1 소개

3 Code Collab 주소

4 inception module

inception module

maxpooling다음에 1\*1convolution 이 들어가는 이유는

maxpooling은 채널수를 조정할 수 없기 때문

5 bias\_initializer

a bias value allows you to shift the activation function to the left or right,

A simpler way to understand what the bias is: it is somehow similar to the constant b of a linear function

y = ax + b

It allows you to move the line up and down to fit the prediction with the data better.

Axis 합치는 차원

6 Cifar 10 data set 머신러닝용 이미지 분류 셋

224 \* 224 전처리 과정

트레이닝셋, 테스트셋

sub\_text="점심 먹으러 갈래 메뉴는 햄버거 최고야"

[[0. 0. 1. 0. 0. 0. 0. 0.] #인덱스 2의 원-핫 벡터

[0. 0. 0. 0. 0. 1. 0. 0.] #인덱스 5의 원-핫 벡터

[0. 1. 0. 0. 0. 0. 0. 0.] #인덱스 1의 원-핫 벡터

[0. 0. 0. 0. 0. 0. 1. 0.] #인덱스 6의 원-핫 벡터

[0. 0. 0. 1. 0. 0. 0. 0.] #인덱스 3의 원-핫 벡터

[0. 0. 0. 0. 0. 0. 0. 1.]] #인덱스 7의 원-핫 벡터

8 padding이 same 이면 입력과 출력을 같게

9 Local Response Normalization

측면억제 : 신경망 관련용어

흰색 선에 집중 안하면 회색의 점이 보이는게 측면억제에 의한 것

Alexnet에서 이후 Inception모델 ReLU함수 쓰는 것이 CONV POOLING시 하나의 픽셀값이 주변의 픽셀에 영향을 미쳐 이런부분을 방지하기 위해 정규화를 해줌

12 inception\_4a에서 auxiliary Learning 1

Inception\_4d에서 auxiliary Learning 2

논문에서 주의깊게 안본부분 이었습니다.

Dense : 밀집층을 구현할 때 쓰이는 함수

Drop out : 정규화

21학습시키기

Epoch : 똑 같은 시험지를 몇번 풀어볼지 (Epoch값이 크면 overfitting)

Batch : 몇 개를 풀어보고 해답을 맞춰볼지

배치 사이즈가 큰 경우 기울기를 계산하기 위해 더 많은 데이터를 사용하게 되므로 우리가 최적화 시켜야하는 전체 학습용 데이터를 사용한 해공간의 기울기 값과 유사한 기울기를 사용하므로 최적화가 더 수월해질 수 있다

반면에 배치 사이즈가 작은 경우 상대적으로 부정확한 기울기를 사용한다는 단점이 있지만, 한번의 업데이트에 적은 계산 비용이 들어가 한번 업데이트 할 동안 여러번의 업데이트를 수행할 수 있고, 기울기의 부정확한 면이 랜덤성으로 작용해 실제 기울기가 낮은 구간이나 극소점, 안장점에서 쉽게 벗어날 가능성이 있다는 장점이 있다.

SGD Stochastic gradient descent(SGD)

추출된 데이터 한 개에 대해 error gradient 를 계산 Gradient descent 알고리즘을 적용하는 방법.

Gradient Descent  
네트워크에서 내놓은 결과값과 실제값 사이의 차이를 정의하는 lossfunction 값을 최소화하기 위해 기울기를 이용

Lossfunction의 값을 최소화하는 값을 찾는것으로 기울기 반대방향으로 이동하는 것을 반복

BGD, SGD 전체데이터를 쓰는게 BGD

아직은 이해 못할 심오한 개념들이 많이 쓰여서 이해를 못했음