CS231n 3강. **Loss Functions and Optimization**

-Linear Classifier에서 가장 좋은 행렬 W를 구하는데 어떻게 트레이닝데이터를 활용해야 하는가?

지금만든 W가 좋은지 나쁜지 정량화 할 방법: 손실함수

텍스트, 스크린샷, 자동차, 폰트이(가) 표시된 사진

자동 생성된 설명

개구리 클래스는 잘 분류되지 못함

X: 이미지, Y: 예측하고자 하는 것

레이블 y는 1~10사이의 정수(프로그래밍 언어에따라 0~9) 이 정수값은 각 이미지 x의 정답 카테고리를 의미한다.

손실함수 L: N개 샘플들의 loss 평균

W의 공간을 탐색하며 Loss가 최소인 W를 찾는다.

-멀티 클래스 SVM의 손실함수

텍스트, 폰트, 스크린샷, 대수학이(가) 표시된 사진

자동 생성된 설명텍스트, 폰트, 스크린샷, 라인이(가) 표시된 사진

자동 생성된 설명

x: s\_yi, y: L\_i

1. 정답인 클래스의 결과값 s\_ji 와 나머지 클래스를 비교한다.
2. 정답보다 다른 클래스 점수가 더 높으면 이게 LOSS
3. loss에서 safety margin값을 추가 (적어도 이보다는커야함을의미)
4. loss값이 0보다 작은 음수이면 포함하지않음

-예시: 고양이는 3.2의 값을 가지고 나머지 자동차와 개구리에 대한 loss값

Loss(class:cat) = 0

Loss(class:car) = max(0, 5.1-3.2+1) = 0

Loss(class:frog) = max(0,-1.7-3.2+1)=2.9

* L\_i = 2.9 +0 = 2.9

텍스트, 스크린샷, 자동차, 차량이(가) 표시된 사진

자동 생성된 설명

같은 방법으로 나머지 두 사진에 대해 반복한 결과

Multiclass SVM 값은 5.27

-특징:

정답 스코어가 safety margin 을 만족하는 범위이면 loss값에 변화 안줌

Loss = 0일 때 w파라미터에 임의의 상수를 곱해도 loss=0

위 점에서 multiclass SVM loss는 유일하지 않다

모든 s=0이면 loss= 클래스 수 -1 (loss function을 디버그하기 좋다)

loss최소값은 0 최댓값은 무한대

-Regulation: 과적합 해결방법

과적합: train 에 대해서는 좋은 성능을 가지게 학습되지만 test 데이터에 대해서는 성능이 떨어지는 현상

Regulation은 w값에 대하여 제약을 가한다.

* Weight decay : 특정 가중치 비이상적으로 커지는 것 방지
* Local noise, outlier 의 영향을 덜 받게 한다.

텍스트, 폰트, 도표, 라인이(가) 표시된 사진

자동 생성된 설명

-Regulation function(=R() )의 예시

종류는 L2 regularization, L1 regularization, Elasic net, Dropout등

텍스트, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명

-예시:

텍스트, 폰트, 스크린샷, 명함이(가) 표시된 사진

자동 생성된 설명

두개의 weight값 w1, w2는 x 에 대해 같은 스코어를 가지지만 분포형태는 다르다.

L1을 적용 -> 같은 loss값을 가짐

L2를 적용 -> w1 loss = 1, w2 loss = 4x(0.5)^2 = 1/4

-특징:

L1: weight값이 0으로 수렴하는 것이 많은 형태 희소행렬이라부름. (어떤특징은무시함)

L2: weight값이 큰 것은 점점 줄이며 대부분값을 0에 가까운 가우시안분포를 가짐(특징 조금 참고)

-Softmax Classifier (= Logistic Regression)

텍스트, 스크린샷, 폰트, 라인이(가) 표시된 사진

자동 생성된 설명

위 Mutinominal SVM과 다르게 logistic 한 값을 사용, Multi SVM에서 정답 스코어값이 다른값보다 safety 마진만큼 크면 되는점과 달리 정답값의 분포비중을 보고 loss 값을 결정한다.

-예시: loss값을 가지는 과정

텍스트, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명

1. 스코어 값에 exponental취함
2. 확률분포값을 구함
3. 이후 정답제외 나머지 클래스에 대해 -log를 취해 더함

-SVM(hinge loss)과 Softmax(cross-entropy loss)의 차이

텍스트, 스크린샷, 도표, 번호이(가) 표시된 사진

자동 생성된 설명

-Optimization: 어떻게 좋은 weight값을 찾아가야할지 다룸

-Random 서치: 아무기준없음

텍스트, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명

텍스트, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명

-Local geometry 경사하강법: 주변보다 낮은곳으로 조금씩 움직임

실제로 weight 값 기울기 구하는 법: 미분

텍스트, 스크린샷, 폰트, 대수학이(가) 표시된 사진

자동 생성된 설명

-Numerical 기울기: 근사값, 느림, 코드짜기 쉬움

-Anaytic 기울기: 정확한값을 구함 빠름, 코드짜기 어려움

-Stochastic Gradient Descent(SGD): 전체 데이터셋 기울기 대신 미니배치라는 작은 트레이닝샘플로 나누어 학습하는 방법

-특징변환:

-컬러 히스토그램:

-Orented Gradient의 히스토그램

-Bag of Words