CS231n- Lecture 3

February 14, 2017

1 Optimization

1.1 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

$$L_i = \sum_{j \neq y_i} max(0, s_j - s_{y_i} + 1)$$

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	10.9

Q: What if the sum was instead over all class?(i.e including $j = y_i$)
A: Score is just being inflated by one as $j = y_i$

Q: What if mean was used instead of sum?
A: Pointless as it is only a difference of $c_i * L_i$

Q: What if instead you use $L_i = \sum_{j \neq y_i} max(0, s_j - s_{y_i} + 1)^2$
A: This changes the loss function.

Q: Min/Max values?
A: Min=0. Max= Infinite

Q: Usually at initialization W are small numbers so all s = 0. What is the loss?
A: Loss becomes $number\ of\ classes - 1$. This serves as a good sanity check.

Example numpy code for:

$$L_i = \sum_{j \neq y_i} max(0, s_j - s_{y_i} + 1)$$

```
def L_i_vectorized(x, y, W):
        scores= W.dot(x)
        margins=np.maximum(0, scores - scores[y] + 1)
        margins[y] = 0
        loss_i = np.sum(margins)
        return loss_i
```

We now have

$f(x, W) = Wx$

$L = \frac{1}{N} \sum_{j \neq y_i} max(0, s_j - s_{y_i} + 1)$

Suppose we found a W such that L=0. Is this W unique?
No. If we double W. Loss is the same. Same goes if you multiply it with a number $n : n >= 1$
So there is a huge subspace of W for which this is optimal.
Weight Regularisation
$L = \frac{1}{N} \sum_{i=1}^{N} \sum_{j \neq y_i} max(0, f(x_i; W)_j - f(x_i, W)_{y_i} + 1) + \lambda R(W)$
In Common use:
L2 Regularization: $R(W) = \sum_k \sum_l W_{k,l}^2$
L1 Regularization: $R(W) = \sum_k \sum_l |W_{k,l}|$
Elastic Net(L1+ L2): $R(W) = \sum_k \sum_l \beta W_{k,l}^2 + |W_{k,l}|$
Max norm regularization(will see later)
Dropout(Will see later)
L2 is most popular. λ indicate the regularisation strength.