

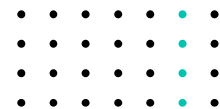
10

인공신경망 튜닝

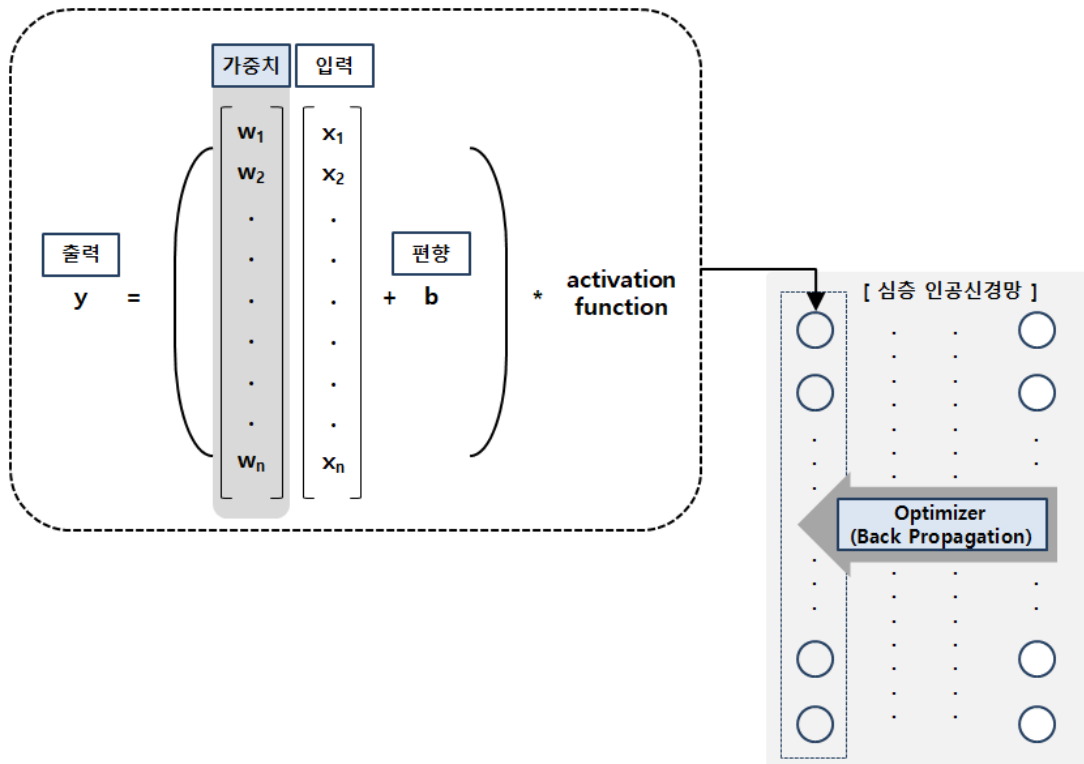
2. 튜닝 II



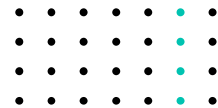
인공신경망 튜닝 II 가중치 초기화



인공신경망 가중치

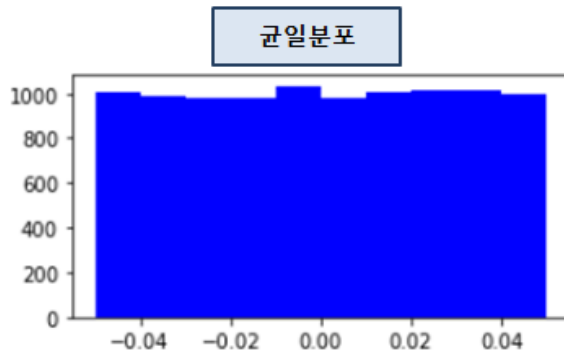
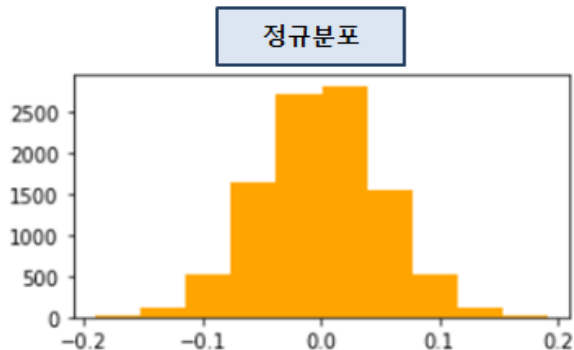


