h_1 X_1 score prob

$$X \rightarrow XW_1 + b_1 \rightarrow ReLU(h) \rightarrow X_1 W_2 + b_2 \rightarrow softMax(score)$$

loss sum_loss

$$\rightarrow -\log(prob_{:,y_i}) \rightarrow \sum loss + R_2$$

$$\frac{\partial sum\_loss}{\partial W_1} = \frac{\partial sum\_loss}{\partial loss} \cdot \frac{\partial loss}{\partial prob} \frac{\partial prob}{\partial score} \frac{\partial score}{\partial X_1} \cdot \frac{\partial X_1}{\partial h_1} \cdot \frac{\partial h_1}{\partial W_1}$$

$$= 1 \cdot (prob \cdot W_2^T) \cdot (h_1 > 0) \cdot X^T$$