Softmax求导

Softmax的输入：

Softmax的输出：，其中

交叉熵损失（单个样本）：

，其中为当前样本的one-hot标签

求



先求 ：

1. 当i=j时



1. 当i≠j时



之后带入：



观察上式，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，当i=j时， ，当i≠j时，后面一项等于0，此时也等于0，所以可以统一写成