信号处理作业

SA17011031，黄锦静，419086888@qq.com

2018 年 5 月 4 日

Given overall network function $y_k(\boldsymbol{x},\boldsymbol{w}) = f(\sum_{j=0}^{M} w_{kj}^{(2)} h(\sum_{i=0}^{D} w_{ji}^{(1)} x_i))$ and cross-entropy error function for the multiclass classification problem $E(\boldsymbol{w}) = -ln\left(\prod_{k=1}^{K} y_k^{t_k}\right) = -\sum_{k=1}^{K} t_k ln\, y_k$, where $t_k = 0, 1^K$ and $\sum_{k=1}^{K} t_k = 1$, $f(z_k) = \frac{exp(z_k)}{\sum_{l=1}^{K} exp(z_l)}$ is the softmax function, $h(a) = tanh(a) = \frac{exp(a)-exp(-a)}{exp(a)+exp(-a)}$. Compute the derivatives of error function w.r.t the first-layer and second-layer weights with Back Propagation $\frac{\partial E}{\partial w_{ji}^{(1)}}$ and $\frac{\partial E}{\partial w_{kj}^{(2)}}$.

设 $z_k = \sum_{j=0}^{M} w_{kj}^{(2)} h(\sum_{i=0}^{D} w_{ji}^{(1)} x_i)$, $a = \sum_{i=0}^{D} w_{ji}^{(1)} x_i$.

$$\frac{\partial E}{\partial w_{kj}^{(2)}}$$
$$= \frac{\partial E}{\partial y_k} \times \frac{\partial y_k}{\partial z_k} \times \frac{\partial z_k}{\partial w_{kj}^{(2)}}$$
$$= -\frac{t_k}{y_k} \times \frac{exp(z_k)(\sum_{l=1}^{K} exp(z_l)) - exp(z_k)^2}{(\sum_{l=1}^{K} exp(z_l))^2} \times h(\sum_{i=0}^{D} w_{ji}^{(1)} x_i)$$
$$= -\frac{t_k \sum_{l=1}^{K} exp(z_l)}{exp(z_k} \times \frac{exp(z_k)(\sum_{l=1}^{K} exp(z_l)) - exp(z_k)^2}{(\sum_{l=1}^{K} exp(z_l))^2} \times h(\sum_{i=0}^{D} w_{ji}^{(1)} x_i)$$
$$= -\frac{t_k(\sum_{l=1}^{K} exp(z_l) - exp(z_k))}{\sum_{l=1}^{K} exp(z_l)} \times h(\sum_{i=0}^{D} w_{ji}^{(1)} x_i)$$
$$= -\frac{t_k(\sum_{l=1}^{K} exp(\sum_{j=0}^{M} w_{lj}^{(2)} h(\sum_{i=0}^{D} w_{ji}^{(1)} x_i)) - exp(\sum_{j=0}^{M} w_{kj}^{(2)} h(\sum_{i=0}^{D} w_{ji}^{(1)} x_i)))}{\sum_{l=1}^{K} exp(\sum_{j=0}^{M} w_{lj}^{(2)} h(\sum_{i=0}^{D} w_{ji}^{(1)} x_i))} \times h(\sum_{i=0}^{D} w_{ji}^{(1)} x_i)$$

$$\frac{\partial E}{\partial w_{ji}^{(1)}}$$
$$= \frac{\partial E}{\partial h_j} \times \frac{\partial h_j}{\partial a} \times \frac{\partial a}{\partial w_{ji}^{(1)}}$$
$$= \sum_{k=1}^{K} \frac{\partial E}{\partial y_k} \times \frac{\partial y_k}{\partial z_k} \times \frac{\partial z_k}{\partial h_j} \times \frac{\partial h_j}{\partial a} \times \frac{\partial a}{\partial w_{ji}^{(1)}}$$
$$= \sum_{k=1}^{K} \frac{\partial E}{\partial y_k} \times \frac{\partial y_k}{\partial z_k} \times w_{kj} \times \frac{(exp(a)+exp(-a))^2 - (exp(a)-exp(-a))^2}{(exp(a)+exp(-a))^2} \times x_i$$
$$= \sum_{k=1}^{K} \frac{\partial E}{\partial y_k} \times \frac{\partial y_k}{\partial z_k} \times w_{kj} \times \frac{4}{exp(2a)+exp(-2a)+2} \times x_i$$
$$= \sum_{k=1}^{K} \frac{\partial E}{\partial y_k} \times \frac{\partial y_k}{\partial z_k} \times w_{kj} \times \frac{4}{exp(2\sum_{i=0}^{D} w_{ji}^{(1)} x_i) + exp(-2\sum_{i=0}^{D} w_{ji}^{(1)} x_i) + 2} \times x_i$$
$$= \sum_{k=1}^{K} -\frac{t_k(\sum_{l=1}^{K} exp(\sum_{j=0}^{M} w_{lj}^{(2)} h(\sum_{i=0}^{D} w_{ji}^{(1)} x_i)) - exp(\sum_{j=0}^{M} w_{kj}^{(2)} h(\sum_{i=0}^{D} w_{ji}^{(1)} x_i)))}{\sum_{l=1}^{K} exp(\sum_{j=0}^{M} w_{lj}^{(2)} h(\sum_{i=0}^{D} w_{ji}^{(1)} x_i))}$$
$$\times w_{kj} \times \frac{4}{exp(2\sum_{i=0}^{D} w_{ji}^{(1)} x_i) + exp(-2\sum_{i=0}^{D} w_{ji}^{(1)} x_i) + 2} \times x_i$$