

제7장: 딥러닝 최적화와 모형진단

Soyoung Park

Pusan National University
Department of Statistics

`soyoung@pusan.ac.kr`

최적화와 딥러닝 모형진단

1

2

3

4

5

출력층과 손실함수 - 회귀(regression)

출력층과 손실함수 - 분류(classification)

출력층과 손실함수 - 분류(classification)

출력층과 손실함수 - 분류(classification)

출력층과 손실함수 - 분류(classification)

출력층과 손실함수 - 분류(classification)

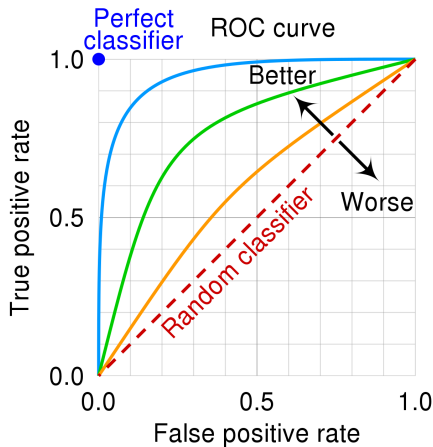
분류에서의 정밀도

- Sensitivity(민감도), True positive rate, Recall
- Specificity(특이도), True negative rate
- False positive rate(FRP) = $1 - \text{Specificity}$
- Accuracy
- Precision
- Error Rate

1

¹<https://loskatsu.github.io/machine-learning/stat-roc-curve>

ROC curve



2

²https://en.wikipedia.org/wiki/Receiver_operating_characteristic

출력층의 활성화함수, 손실함수, 정밀도

표 3-1 출력층의 활성화함수, 손실함수, 그리고 정밀도

목적	출력층의 활성함수	손실함수	정밀도
회귀	항등	MSE 또는 SSE	MSE, MAE, MAPE
이항분류	sigmoid	binary crossentropy	accuracy, precision, recall
다항분류	softmax	binary crossentropy categorical crossentropy	accuracy, precision, recall

학습률과 모수최신화

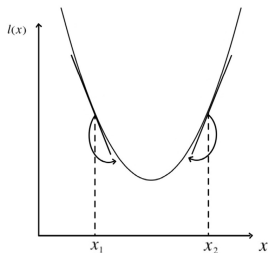


그림 3-1 손실함수의 최적화

역전파의 과정

- 1 모수의 초기를 임의로 부여하고, 딥러닝의 모형구조대로 차례로 움직여 손실함수를 구한다 (입력층 \rightarrow 은닉층 \rightarrow 출력층)
- 2 손실함수의 최소화를 위한 미분을 역순으로 진행해 모수를 최신화한다 (출력층 \rightarrow 은닉층 \rightarrow 입력층)
- 3 위의 과정을반복해 최적의 모수를 구한다 (과대적합 발생하기 전까지)

역전파의 과정 - (1) 순전파 (forward propagation)

역전파의 과정 - (1) 순전파 (forward propagation)

역전파의 과정 - (1) 순전파 (forward propagation)

역전파의 과정 - (2) 역전파 (back propagation)

역전파의 과정 - (2) 역전파 (back propagation)

역전파의 과정 - (2) 역전파 (back propagation)

역전파의 과정 - (2) 역전파 (back propagation)

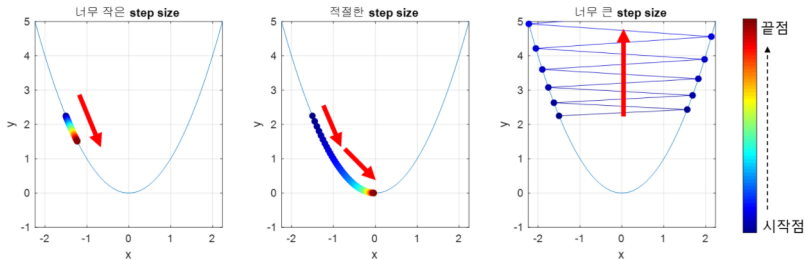
역전파의 과정 - (2) 역전파 (back propagation)

역전파의 과정 - (2) 역전파 (back propagation)

역전파의 과정 - (2) 역전파 (back propagation)

최적화알고리즘

Learning rate



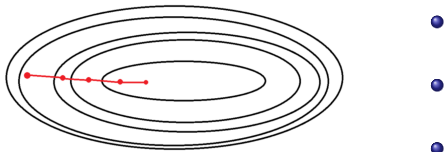
3

³https://angeloyeo.github.io/2020/08/16/gradient_descent.html

기울기하강법(Gradient Descent)

