

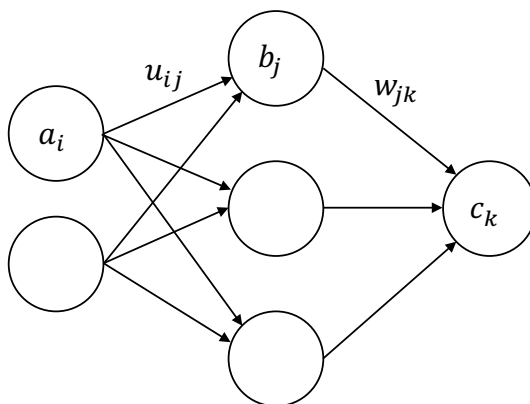
# CS 446: Machine Learning

## 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  $\theta$  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:

- (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  $\delta_k$  as a function of  $c_k$ ,  $t_k$ ,  $g'$  and  $h_k$ .

Your answer:

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

Your answer:

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

Your answer:

- (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  $\psi_j$  as a function of  $\delta_k$ ,  $w_{jk}$ ,  $g'$  and  $z_j$ .

Your answer:

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

Your answer:

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

Your answer: