확률적 경사 하강법

경제학과 2015231035 하지민

점진적인 학습

경사 하강법 (Gradient Descent)

함수의 기울기(경사)를 구하여 기울기의 절댓값이 낮은 쪽으로 이동시켜 극값(최적값)을 찾는 방법
즉, 비용함수를 최소화 하기 위해 경사를 반복적으로 하강해가면서 파라미터를 조정해 나가는 것

확률적 경사 하강법 (Stochastic Gradient Descent)

랜덤으로 단 한 개의 데이터를 추출하여(배치 크기 1) 기울기를 얻어냄. 이러한 과정을 반복해서 학습하여 최적점을 찾아내는 것이 확률적 경사 하강법

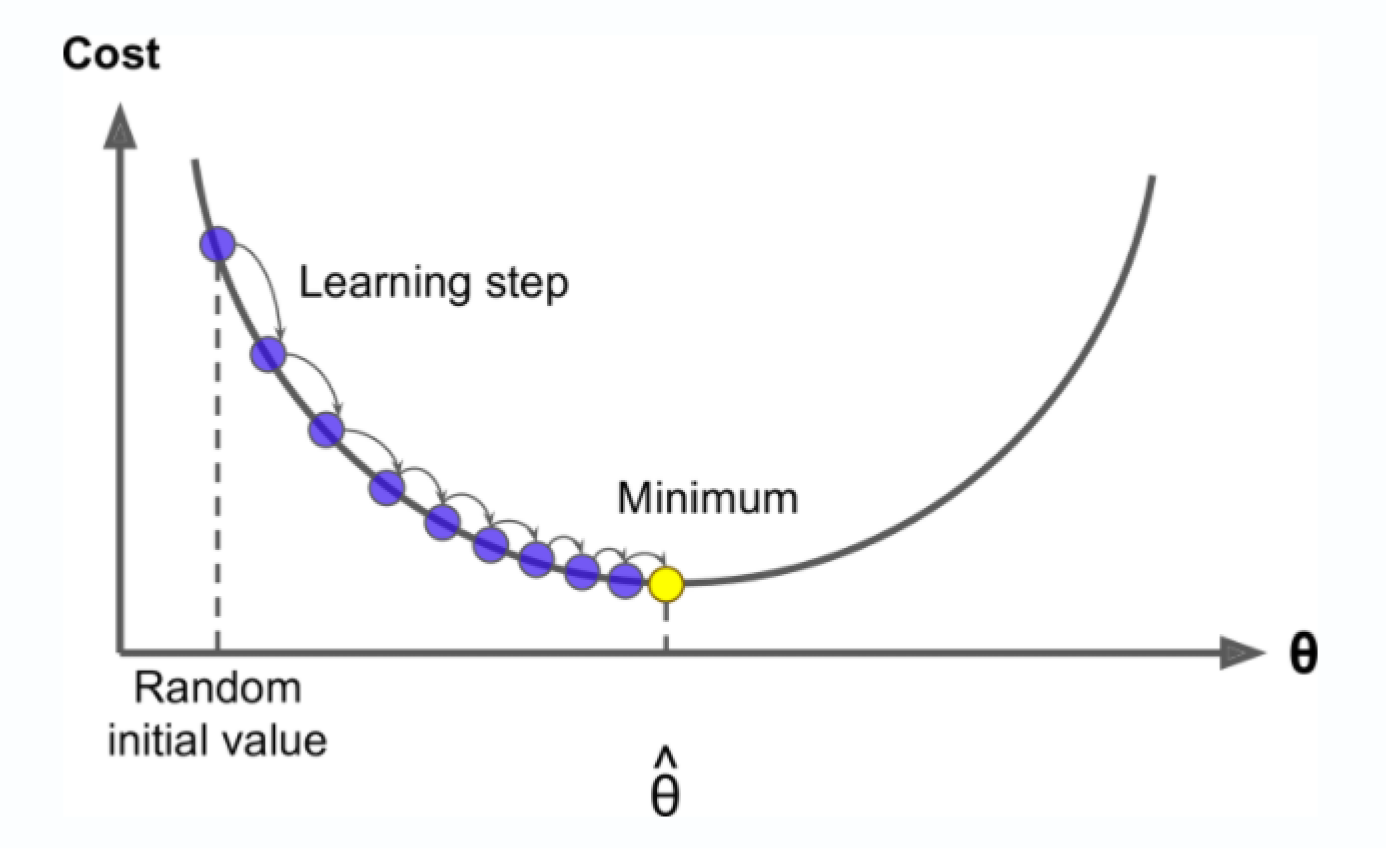
****배치(batch)****

모델 학습의 반복 1회, 즉 경사 업데이트 1회에 사용되는 트레이닝 데이터의 집합

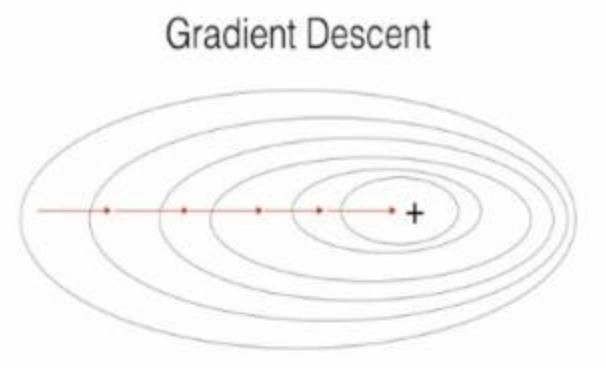
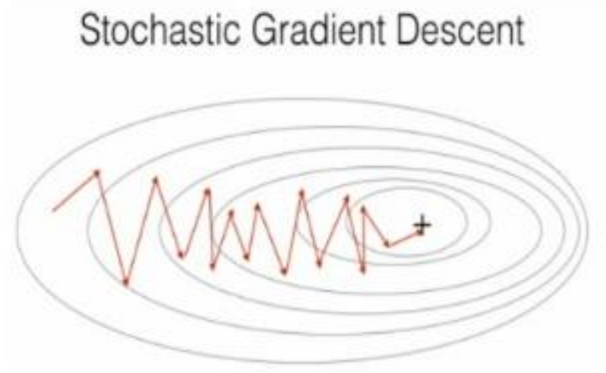
