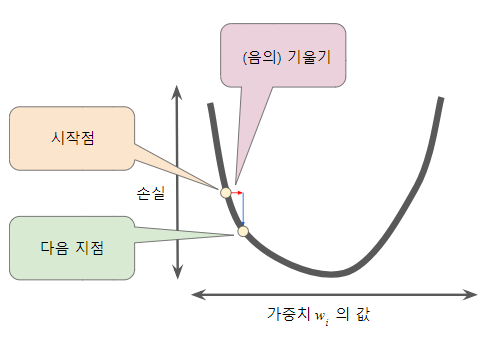
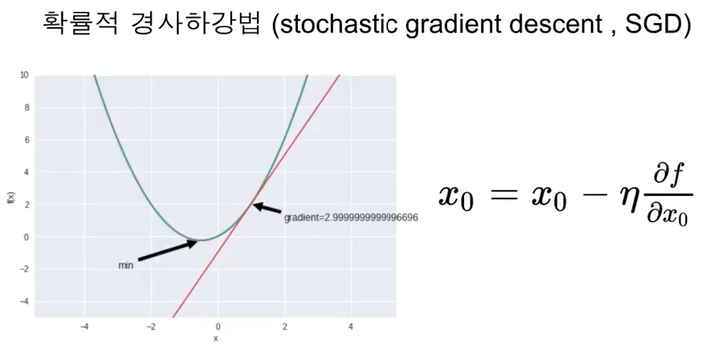
**SGD( Stochastic Gradient Descent)**

**Gradient descent ( 경사하강법 ) 이란?**

간단하게 말해, 미분(기울기)가 최소가 되는 점을 찾아 알맞은 weight(가중치 매개변수) 찾는 방법이다. => 손실(cost) 줄이는 방법!





기울기

{\ displaystyle \ eta} : step size ( 기계 학습에서 *learning rate* 라고도 함)

경사하강법의 첫 번째 단계는 w1에 대한 시작 값(시작점)을 선택하는 것입니다. 시작점은 별로 중요하지 않습니다. 따라서 많은 알고리즘에서는 w1을 0으로 설정하거나 임의의 값을 선택합니다. 다음 그림에서는 0보다 조금 큰 시작점을 지정했습니다.

그런 다음 경사하강법 알고리즘은 시작점에서 손실 곡선의 기울기를 계산합니다. 간단히 설명하자면 **기울기**는 편미분의 벡터로서, 어느 방향이 '더 정확한지' 혹은 '더 부정확한지' 알려줍니다. 그림에 나온 것과 같이 단일 가중치에 대한 손실의 기울기는 미분 값과 같습니다.

손실 함수 곡선의 다음 지점을 결정하기 위해 경사하강법 알고리즘은 다음과 같이 기울기의 크기의 일부를 시작점에 더합니다. **기울기 보폭(Learning rate)을 통해 손실 곡선의 다음 지점으로 이동합니다.** 그런 다음 경사하강법은 이 과정을 반복해 최소값에 점점 접근합니다.

**SGD란?**

배치 크기가 1인 [**경사하강법**](https://developers.google.com/machine-learning/crash-course/glossary?hl=ko#gradient_descent) 알고리즘입니다. 즉, 확률적 경사하강법은 데이터 세트에서 **무작위**로 균일하게 선택한 하나의 예에 의존하여 각 단계의 예측 경사를 계산합니다.

경사하강법에서 **배치**는 단일 반복에서 기울기를 계산하는 데 사용하는 예의 총 개수입니다. 지금까지는 배치가 전체 데이터 세트라고 가정했습니다. 하지만 Google 규모의 작업에서는 데이터 세트에 수십억, 수천억 개의 예가 포함되는 경우가 많습니다. 또한 대개 Google 데이터 세트에는 엄청나게 많은 특성이 포함되어 있습니다. 따라서 배치가 거대해질 수 있습니다. 배치가 너무 커지면 단일 반복으로도 계산하는 데 오랜 시간이 걸릴 수 있습니다.

무작위로 샘플링된 예가 포함된 대량의 데이터 세트에는 중복 데이터가 포함되어 있을 수 있습니다. 실제로 배치 크기가 커지면 중복의 가능성도 그만큼 높아집니다. 적당한 중복성은 노이즈가 있는 기울기를 평활화하는 데 유용할 수 있지만, 배치가 거대해지면 예측성이 훨씬 높은 값이 대용량 배치에 비해 덜 포함되는 경향이 있습니다.

만약에훨씬 적은 계산으로 적절한 기울기를 얻을 수 있다면 어떨까요? 데이터 세트에서 예를 무작위로 선택하면 (노이즈는 있겠지만) 훨씬 적은 데이터 세트로 중요한 평균값을 추정할 수 있습니다. **확률적 경사하강법**(**SGD**)은 이 아이디어를 더욱 확장한 것으로서, 반복당 하나의 예(배치 크기 1)만을 사용합니다. **'확률적(Stochastic)'이라는 용어는 각 배치를 포함하는 하나의 예가 무작위로 선택된다는 것을 나타냅니다.**

단점: 반복이 충분하면 SGD가 효과는 있지만 노이즈가 매우 심합니다. 확률적 경사하강법의 여러 변형 함수의 최저점에 가까운 점을 찾을 가능성이 높지만 항상 보장되지는 않습니다.

추가)

**미니 배치 확률적 경사하강법**(**미니 배치 SGD**)는 전체 배치 반복과 SGD 간의 절충안입니다. 미니 배치는 일반적으로 무작위로 선택한 10개에서 1,000개 사이의 예로 구성됩니다. 미니 배치 SGD는 SGD의 노이즈를 줄이면서도 전체 배치보다는 더 효율적입니다.

간단한 설명을 위해 단일 특성에 대한 경사하강법에 중점을 두었습니다. 물론 경사하강법은 여러 개의 특성을 갖는 특성 세트에도 사용 가능합니다.

출처: 구글 머신러닝 단기 집중과정

(<https://developers.google.com/machine-learning/crash-course/reducing-loss/gradient-descent?hl=ko>)