

$X$   $(N, D)$  这里  $(1000, 3073)$

$W$   $(D, C)$  这里  $(3073, 10)$

$y$   $(N, 1)$  这里  $(1000, 1)$

$$\text{loss} = \sum_{i=0}^{N-1} \sum_{\hat{j}=0}^{C-1} (S_{i\hat{j}} - S_{iy_i} + 1) \quad (\hat{j} \neq y_i)$$

这里  $S_{i,:} = X_{i,:} W$   $(1, 10)$

$$\therefore S_{i\hat{j}} = X_{i,:} W_{:, \hat{j}} = \sum_{k=0}^{D-1} X_{ik} W_{k\hat{j}}$$

有  $\text{loss}_i = \sum_{\hat{j}=0}^{C-1} \max(0, X_{i,:} W_{:, \hat{j}} - X_{i,:} W_{:, y_i} + 1)$

$$\text{loss}_{i\hat{j}} = \sum_{k=0}^{D-1} \max(0, X_{ik} W_{k\hat{j}} - X_{ik} W_{ky_i} + 1)$$

若比0小，则结果为0，不产生梯度。



若比0大,  $\text{loss}_{ij} = \sum_{k=0}^{D-1} (X_{ik} W_{kj} - X_{ik} W_{kj} y_i + 1)$

$$\frac{\partial \text{loss}_{ij}}{\partial W_{:,j}} = X_i^T$$

$$\text{loss}_i = \sum_{j=0}^{C-1} \text{loss}_{ij}$$

$$\Rightarrow$$

$$\frac{\partial \text{loss}_i}{\partial W_{:,j}} = X_i^T$$

$$\frac{\partial \text{loss}_{ij}}{\partial W_{:,y^i}} = -X_i^T$$

$$\frac{\partial \text{loss}_i}{\partial W_{:,y^i}} = -X_i^T$$