Code Illustration

Zhedong Zheng

August 9, 2017

1 Forward Loss

In this doc, the label and equation may be informal. But we think it can help you to understand our code in a short time.

The classical softmax loss is as follow:

$$Loss = -log(P(c)) = -log(\frac{e^{X_c - X_{max}}}{\sum e^{X - X_{max}}}) \tag{1}$$

$$= -X_c + X_{max} + log(\sum e^{X - X_{max}}), \tag{2}$$

where c is ground truth class label.

The label smooth loss is as follow:

$$Loss = -\frac{1}{K}\sum_{i=1}^{K} log(P) \tag{3}$$

$$= -\frac{1}{K}\sum_{i=1}^{K} log(\frac{e^{X_i - X_{max}}}{\sum_{j=1}^{K} e^{X_j - X_{max}}}) \tag{4}$$

$$= -\frac{1}{K}\sum_{i=1}^{K} log(e^{X_i - X_{max}}) + \frac{1}{K} \times K \times log(\sum_{j=1}^{K} e^{X - X_{max}}) \tag{5}$$

$$= -\frac{1}{K}\sum_{i=1}^{K} (X_i - X_{max}) + log(\sum e^{X - X_{max}}). \tag{6}$$

In the paper, we use the Eq. 1 and Eq. 3;
In the code, we actually use the Eq. 2 and Eq. 6.

2 Backward Gradient

According to Eq. 2, The classical softmax gradient is as follow: (We denote that $E = e^{X_i - X_{max}}$)

$$Loss = -X_c + X_{max} + log(\sum E) \tag{7}$$

$$(log(\sum E))' = \frac{(\sum E)'}{\sum E} = \frac{E}{\sum E} \tag{8}$$

$$(-X(c))' = -1 \tag{9}$$

According to Eq. 6, The labelsmooth gradient is as follow:

$$Loss = -\frac{1}{K}\sum_{i=1}^{K}(X_i - X_{max}) + log(\sum E) \tag{10}$$

$$Loss' = -\frac{1}{K} + \frac{E}{\sum E} \tag{11}$$