인공신경망 튜닝 II 가중치 초기화

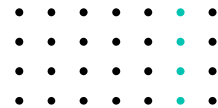


정규분포와 균일분포

- 정규분포 : 특정 값을 중심으로 데이터가 종모양 대칭형으로 분포한 구조
- 균일분포 : 특정 범위에서 데이터가 균일하게 분포하고 있는 형태



인공신경망 튜닝 II 가중치 초기화



Glorot 초기화와 He 초기화

- Glorot 초기화와 He 초기화 모두 데이터를 정규분포 또는 균일분포에 따라 초기화할 수 있는 방법을 제공하고

균등 분포

정규 분포

Glorot

$$r = \sqrt{\frac{6}{n_{in} + n_{out}}}$$

$$\sigma = \sqrt{\frac{6}{n_{in} + n_{out}}}$$

He

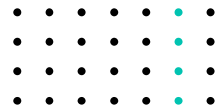
$$r = \sqrt{2} \sqrt{\frac{6}{n_{in} + n_{out}}}$$

$$\sigma = \sqrt{2} \sqrt{\frac{6}{n_{in} + n_{out}}}$$

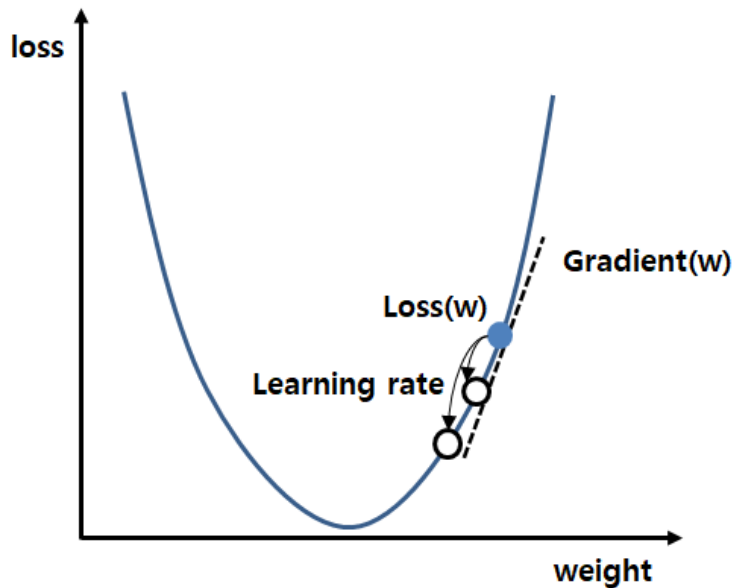
- R : 값의 범위
- σ : 분산
- n_{in} : 입력 노드 개수
- n_{out} : 출력 노드 개수



인공신경망 튜닝 II 최적화 알고리즘



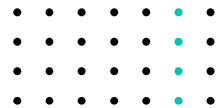
경사하강법



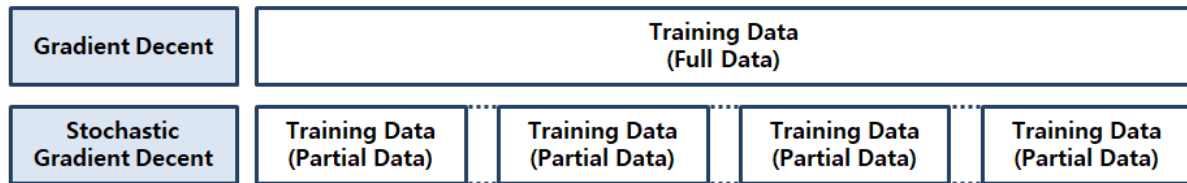
경사하강법은 편미분을 통해 구한 기울기로 함수 (손실 함수) 의 최저점을 찾아가는 알고리즘



