

CUAI Hands-On ML chapter 4 문제

4.1-4.2 묶어서 두 문제

(1) Gradient Descent와 Normal Equation의 장단점을 각각 쓰세요

Normal Equation의 경우 장점으로는 learning rate를 고르지 않아도 되고 반복하지 않아도 된다. 단점으로는 공식의 내용을 적용해야 하는데 feature수가 클 경우에는 시간이 gradient descent보다 오래 걸린다.

Gradient Descent의 경우 장점으로는 feature 수가 많을 때에도 제대로 동작한다는 장점이 있으며, 단점으로는 적당한 learning rate를 골라야 하며 많은 반복을 진행해야 한다는 점이 있다.

(2) 확률적 경사하강법의 장단점을 배치 경사 하강법과 비교하여 쓰세요.

장점: 배치 하강법의 경우 매 스텝에서 전체 훈련 세트를 사용해 그레디언트를 계산한다. 그래서 훈련 세트가 커지면 매우 느려진다. 확률적 경사하강법은 매 스텝에서 한 개의 샘플을 무작위로 선택하고 그 하나의 샘플에 대한 그레디언트를 계산하므로 매 반복에서 다뤄야 할 데이터가 매우 적어 한 번에 하나의 샘플을 처리하면 알고리즘이 훨씬 빠르다.

단점: 확률적이므로 이 알고리즘은 배치 경사 하강법보다 훨씬 불안정하다. 시간이 지나면 최소값에 매우 근접하겠지만 최소값에 안착하지는 못할 것이고 최적치로 구해지기 어렵다. 또한, 배치 경사 하강법보다 전역 최솟값을 찾을 가능성이 높다.

4.3 PolynomialFeatures를 사용하면 특성이 여러 개일 때 다항 회귀에서 특성 간의 관계를 찾을 수 있다. 그 이유는?

PolynomialFeatures가 주어진 차수까지 특성 간의 모든 교차항을 추가하기 때문이다. A, B 두 특성이 있을 때 degree=3으로 하면 a^2 , a^3 , b^2 , b^3 뿐만 아니라 ab , a^2b 등도 특성으로 추가한다.

4.4 o,x 문제, 편향/분산 트레이드오프

(1) 모델의 복잡도가 커지면 통상적으로 분산이 늘어나고 편향이 줄어든다. ()

(2) 분산은 훈련 데이터에 있는 큰 변동에 의해 나타난다. ()

(3) 편향은 잘못된 가정으로 인한 것이다. 편향이 큰 모델은 훈련 데이터에 과대적합되기 쉽다. ()

풀이: $o, x, x //$ 분산은 작은 변동에 모델이 과도하게 민감해서 나타남. 편향이 큰 모델은 훈련 데이터에 과소적합되기 쉬움

4.5 특성이 몇 개 뿐일 때, 라쏘나 엘라스틱넷을 선호하는데 이 이유는? 그리고 라쏘보다 엘라스틱넷을 선호하는 경우는 언제인지 쓰시오.

답: 라쏘나 엘라스틱넷의 경우 불필요한 특성의 가중치를 0으로 만들어준다. 또한 특성 수가 훈련 샘플 수보다 많거나 특성 몇 개가 강하게 연관된 경우 라쏘가 문제를 일으킬 수 있어 엘라스틱넷을 선호한다.

4.6 로지스틱 회귀가 어떻게 작동하는지 간단히 쓰시오.(선형회귀와 비교하여 공통점 차이점 중심으로)

답: 선형 회귀 모델과 같이 로지스틱 회귀 모델은 입력 특성의 가중치 합을 계산한다. 그리고 편향을 더한다. 대신 선형회귀처럼 바로 결과를 출력하지 않고 결과값의 로지스틱을 출력한다. 로지스틱 회귀 모델의 경우 분류에서도 사용할 수 있으며 0과 1 사이의 값을 출력하는 시그모이드 함수를 통해 이 값(샘플이 양성 클래스에 속할 확률)이 0.5보다 크거나 같으면 1로 예측, 그게 아니면 0으로 예측한다.