- 전체data로부터 계산된 손실함수를 최소화하는 과정
- 손실함수가 convex함수이고, 국소최소값(local minimum)이 존재하지 않을 때 매우 유용하며 우수한 성능을 보임
- 그러나, 딥러닝 손실함수는 하나 이상의 국소최소값이 존재 할 뿐만아니라, 안장점 (saddle point)이 존재하여 기울기 하강법은 일반적으로 적절하지 않음

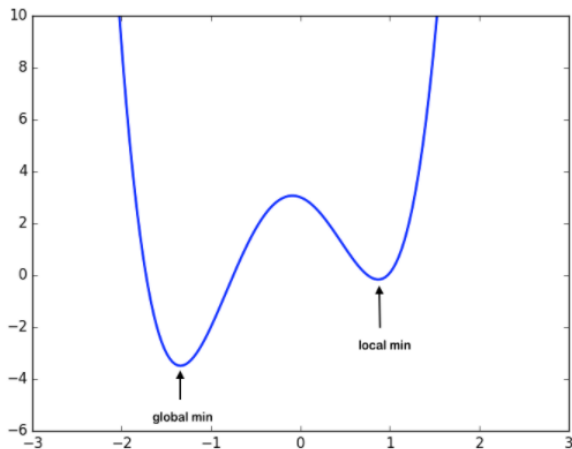
(1) 배치 경사 하강법 (Batch Gradient Descent)



4

⁴<https://skyil.tistory.com/68>

(1) 배치 경사 하강법 (Batch Gradient Descent)



5

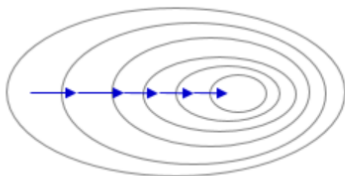
<https://>

(2) 확률적 경사 하강법 (Stochastic Gradient Descent)

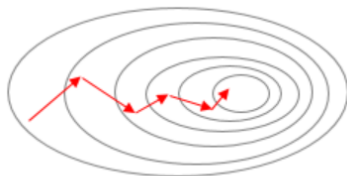
•

•

•



경사 하강법

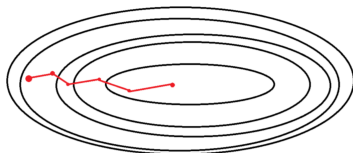


SGD

6

⁶<https://truman.tistory.com/164>

(3) 미니 배치 경사 하강법 (Mini-Batch Gradient Descent)



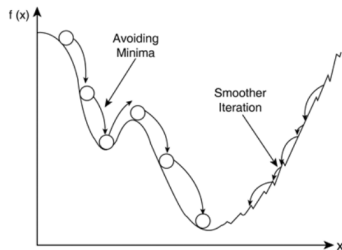
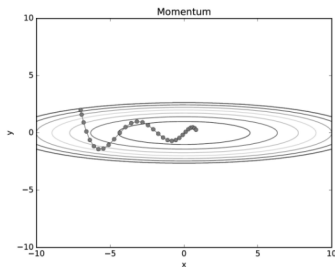
MSGD의 시각화

7

⁷<https://skyil.tistory.com/68>

(4) 모멘텀 (Momentum)

$$v_i = \alpha v_{i-1} + \eta \nabla f(x_i) \quad x_{i+1} = x_i + v_i$$



8

⁸<https://east-rain.github.io/docs/Deep%20Learning/basic%20deeplearning/optimization.html>

그외 알고리즘..

- Adagradoptimizer
- Rprop 최적화
- RMSprop
- AdaDelta
- Adam (Adaptive moment)
- Nesterov
- Nadam(Nesterov adam)

딥러닝 모형의 진단과 일반화

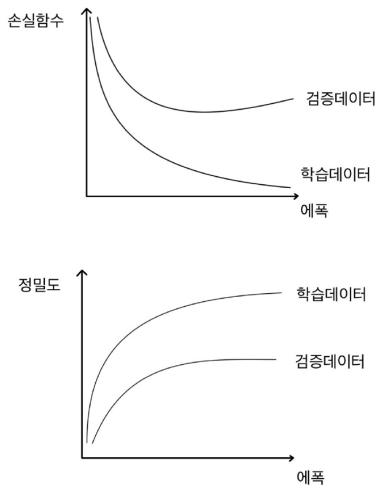


그림 3-4 학습데이터와 검증데이터의 손실함수와 정밀도