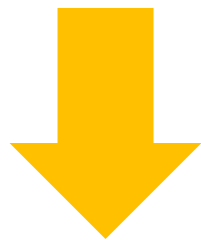


cs231n 3강

Lost Function & Optimization

Lost function(손실함수): 예측하여 각 클래스에 대한 점수를 정답 값과 비교했을때 차이가 얼마큼 나는지 정도를 보여주는 함수



Optimization(최적화): Lost function의 결과를 최소화하기 위하여 parameter를 찾는 과정

Lost function(손실함수): 예측하여 각 클래스에 대한 점수를 정답 값과 비교했을때 차이가 얼마큼 나는지 정도를 보여주는 함수



cat	3.2	1.3	2.2
car	5.1	4.9	2.5
frog	-1.7	2.0	-3.1

여러 개의 class에 대하여 결과를 도출해야하므로
Multiclass SVM loss

Multiclass SVM loss:

Given an example (x_i, y_i)
where x_i is the image and
where y_i is the (integer) label,

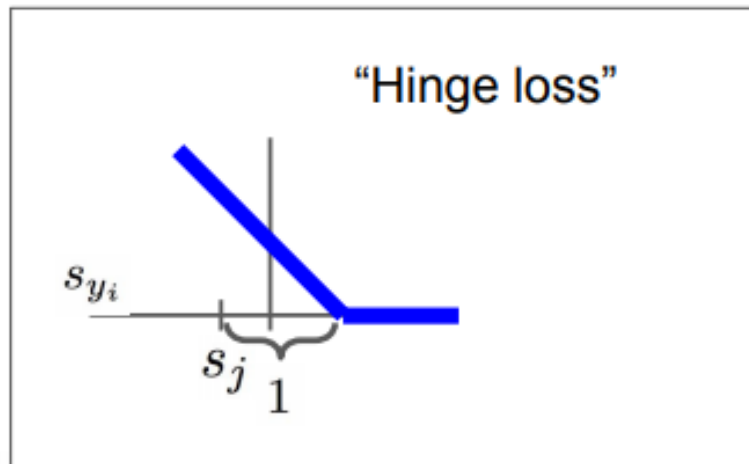
and using the shorthand for the
scores vector: $s = f(x_i, W)$

the SVM loss has the form:

$$L_i = \sum_{j \neq y_i} \begin{cases} 0 & \text{if } s_{y_i} \geq s_j + 1 \\ s_j - s_{y_i} + 1 & \text{otherwise} \end{cases}$$
$$= \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)$$

<SVM loss의 그래프>

Multiclass SVM loss:



Suppose: 3 training examples, 3 classes.
With some W the scores $f(x, W) = Wx$ are:



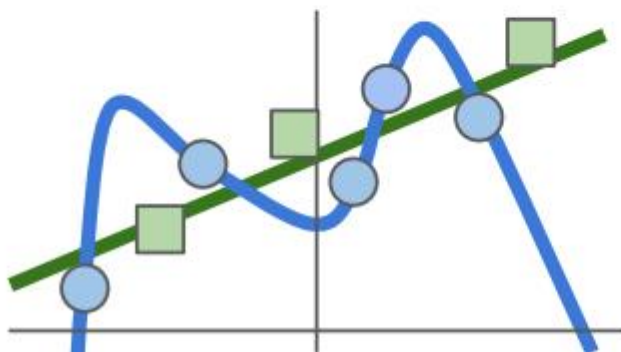
cat	3.2	1.3	2.2
car	5.1	4.9	2.5
frog	-1.7	2.0	-3.1
Losses:	2.9	0	12.9

<SVM loss의 그래프 – hinge loss>

- 이미 어떤 라벨에 대하여 정답인 값이 정답이 아닌 값보다 매우 크면 그 값이 조금 바뀐다 하더라도 loss는 변하지 않는다.
- 그래프와 같이 loss의 최솟값은 0이고, 최댓값은 무한대의 값이다.
- Squared hinge loss를 사용한다면 분류가 잘되었는지의 차이를 더 크게 확인할 수 있다.
- 같은 loss 값을 가지는 여러 개의 W 가 존재할 수 있다.

Regularization(정규화): overfitting이 일어나지 않는 단순한 model로 만드는 방법

$$L(W) = \underbrace{\frac{1}{N} \sum_{i=1}^N L_i(f(x_i, W), y_i)}_{\text{Data loss: Model predictions should match training data}} + \underbrace{\lambda R(W)}_{\text{Regularization: Model should be "simple", so it works on test data}}$$



- L1 regularization
- L2 regularization
- Elastic net(L1+L2)
- Max norm regularization
- Dropout
- Batch normalization, stochastic depth

Softmax Loss function

Softmax Classifier (Multinomial Logistic Regression)



$$L_i = -\log\left(\frac{e^{s_{y_i}}}{\sum_j e^{s_j}}\right)$$

unnormalized probabilities

cat
car
frog

3.2

5.1

-1.7

exp

24.5

164.0

0.18

normalize

0.13

0.87

0.00

$$L_i = -\log(0.13) = 0.89$$

unnormalized log probabilities

probabilities

모든 각 카테고리에 대하여 결과 값을 내고 확률로 나타내기-> 정규화로 0-1사이 값으로 만들기

Optimization(최적화): Lost function의 결과를 최소화하기 위하여 parameter를 찾는 과정(W 구하기)

Optimization의 방법:

1. 임의로 W 선택하기(X)
2. Gradient Descent, Stochastic Gradient Descent

$$L(W) = \frac{1}{N} \sum_{i=1}^N L_i(x_i, y_i, W) + \lambda R(W)$$

$$\nabla_W L(W) = \frac{1}{N} \sum_{i=1}^N \nabla_W L_i(x_i, y_i, W) + \lambda \nabla_W R(W)$$