****에포크(epoch)****

확률적 경사 하강법에서 훈련 세트를 한 번 모두 사용하는 과정

경사 하강법



경사 하강법 비교

	경사하강법	확률적 경사하강법
1회의 학습에 사용되는 데이터	모든 데이터 사용	랜덤으로 추출된 1개의 데이터 사용(중복 선택 가능)
반복에 따른 정확도	학습이 반복 될 수록 최적해에 근접	학습이 반복 될 수록 최적해에 근접
노이즈	거의 없음	비교적 노이즈가 심함
해를 찾는 과정의 이미지 비교		

한 번 학습할 때 마다 모든 데이터를 계산하여 최적의 한 스텝을 나아가는 경사 하강법과 달리, 확률적 경사 하강법은 랜덤하게 추출한 하나의 데이터만 계산하여 빠르게 다음 스텝으로 넘어 감.

그 결과 더 빠르게 최적점을 찾을 수 있게 되었지만 그 정확도는 낮아짐

손실 함수

손실 함수

예상한 값과 실제 타깃값의 차이를 함수로 정의한 것(비용함수, 목적함수로도 불림)

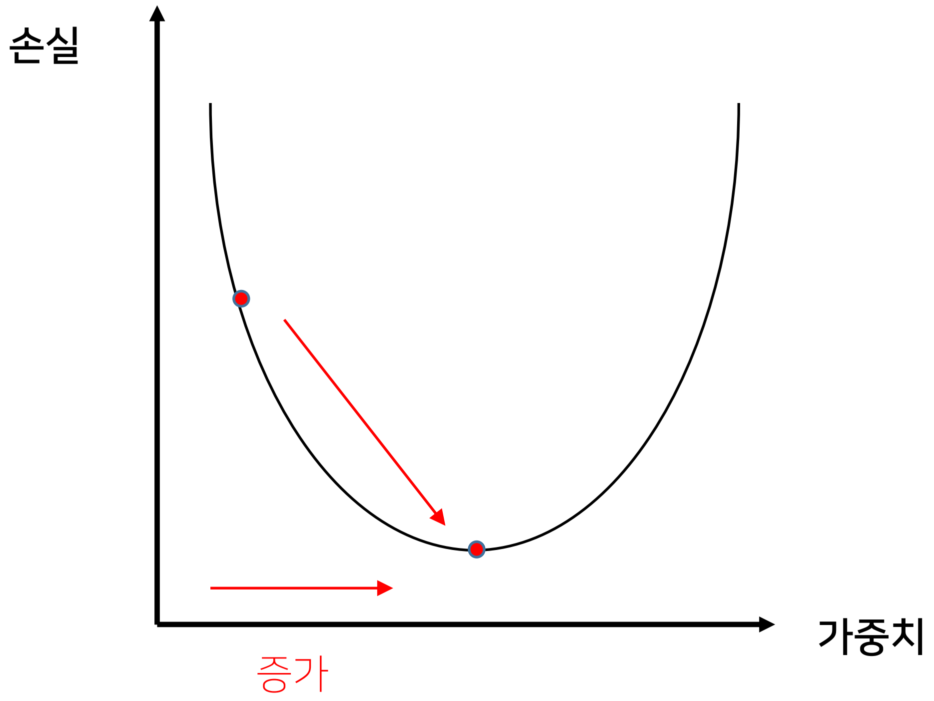
평균 제곱 오차를 많이 사용.(선형회귀 모델에 적합)