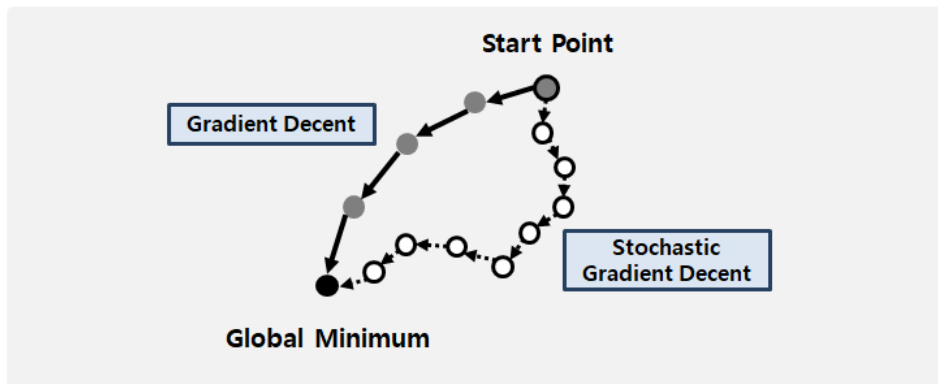
인공신경망 튜닝 II 최적화 알고리즘



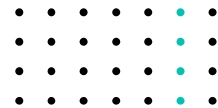
경사하강법과 확률적 경사하강법



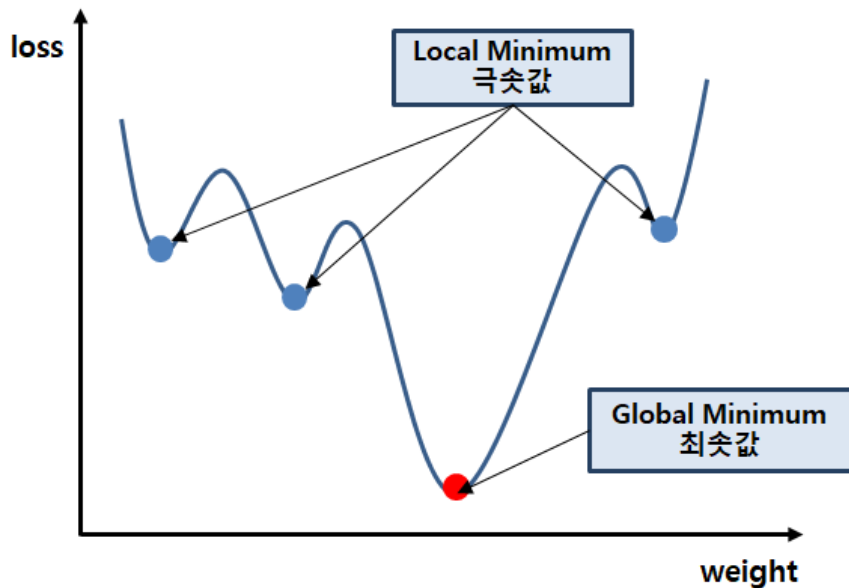
확률적 경사하강법: 데이터를 샘플링 해서 작은 데이터를 학습하고, 이 과정을 여러 번 반복 하면서 확률적으로 전체 데이터를 학습하는 것과 비슷한 효과를 나타내게 한다는 것



인공신경망 튜닝 II 최적화 알고리즘



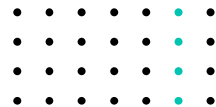
지역 최소값과 전역 최소값



- 극솟값(Local Minimum)
 - 손실함수 곡선에서 중간중간 오목하게 들어간 부분
- 최솟값(Global Minimum)
 - 중간에 가장 크게 오목하게 들어간 부분



인공신경망 튜닝 II 최적화 알고리즘



다양한 최적화 알고리즘

모멘텀 (Momentum)

- 가중치 변경 값에 관성(momentum)을 추가
- 관성 때문에 기울기가 0인 부분에서도 가중치 업데이트 가능

아다그래드 (AdaGrad)

- Adaptive Gradient
- 학습 속도(learning rate)를 매개 변수에 맞게 조정
- 빈번한 매개변수 학습 속도 낮추고, 업데이트 횟수 줄임

알엠에스프롭 (RMSProp)

- 시간이 지남에 따라 학습 속도가 줄어드는 AdaGrad 알고리즘 단점 개선
- 최근 학습한 데이터가 학습 속도 변경에 좀 더 많이 영향을 미치도록 설계

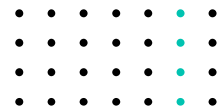
아담 (Adam)

- Momentum 알고리즘 + RMSProp 알고리즘
- 가중치와 학습 속도를 모두 변경하면서 안정적으로 손실함수 최솟값 학습



인공신경망 튜닝 II

노드와 은닉층 개수에 대한 논의



일반적 논의	단일 은닉층으로 모든 데이터 표현 가능
	보다 많은 은닉층과 노드들이 보다 복잡한 데이터 표현 가능
	은닉층에 너무 적은 노드를 사용하면 과소적합(underfitting) 발생
	은닉층에 너무 많은 뉴런을 사용하면 과대적합(overfitting) 발생
경험적 논의	은닉층 노드 수는 입력 레이어 크기와 출력 레이어 크기 사이
	은닉층 노드 수는 입력 레이어 크기의 2/3에 출력 레이어 크기를 더한 값
	은닉층 노드 수는 입력 레이어 크기의 두 배 미만

