**모델훈련**

모델을 훈련 시킨다는 것은 모델이 훈련 세트에 가장 잘 맞도록 모델 파라미터를 설정하는 것

이를 위해 먼저 모델이 훈련데이터에 얼마나 잘 들어맞는지 측정해야 한다.

회귀에서 가장 널리 사용되는 성능측정 지표는 평균 제곱근 오차(RMSE)이다. -> RMSE를 최소화하는 θ를 찾아야한다.

<선형회귀>: 일차방정식

1. 정규방정식 : 비용함수를 최소화하는 θ값을 찾기위한 해석적(수학공식) 방법

단점: 훈련할때 샘플 또는 특성이 두배 들어나면 걸리는 시간도 8배 증가한다.

예측하려는 샘플 또는 특성이 두배 들어나면 걸리는 시간도 두배 증가한다.(이미 만들어진 상태)

* θ: 기울기!

1. 경사하강법: 여러 종류의 문제에서 최적의 해법을 찾을 수 있는 매우 일반적인 최적화 알고리즘이다. (수치적)

기본 아이디어: 비용함수를 최소화하기위해 반복해서 파라미터(gradient)를 조정해가는 것

단점: 모든 비용함수가 매끈한 그릇과 같지 않아 지역 최솟값(local minimum)에 수렴할 수도 있다. ( 선형회귀의 MSE 비용함수는 매끈한 볼록함수 )

* 배치 경사하강법: θ가 조금 변경될 때 비용함수가 얼마나 바뀌는지 계산해야 합니다. (편도함수) -> 매 경사 하강법 스텝에서 전체 훈련 세트 X에 대해 계산

절절한 학습률을 찾으려면? 그리드 탐색을 사용(but, 반복횟수제한해야함)

반복횟수 어떻게 지정? 너무 작으면 최적점에 도착하기전에 알고리즘이 멈춤, 너무 크면 모델 파라미터가 변하지 않는 동안 시간을 낭비하게 됩니다. -> 그래디언트 벡터의 노름(예측값의 벡터와 타깃값의 벡터 사이의 거리를 재는 방법)이 어떤 값(허용오차)보다 작아지면 경사하강법이 거의 최솟값에 도달한 것이므로 알고리즘을 중지

* 확률적 경사하강법: 매 스템에서 딱 한 개의 샘플을 무작위로 선택하고 그 하나의 샘플에 대한 그래디언트를 계산한다.

단점: 불안정 -> 최적의 최솟값에 안착하지 못한다. but 배치보다 전역 최솟값을 찾을 확률이 높다.

해결방법: 학습률을 점진적으로 감소시키는 것. (학습 스케쥴링)

한 반복에서 샘플 수만큼 되풀이되고 각 반복을 epoch(에폭)이라고 한다.

* 미니 배치 경사하강법: 각 스텝에서 전체 훈력 세트나 하나의 샘플을 기반으로 그래디언트를 계산하는 것이 아니라 임의의 작은 샘플 세트(미니배치)에 대해 그래디언트를 계산한다. -> GPU를 사용해 얻는 성능 향상
* SGD & 미니배치 경사하강법을 외부 메모리 학습을 지원하는 알고리즘이라고 하는데?

: 학습 알고리즘을 사용해 컴퓨터의 메인 메모리에 맞지 않는 거대한 데이터 집합의 시스템을 학습할 수도 있다. 데이터의 일부를 불러오고 해당 데이터에 대해 학습 단계를 실행한다. 나머지 데이터에 대해 모두 수행될 때까지 과정을 반복한다.

\* SGD와 달리 배치방법이 직선으로 가는 이유? -> 배치는 평균으로 이동!

\* 정규화하는 이유?

1. GD 빠르게 훈련가능 2. 정규화를 꼭 써야하는 모델 K-means, K-nn: 한쪽에 민감하지 않도록 scale을 맞춘다.(유클리드 거리를 사용하는 알고리즘의 경우)

\* regression classification -> 둘다 GD로 사용 but 비용함수가 다름

<다중회귀>: 비선형 데이터를 학습하는데 선형 모델을 사용할 수 있다.

어떻게 모델이 데이터에 과대적합 또는 과소적합되었는지 알 수 있을까?

1. 교차검증

* 훈련데이터에서 성능이 좋지만 교차검증 점수가 나쁘면 모델이 과대적합된 것이다.
* 양쪽모두 좋지 않다면 과소적함

1. 학습곡선 살펴보기

* 훈련세트와 검증 세트의 모델 성능을 훈련 세트 크기의 함수로 나타낸다.

\* 과대적합 모델을 개선하는 방법은 검증 오차에 근접할 때까지 더 많은 훈련 데이터를 추가하는 것, 모델을 제한하는 것 -> 다중회귀: 차수 감소, 선형회귀: 가중치 제한 -> reg parameter 조절!

과소적합 모델을 개선하는 방법: 다중회귀: 차수 증가, 선형회귀: 가중치 제한

<릿지회귀>: 실제로 많이 쓰임

<로지스틱 회귀>: 샘플이 특정 클래스에 속할 확률을 추정하는데 사용한다. (이진 분류기)

로지스틱은 0과 1사이의 값을 출력하는 시그모이드 함수이다.

훈련목적: 양성 샘플(y=1)에 대해 높은 확률을 추정하고 음성샘플(y=0)에 대해서는 낮은 확률을 추정하는 모델의 파라미터 벡터를 찾는 것이다.

<오차의 종류>

1. 편향(잘못된 가정으로 인해 발생)
2. 분산(모델이 과도하게 민감하기 때문에 발생)
3. 줄일 수 없는 오차(노이즈 때문에 발생)

* 편향/분산 트레이드 오프: 모델의 복잡도가 커지면 통상적으로 분산이 늘어나고 편향은 준다. 반대로 모델의 복잡도가 줄어들면 편향이 커지고 분산이 작아진다.