도 높다.

학습 스케줄: 훈련이 지속될수록 학습률을 조금씩 줄이는 기법으로, 일반적으로 훈련 에포크가 진행될수록 학습률이 조금씩 작아지도록 설정한다. 요동치는 파라미터를 제어하기 위해 학습률을 학습 과성동안 천천히 줄어들게 만들 수 있다./학습률이 너무 빨리 줄어들면 지역 최솟값에 갇힐 수 있으며, 학습률이 너무 느리게 줄어들면 전역 최솟값에 제대로 수렴하지 못하고 맴돌 수 있는 주의사항이 있다.

미니배치 경사하강법: 배치 크기를 어느 정도 크게 하여 확률적 경사하강법보다 파라미터의 움직임이 덜 불규칙적이게 되고, 배치 경사하강법보다 빠르게 학습한다. 학습 스케줄을 잘 활용하면 최솟값에 수렴한다./SGD에 비해 지역 최솟값에 수렴할 위험도가 보다 큼.

6. 다항회귀: 비선형 데이터를 선형 회귀를 이용하여 학습하는 기법

단점: 몇 차 다항 회귀를 사용해야 할지 일반적으로 알 수 없으며, 심층 신경망처럼 비선형 데이터를 분석하는 보다 좋은 모델이 개발되어 다항 회귀를 사용할 필요가 없음.

7. 모델 성능 평가

검증 데이터셋 활용: 훈련 점수와 검증 데이터셋 점수 모두 낮은 경우에는 과소적합, 훈련 점수는 높지만 검증 데이터셋 점수가 상대적으로 많이 낮은 경우에는 과대적합으로 판단한다.

학습 곡선: 훈련셋과 검증셋에 대한 모델 성능을 비교하는 그래프로, x축은 훈련셋의 크기, y축은 훈련셋 크기에 따른 모델성능, 훈련 점수와 검증 점수를 사용하여 나타낸다. 학습 곡선의 모양에 따라 과소적합/과대적합을 판정할 수 있다.

8. 모델 규제와 조기 종료

모델 규제

-릿지 회귀: 가중치의 절댓값을 최대한 작게 유지하고, 모델의 분산을 줄이지만, 편향은 커진다.

-라쏘 회귀: 중요하지 않은 특성의 가중치를 0으로 만들고, 자유도가 줄어들어 모델의 분산이 줄어들지만, 편향은 커진다.

-엘라스틱 넷: 릿지 회귀와 라쏘 회귀를 혼합한 방법이다.

-규제 선택: 대부분의 경우 약간이라도 규제를 사용하는 것을 추천한다. 유용한 속성이 많지 않다고 판단되는 경우 불필요한 속성의 가중치를 0으로 만들기 위해 라쏘 규제나 엘라스틱넷을 활용한다. 일반적으로 엘라스틱 넷을 많이 추천한다.

조기 종료: 모델이 훈련 중에 훈련셋에 너무 과하게 적응하지 못하도록 하는 가장 일반적인 규제 기법으로, 에포크가 남아있다 하더라도 검증셋에 대한 비용함수의 값이 줄어들다가 다시 커지는 순간 훈련을 종료한다. 검증셋에 대한 비용 함수의 곡선이 진동할 수 있기에 검증 손실이 한동안 최솟값보다 높게 유지될 때 훈련을 멈추고 기억해둔 최적의 모델을 사용한다.

9. 로지스틱 회귀와 소프트맥스 회귀

회귀 모델을 분류 모델로 활용한다. 이진 분류에서는 로지스틱 회귀를 사용하며, 다중 클래스 분류에서는 소프트맥스 회귀를 사용한다.