Softmax 求导

Softmax 的输入： $f = [f_1, f_2, f_3, ..., f_n]$

Softmax 的输出： $p = [p_1, p_2, p_3, ..., p_n]$，其中 $p_i = \dfrac{e^{f_i}}{\sum_{j=1}^{n} e^{f_j}}$

交叉熵损失（单个样本）：

$\mathrm{L} = -\sum_{i=1}^{n} y_i \log(p_i)$，其中 $\mathrm{y} = [y_1, y_2, y_3, ..., y_n]$ 为当前样本的 one-hot 标签

求 $\dfrac{\partial \mathrm{L}}{\partial f_j}$， $\mathrm{j} \in [1, n]$

$$\frac{\partial L}{\partial f_j} = \frac{\partial L}{\partial p_i} \cdot \frac{\partial p_i}{\partial f_j} = -\sum_{i=1}^{n} \frac{y_i}{p_i} \cdot \frac{\partial p_i}{\partial f_j}$$

先求 $\dfrac{\partial p_i}{\partial f_j}$ ：

（1） 当 i=j 时

$$\begin{aligned}
\frac{\partial p_i}{\partial f_j} &= \partial \left(\frac{e^{f_i}}{\sum_{j=1}^{n} e^{f_j}} \right) / \partial f_j \\
&= \frac{e^{f_i} \cdot \sum - e^{f_i} \cdot e^{f_i}}{\sum \cdot \sum} \\
&= \frac{e^{f_i}}{\sum} \cdot \left(1 - \frac{e^{f_i}}{\sum} \right) \\
&= p_i \cdot (1 - p_i)
\end{aligned}$$

（2） 当 i≠j 时

$$
\begin{aligned}
\frac{\partial p_i}{\partial f_j} &= \partial\left(\frac{e^{f_i}}{\sum_{j=1}^{n} e^{f_j}}\right) / \partial f_j \\
&= \text{-}e^{f_i} \cdot \frac{1}{\sum^2} \cdot e^{f_j} \\
&= -\frac{e^{f_i}}{\sum}\frac{e^{f_j}}{\sum} \\
&= -p_i \cdot p_j
\end{aligned}
$$

之后带入：

$$
\begin{aligned}
\frac{\partial L}{\partial f_j} &= \frac{\partial L}{\partial p_i} \cdot \frac{\partial p_i}{\partial f_j} \\
&= -\sum_{i=1}^{n} \frac{y_i}{p_i} \cdot \frac{\partial p_i}{\partial f_j} \\
&= \text{-}\left[y_i\left(1-p_i\right)\right]_{i=j} + \sum_{i \neq j}^{n} y_i \cdot p_j \\
&= \sum_{i,i \neq j}^{n} y_i \cdot p_j + [y_i\, p_i]_{i=j} - [y_i]_{i=j} \\
&= \sum_{i}^{n} y_i \cdot p_j - [y_i]_{i=j} \\
&= p_j - y_j
\end{aligned}
$$

观察上式，y_i 只有在正确的标签位置为 1，其他位置都是 0，所以第一项只剩下了 p_j，当 i=j 时，$y_i = y_j = 1$，当 i≠j 时，后面一项等于 0，此时 y_j 也等于 0，所以可以统一写成 $p_j - y_j$