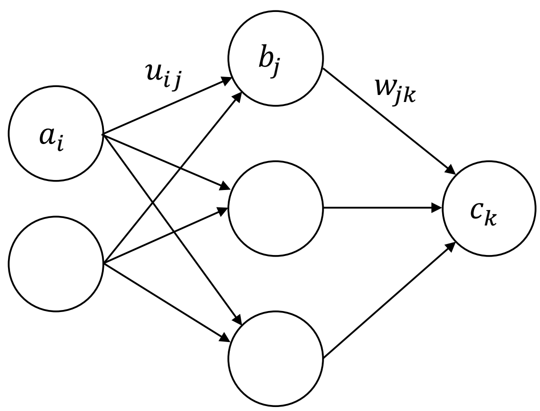
University of Illinois Spring 2018

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:

$$
\begin{aligned}
g\prime(x) &= \frac{d}{dx}\sigma(x) \\
&= \frac{d}{dx}(1+e^{-x})^{-1} \\
&= -1(1+e^{-x})^{-2}e^{-x}(-1) \\
&= \frac{e^{-x}}{(1+e^{-x})^{-2}} \\
&= \frac{1}{1+e^{-x}}\frac{e^{-x}}{1+e^{-x}} \\
&= \frac{1}{1+e^{-x}}\frac{1+e^{-x}-1}{1+e^{-x}} \\
&= \frac{1}{1+e^{-x}}\left(1-\frac{1}{1+e^{-x}}\right) \\
&= \sigma(x)\left(1-\sigma(x)\right) \\
&= \sigma(x)-\sigma(x)^2
\end{aligned}
$$

(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:

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

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

Your answer:

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

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

Your answer:

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

(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: Note that

$$b_j = g(z_j)$$

.

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

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

Your answer: Note that

$$\frac{\partial z_j}{\partial u_{ij}} = \frac{\partial}{\partial u_{ij}} \left(e_j + \sum_i a_i u_{ij} \right) = a_i$$

Now

$$\begin{aligned} \frac{\partial E}{\partial u_{ij}} &= \frac{\partial E}{\partial z_j} \frac{\partial z_j}{\partial u_{ij}} \\ &= \psi_j a_i \end{aligned}$$

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

Your answer: Note that

$$\frac{\partial z_j}{\partial e_j} = \frac{\partial}{\partial e_j} \left(e_j + \sum_i a_i u_{ij} \right) = 1$$

$$\begin{aligned} \frac{\partial E}{\partial e_j} &= \frac{\partial E}{\partial z_j} \frac{\partial z_j}{\partial e_j} \\ &= \psi_j \end{aligned}$$