SoftMax以及CrossEntropy求导

数学推导作业

Softmax求导

原始形式:

$$S_i = \frac{e^{a_i}}{\sum_j e^{a_j}}$$

求导:

记:

$$\Sigma = \sum_j e^{a_j}$$

则求导过程:

$$\frac{\partial S_i}{\partial a_j} = \frac{\frac{\partial e^{a_i}}{\partial a_j} \cdot \Sigma - \frac{\partial \Sigma}{\partial a_j} \cdot e^{a_i}}{\Sigma^2}$$

当 i == j 时:

$$\begin{aligned} \frac{\partial S_i}{\partial a_j} = \frac{e^{a_i} \cdot \Sigma - e^{a_j} \cdot e^{a_i}}{\Sigma^2} = \frac{e^{a_i}}{\Sigma} \cdot \frac{\Sigma - e^{a_j}}{\Sigma} = S_i \cdot (1 - S_j) \\ = S_i \cdot (1 - S_i) \end{aligned}$$

当 i != j 时:

$$\frac{\partial S_i}{\partial a_j} = -\frac{e^{a_j} \cdot e^{a_i}}{\Sigma^2} = -S_i \cdot S_j$$

Softmax + CrossEntropy的求导

原始形式:

$$S_i = \frac{e^{a_i}}{\sum_j e^{a_j}}$$

$$L = -\sum y_i \log S_i$$

求导:

L对Si求偏导，n代表n个类别

$$\frac{\partial L}{\partial S_i} = -y_i \cdot \frac{1}{S_i}$$

$$\sum_{i=0}^{n} y_i = 1$$

L对ai求偏导

$$\begin{aligned} \frac{\partial L}{\partial a_i} &= \sum_j \frac{\partial L}{\partial S_j} \cdot \frac{\partial S_j}{\partial a_i} \\ &= \frac{\partial L}{\partial S_i} \cdot \frac{\partial S_i}{\partial a_i} + \sum_{j!=i} \frac{\partial L}{\partial S_j} \cdot \frac{\partial S_j}{\partial a_i} \\ &= -\frac{y_i}{S_i} \cdot S_i(1 - S_i) + \sum_{j!=i} -\frac{y_j}{S_j} \cdot (-1) S_i S_j \\ &= -y_i + y_i S_i + \sum_{y!=i} y_j \cdot S_i \\ &= S_i - y_i \end{aligned}$$