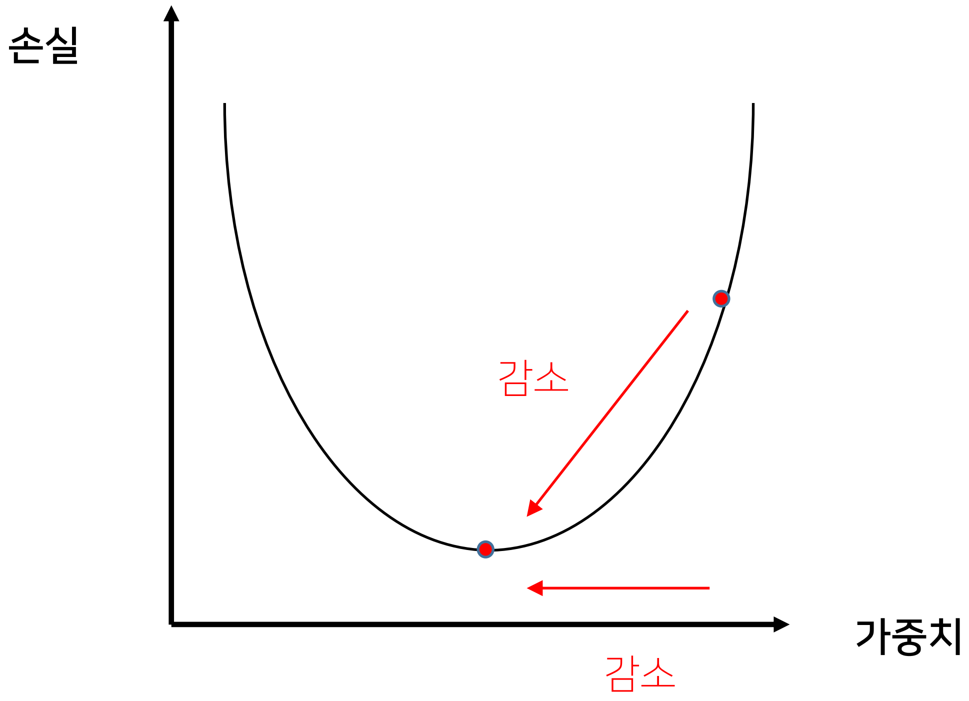
$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

차이가 음수이든 양수이든 상관없이 차이의 크기만 고려하고자 하기 때문

손실 함수

최적의 가중치 = 예측값과 실제 값의 차이를 최소화하는 가중치 = 손실 함수 값을 최솟값으로 만드는 가중치





접선의 기울기가 음수이면 w를 증가시키고, 접선의 기울기가 양수이면 w를 감소시킴
W_new = w – 접선 기울기(손실함수 미분값)

손실 함수에서 가중치와 절편의 업데이트

가중치에 대해 제곱오차 미분

오차역전파

절편에 대해 제곱오차 미분

$$SE = (y-\hat{y})^2$$
$$= (y-(wx+b))^2$$

① $\hat{y} = wx+b$

② $\frac{dSE}{dw} = \frac{d}{dw}(y-\hat{y})^2 = 2(y-\hat{y})(-\frac{d}{dw}\hat{y})$

$$= 2(y-\hat{y})(-x) = -2(y-\hat{y})x$$

$$W_{new} = w-(-2(y-\hat{y})x)$$

* 여기서 2는 이동하는 정도에만 영향
방향성에는 영향X
2를 없애줘도 가중치 찾는데 무방

$$W_{new} = W - \frac{dSE}{dw} = w+(y-\hat{y})x$$

① $\hat{y} = wx+b$

② $\frac{dSE}{db} = \frac{d}{db}(y-\hat{y})^2 = 2(y-\hat{y})(-\frac{d}{db}\hat{y})$

