SVM 以Hinge Loss 为例

Hinge Loss Function

$$L_i = \sum_{j \neq y_i} [\max(0, x_i w_j - x_i w_{y_i} + \Delta)]$$

- i iterates over all N examples .
- j iterates over all C classes .
- L_i is loss for classifiying a single example x_i (row vector) .
- w_j is the weights (column vector) for computing the score of class j .
- y_i is the index of the correct class of x_i .
- Δ is a margin parameter

$$\nabla_w \mathcal{L}_i = \begin{bmatrix} \frac{d\mathcal{L}_i}{dw_1} & \frac{d\mathcal{L}_i}{dw_2} & \cdots & \frac{d\mathcal{L}_i}{dw_C} \end{bmatrix} = \begin{bmatrix} \frac{d\mathcal{L}_i}{dw_{11}} & \frac{d\mathcal{L}_i}{dw_{21}} & \dots & \frac{d\mathcal{L}_i}{dw_{y_i 1}} & \dots & \frac{d\mathcal{L}_i}{dw_{C1}} \\ \vdots & \ddots & & & & \\ \frac{d\mathcal{L}_i}{dw_{1D}} & \frac{d\mathcal{L}_i}{dw_{2D}} & \dots & \frac{d\mathcal{L}_i}{dw_{y_i D}} & \dots & \frac{d\mathcal{L}_i}{dw_{CD}} \end{bmatrix}$$

$$\begin{aligned} \mathcal{L}_i = & \max(0, x_{i1} w_{11} + x_{i2} w_{12} \ldots + x_{iD} w_{1D} - x_{i1} w_{y_i 1} - x_{i2} w_{y_i 2} \ldots - x_{iD} w_{y_i D} + \Delta) + \\ & \max(0, x_{i1} w_{21} + x_{i2} w_{22} \ldots + x_{iD} w_{2D} - x_{i1} w_{y_i 1} - x_{i2} w_{y_i 2} \ldots - x_{iD} w_{y_i D} + \Delta) + \\ & \vdots \\ & \max(0, x_{i1} w_{C1} + x_{i2} w_{C2} \ldots + x_{iD} w_{CD} - x_{i1} w_{y_i 1} - x_{i2} w_{y_i 2} \ldots - x_{iD} w_{y_i D} + \Delta) \end{aligned}$$

For a general case , if $(x_i w_1 - x_i w_{y_i} + \Delta) > 0$

$$\frac{d\mathcal{L}_i}{dw_{11}} = x_{i1}$$

using an indicator function:

$$\frac{d\mathcal{L}_i}{dw_{11}} = \mathbb{I}(x_i w_1 - x_i w_{y_i} + \Delta > 0) x_{i1}$$

同样的

$$\begin{aligned} \frac{d\mathcal{L}_i}{dw_{12}} &= \mathbb{I}(x_i w_1 - x_i w_{y_i} + \Delta > 0) x_{i2} \\ \frac{d\mathcal{L}_i}{dw_{13}} &= \mathbb{I}(x_i w_1 - x_i w_{y_i} + \Delta > 0) x_{i3} \\ &\vdots \\ \frac{d\mathcal{L}_i}{dw_{1D}} &= \mathbb{I}(x_i w_1 - x_i w_{y_i} + \Delta > 0) x_{iD} \end{aligned}$$

因此

$$
\begin{aligned}
\frac{d\mathcal{L}_i}{dw_j} &= \mathbb{I}(x_i w_j - x_i w_{y_i} + \Delta > 0) \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{iD} \end{bmatrix} \\
&= \mathbb{I}(x_i w_j - x_i w_{y_i} + \Delta > 0) x_i
\end{aligned}
$$

对于 $j = y_i$的特殊情况

$$
\frac{d\mathcal{L}_i}{dw_{y_{i1}}} = -(\ldots) x_{i1}
$$

The coefficent of x_{i1} is the number of classes that meet the desire margin. Mathematically speaking, $\sum_{j \neq y_i} \mathbb{I}(x_i w_j - x_i w_{y_i} + \Delta > 0)$

因此

$$
\begin{aligned}
\frac{d\mathcal{L}_i}{dw_{y_i}} &= -\sum_{j \neq y_i} \mathbb{I}(x_i w_j - x_i w_{y_i} + \Delta > 0) \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{iD} \end{bmatrix} \\
&= -\sum_{j \neq y_i} \mathbb{I}(x_i w_j - x_i w_{y_i} + \Delta > 0) x_i
\end{aligned}
$$