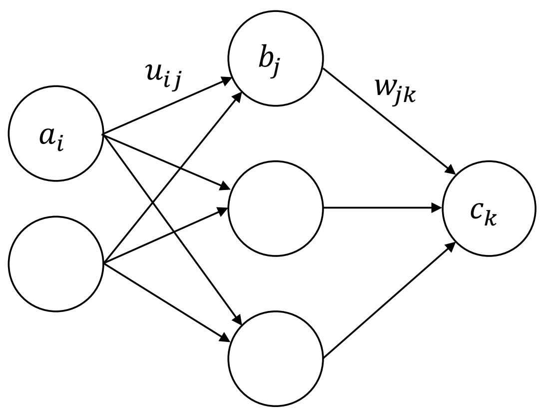
University of Illinois Spring 2018

CS 446: Machine Learning
Homework

Due on Tuesday, February 27, 2018, 11:59 a.m. Central Time

1. **[8 points]** Backpropagation

   Consider the deep net in the figure below consisting of an input layer, an output layer, and a hidden layer. The feed-forward computations performed by the deep net are as follows: every input a_i is multiplied by a set of fully-connected weights u_{ij} connecting the input layer to the hidden layer. The resulting weighted signals are then summed and combined with a bias e_j. This results in the activation signal $z_j = e_j + \sum_i a_i u_{ij}$. The hidden layer applies activation function g on z_j resulting in the signal b_j. In a similar fashion, the hidden layer activation signals b_j are multiplied by the weights connecting the hidden layer to the output layer w_{jk}, a bias f_k is added and the resulting signal h_k is transformed by the output activation function g to form the network output c_k. The loss between the desired target t_k and the output c_k is given by the MSE: $E = \frac{1}{2}\sum_k (c_k - t_k)^2$, where t_k denotes the ground truth signal corresponding to c_k. Training a neural network involves determining the set of parameters $\theta = \{U, W, e, f\}$ that minimize E. This problem can be solved using gradient descent, which requires determining $\frac{\partial E}{\partial \theta}$ for all θ in the model.



   (a) For $g(x) = \sigma(x) = \frac{1}{1+e^{-x}}$, compute the derivative $g'(x)$ of $g(x)$ as a function of $\sigma(x)$.

Your answer:

$$
\begin{aligned}
\sigma(x) &= \frac{1}{1+e^{-x}} \\
\frac{\partial\sigma(x)}{\partial x} &= \frac{\partial}{\partial x}\frac{1}{1+e^{-x}} \\
&= \frac{e^{-x}}{(1+e^{-x})^2} \\
&= \frac{1}{1+e^{-x}}\frac{e^{-x}}{1+e^{-x}} \\
&= \sigma(x)\frac{1+e^{-x}-1}{1+e^{-x}} \\
&= \sigma(x)(\frac{1+e^{-x}}{1+e^{-x}} - \frac{1}{1+e^{-x}}) \\
&= \sigma(x)(1-\sigma(x))
\end{aligned}
$$

(b) We denote by $\delta_k = \frac{\partial E}{\partial h_k}$ the error signal of neuron k in the second linear layer of the network. Compute δ_k as a function of c_k, t_k, g' and h_k.

Your answer:

$$
\begin{aligned}
c_k &= g(h_k) \\
\delta_k &= \frac{\partial E}{\partial h_k} \\
&= \frac{\partial}{\partial h_k}(\frac{1}{2}\sum_k (c_k - t_k)^2) \\
&= \sum_k \big((c_k - t_k)g'(h_k)\big)
\end{aligned}
$$

(c) Compute $\frac{\partial E}{\partial w_{jk}}$. Use δ_k and b_j.

Your answer:

$$\begin{aligned}
h_k &= \sum_j (w_{jk} b_j) + f_k \\
\frac{\partial E}{\partial w_{jk}} &= \frac{\partial E}{\partial h_k} \frac{\partial h_k}{\partial w_{jk}} \\
&= \delta_k \frac{\partial h_k}{\partial w_{jk}} \\
&= \delta_k \frac{\partial}{\partial w_{jk}} \big(\sum_j (w_{jk} b_j) + f_k \big) \\
&= \delta_k \sum_j b_j
\end{aligned}$$

(d) Compute $\frac{\partial E}{\partial f_k}$. Use δ_k.

Your answer:

$$\begin{aligned}
h_k &= \sum_j (w_{jk} b_j) + f_k \\
\frac{\partial E}{\partial f_k} &= \frac{\partial E}{\partial h_k} \frac{\partial h_k}{\partial f_k} \\
&= \delta_k \frac{\partial h_k}{\partial f_k} \\
&= \delta_k \frac{\partial}{\partial f_k} \big(\sum_j (w_{jk} b_j) + f_k \big) \\
&= \delta_k
\end{aligned}$$

(e) We denote by $\psi_j = \frac{\partial E}{\partial z_j}$ the error signal of neuron j in the first linear layer of the network. Compute ψ_j as a function of δ_k, w_{jk}, g' and z_j.

Your answer:

$$
\begin{aligned}
h_k &= \sum_j (w_{jk} b_j) + f_k \\
\frac{\partial h_k}{\partial b_j} &= \sum_j w_{jk}
\end{aligned}
$$

$$
\begin{aligned}
b_j &= g(z_j) \\
\psi_j &= \frac{\partial E}{\partial z_j} \\
&= \frac{\partial E}{\partial h_k} \frac{\partial h_k}{\partial b_j} \frac{\partial b_j}{\partial z_j} \\
&= \delta_k \sum_j \big(w_{jk} g'(z_j) \big)
\end{aligned}
$$

(f) Compute $\frac{\partial E}{\partial u_{ij}}$. Use ψ_j and a_i.

Your answer:

$$
\begin{aligned}
z_j &= \sum_i (u_{ij} a_i) + e_j \\
\frac{\partial E}{\partial u_{ij}} &= \frac{\partial E}{\partial z_j} \frac{\partial z_j}{\partial u_{ij}} \\
&= \psi_j \frac{\partial z_j}{\partial u_{ij}} \\
&= \psi_j \frac{\partial}{\partial u_{ij}} \big(\sum_i (u_{ij} a_i) + e_j \big) \\
&= \psi_j \sum_i a_i
\end{aligned}
$$

(g) Compute $\frac{\partial E}{\partial e_j}$. Use ψ_j.

Your answer:

$$
\begin{aligned}
z_j &= \sum_i (u_{ij} a_i) + e_j \\
\frac{\partial E}{\partial e_j} &= \frac{\partial E}{\partial z_j} \frac{\partial z_j}{\partial e_j} \\
&= \psi_j \frac{\partial z_j}{\partial e_j} \\
&= \psi_j \frac{\partial}{\partial e_j} \big(\sum_i (u_{ij} a_i) + e_j \big) \\
&= \psi_j
\end{aligned}
$$