Gradient computing of Multi-class SVM

Hanhe Lin

Here is the gradient computing I found in Stanford CS231N course. For simplicity, let us use the loss function of multi-class SVM with only one example and without regularization term, i.e.

$$L(w, b) = \sum_{j=1, j \neq y^{(i)}}^{K} \max\left(0, (w^{(j)}x^{(i)} + b^{(j)}) - (w^{(y^{(i)})}x^{(i)} + b^{(y^{(i)})}) + \Delta\right)$$

Compute the gradient with respect to $w^{(y^{(i)})}$, we get:

$$\frac{\partial L}{\partial w^{(y^{(i)})}} = -\left(\sum_{j=1, j \neq y^{(i)}}^{K} \mathbb{1}\{(w^{(j)}x^{(i)} + b^{(j)}) - (w^{(y^{(i)})}x^{(i)} + b^{(y^{(i)})}) + \Delta > 0\}\right) x^{(i)}$$

For the other weights where $j \neq y^{(i)}$, the gradient is:

$$\frac{\partial L}{\partial w^{(j)}} = \mathbb{1}\{(w^{(j)}x^{(i)} + b^{(j)}) - (w^{(y^{(i)})}x^{(i)} + b^{(y^{(i)})}) + \Delta > 0\}x^{(i)}$$