$$= 2(y-\hat{y})(-1) = -2(y-\hat{y})$$

$$b_{new} = b - \frac{dSE}{db} = b+(y-\hat{y})$$

로지스틱 손실 함수

로지스틱 손실 함수

이진분류를 위한 모델

다중 분류일 경우 크로스엔트로피 손실 함수 사용

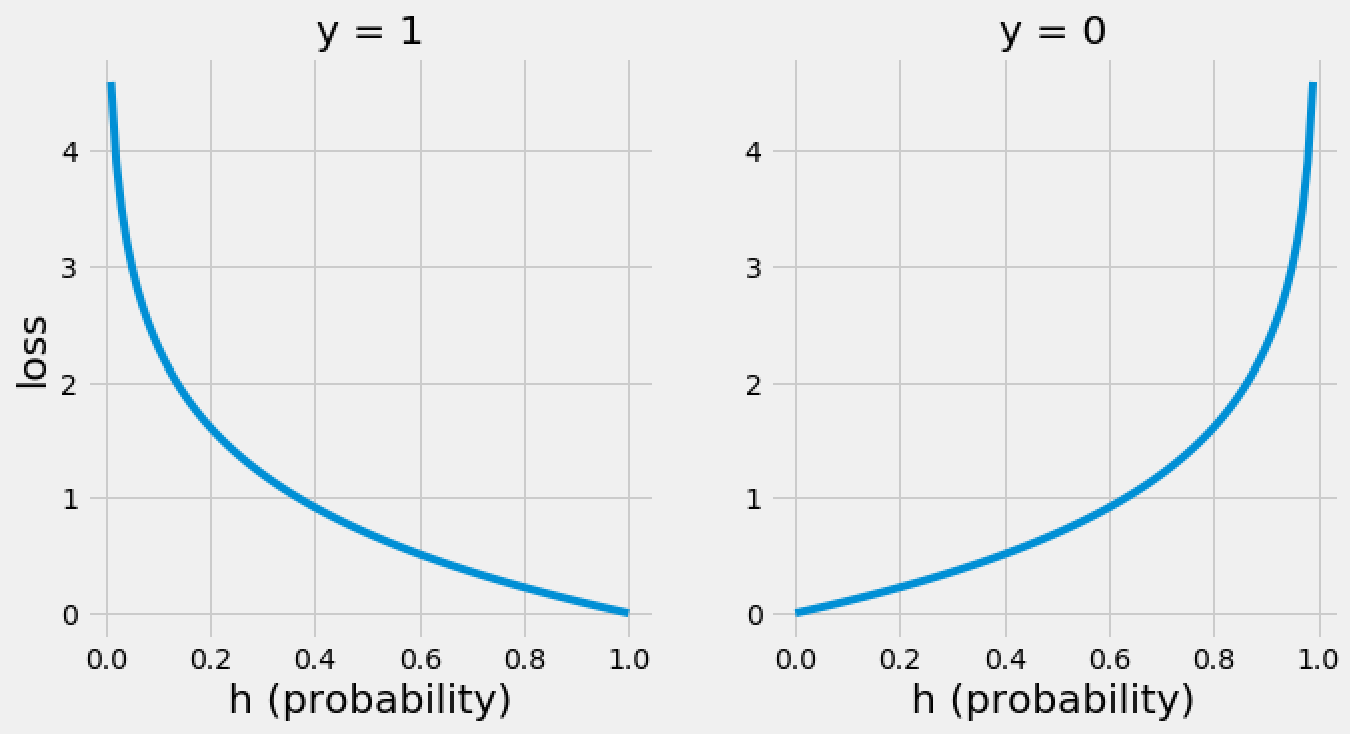
$$L = -(y\log(a) + (1-y)\log(1-a))$$

	L
$y=1$(양성클래스)	$-\log(a)$
$y=0$(음성클래스)	$-\log(1-a)$

이 때 a는 활성화 함수의 결과 값으로 0~1사이의 범위를 갖는다.

로지스틱 손실 함수

양성클래스 음성클래스



L의 값이 최소화 될 때, 샘플이 올바르게 분류되는 방향으로 a 값이 구해짐

즉, L의 값이 최소화될 때 로지스틱 회귀 모델 목표가 달성

로지스틱 손실 함수

	제곱 오차의 미분	로지스틱 손실 함수의 미분
가중치에 대한 미분	$\frac{\partial SE}{\partial w} = -(y - \hat{y})x$	$\frac{\partial L}{\partial w} = -(y - a)x$
절편에 대한 미분	$\frac{\partial SE}{\partial b} = -(y - \hat{y})1$	$\frac{\partial L}{\partial b} = -(y - a)1$

제곱오차의 미분결과와 매우 유사

가중치와 절편을 업데이트하는 방법은 기존 가중치에 미분값을 빼는 것

$$w = w - \left(\frac{\partial L}{\partial w}\right) = w + (y - a)x$$

$$b = b - \left(\frac{\partial L}{\partial b}\right) = b + (y - a)1$$