Exercises

Computational Intelligence Lab

SS 2017

Machine Learning Institute

Dept. of Computer Science, ETH Zürich

Prof. Dr. Thomas Hofmann

Web http://cil.inf.ethz.ch/

Series 8, May 4-5, 2017 (Neural Networks)

Problem 1 (Neural Networks):

1.
$$\forall x \in \mathbb{R},\ s'(x) = -\frac{-e^{-x}}{(1+e^{-x})^2} = \frac{1}{1+e^{-x}} \cdot \frac{e^{-x}}{1+e^{-x}} = s(x)\cdot(1-s(x)).$$

2. Let's write the network in the form $\hat{y}(x_i, \mathbf{w}) = Pr(x_i, \mathbf{w}) = s(f(x_i, \mathbf{w}))$ where s is a sigmoid. Fix a parameter w_j. Then we have

$$\frac{\partial H(\mathbf{w})}{\partial w_j} = \sum_{i=1}^{n} -y_i \frac{\partial}{\partial w_j}\log(Pr(x_i,\mathbf{w})) - (1-y_i)\frac{\partial}{\partial w_j}\log(1-Pr(x_i,\mathbf{w})),$$

i.e.

$$\frac{\partial H(\mathbf{w})}{\partial w_j} = \sum_{i=1}^{n} -y_i \frac{\partial}{\partial w_j}Pr(x_i,\mathbf{w})\cdot\frac{1}{Pr(x_i,\mathbf{w})} - (1-y_i)\frac{\partial}{\partial w_j}(-Pr(x_i,\mathbf{w}))\cdot\frac{1}{1-Pr(x_i,\mathbf{w})},$$

i.e.

$$\frac{\partial H(\mathbf{w})}{\partial w_j} = \sum_{i=1}^{n} -y_i \frac{\partial}{\partial w_j}f(x_i,\mathbf{w})\cdot\frac{s'(f(x_i,\mathbf{w}))}{Pr(x_i,\mathbf{w})} - (1-y_i)\frac{\partial}{\partial w_j}(-f(x_i,\mathbf{w}))\cdot\frac{s'(f(x_i,\mathbf{w}))}{1-Pr(x_i,\mathbf{w})},$$

i.e. using the previous question,

$$\frac{\partial H(\mathbf{w})}{\partial w_j} = \sum_{i=1}^{n} -y_i \frac{\partial}{\partial w_j}f(x_i,\mathbf{w})\cdot(1-s(f(x_i,\mathbf{w}))) - (1-y_i)\frac{\partial}{\partial w_j}(-f(x_i,\mathbf{w}))\cdot s(f(x_i,\mathbf{w})),$$

hence finally

$$\frac{\partial H(\mathbf{w})}{\partial w_j} = \sum_{i=1}^{n} \frac{\partial}{\partial w_j}f(x_i,\mathbf{w})\cdot(s(f(x_i,\mathbf{w}))-y_i),$$

which we can rewrite as

$$\frac{\partial H(\mathbf{w})}{\partial w_j} = \sum_{i=1}^{n} \frac{\partial}{\partial w_j}f(x_i,\mathbf{w})\cdot(\hat{y}(x_i,\mathbf{w})-y_i).$$

3. The MSE for one training example $\mathbf{x} = (x_i)_{1\le i\le d}$ with label y is given by

$$E(\mathbf{w}) = (s(\sum_{j=1}^{m}\sum_{i=1}^{d} w_{j,i}x_i) - y)^2,$$

and its partial derivative with respect to one single parameter $w_{j',i'}$ is given by

$$\frac{\partial}{\partial w_{j',i'}}E(\mathbf{w}) = 2(s(\sum_{j=1}^{m}\sum_{i=1}^{d} w_{j,i}x_i)-y)\cdot(x_{i'}s'(\sum_{j=1}^{m}\sum_{i=1}^{d} w_{j,i}x_i)) = 2x_{i'}s(\sum_{j=1}^{m} z_j)(1-s(\sum_{j=1}^{m} z_j))(s(\sum_{j=1}^{m} z_j)-y),$$

where $z_j = \sum_{i=1}^{d} w_{j,i}x_i$. Notice that this closed-form formula doesn't require us to actually compute any derivative.