CS 446 / ECE 449 Homework 6

Naman Shukla

TOTAL POINTS

8 / 8

QUESTION 1

Backpropagation 8 pts

1.1 a **1 / 1**

✓ **- 0 pts** Correct

- 0.5 pts Mathematically correct, but not expressed as function of sigma

- 0.5 pts Minor error

- 0 pts Incorrect

1.2 b **1 / 1**

✓ **- 0 pts** Correct

- 0.25 pts Undefined notation

- 0.5 pts Minor error

- 1 pts Incorrect

1.3 c **1 / 1**

✓ **- 0 pts** Correct

- 0.5 pts Minor error

- 1 pts Incorrect

1.4 d **1 / 1**

✓ **- 0 pts** Correct

- 0.5 pts Minor error

- 1 pts Incorrect

1.5 e **2 / 2**

✓ **- 0 pts** Correct

- 1 pts Miss the summation

- 1 pts Incorrect

- 2 pts Incorrect

1.6 f **1 / 1**

✓ **- 0 pts** Correct

- 0.5 pts Incorrect

- 1 pts Incorrect

1.7 g **1 / 1**

✓ **- 0 pts** Correct

- 0.5 pts Incorrect

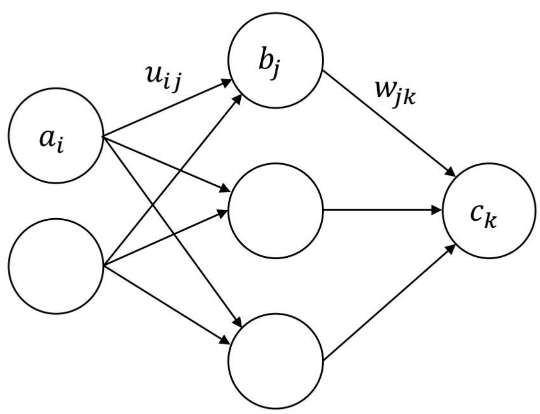
- 1 pts Incorrect

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:

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

(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:

$$E = \frac{1}{2}\sum_k (c_k - t_k)^2$$

$$c_k = g(h_k) \;\text{ and }\; h_k = f_k + \sum_j w_{jk} \cdot b_j$$

$$\delta_k = \frac{\partial E}{\partial h_k} = (c_k - t_k) \cdot g'(h_k)$$

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

Your answer:

$$\frac{\partial E}{\partial w_{jk}} = (c_k - t_k) \cdot g'(h_k) \cdot b_j = \delta_k \cdot b_j$$

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

Your answer:

$$\frac{\partial E}{\partial f_k} = (c_k - t_k) \cdot g'(h_k) \cdot 1 = \delta_k$$

(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.

1.1 a 1 / 1

✓ - 0 pts Correct

- 0.5 pts Mathematically correct, but not expressed as function of sigma

- 0.5 pts Minor error

- 0 pts Incorrect

Your answer:

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

(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:

$$E = \frac{1}{2}\sum_k (c_k - t_k)^2$$

$$c_k = g(h_k) \;\text{ and }\; h_k = f_k + \sum_j w_{jk} \cdot b_j$$

$$\delta_k = \frac{\partial E}{\partial h_k} = (c_k - t_k) \cdot g'(h_k)$$

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

Your answer:

$$\frac{\partial E}{\partial w_{jk}} = (c_k - t_k) \cdot g'(h_k) \cdot b_j = \delta_k \cdot b_j$$

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

Your answer:

$$\frac{\partial E}{\partial f_k} = (c_k - t_k) \cdot g'(h_k) \cdot 1 = \delta_k$$

(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.

1.2 b 1 / 1

- ✓ - **0 pts** Correct
- - **0.25 pts** Undefined notation
- - **0.5 pts** Minor error
- - **1 pts** Incorrect

Your answer:

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

(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:

$$E = \frac{1}{2}\sum_k (c_k - t_k)^2$$

$$c_k = g(h_k) \;\text{ and }\; h_k = f_k + \sum_j w_{jk}\cdot b_j$$

$$\delta_k = \frac{\partial E}{\partial h_k} = (c_k - t_k)\cdot g'(h_k)$$

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

Your answer:

$$\frac{\partial E}{\partial w_{jk}} = (c_k - t_k)\cdot g'(h_k)\cdot b_j = \delta_k \cdot b_j$$

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

Your answer:

$$\frac{\partial E}{\partial f_k} = (c_k - t_k)\cdot g'(h_k)\cdot 1 = \delta_k$$

(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.

1.3 C **1 / 1**

- ✓ **- 0 pts** Correct
- **- 0.5 pts** Minor error
- **- 1 pts** Incorrect

Your answer:

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

(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:

$$E = \frac{1}{2}\sum_k (c_k - t_k)^2$$

$$c_k = g(h_k) \;\text{ and }\; h_k = f_k + \sum_j w_{jk} \cdot b_j$$

$$\delta_k = \frac{\partial E}{\partial h_k} = (c_k - t_k) \cdot g'(h_k)$$

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

Your answer:

$$\frac{\partial E}{\partial w_{jk}} = (c_k - t_k) \cdot g'(h_k) \cdot b_j = \delta_k \cdot b_j$$

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

Your answer:

$$\frac{\partial E}{\partial f_k} = (c_k - t_k) \cdot g'(h_k) \cdot 1 = \delta_k$$

(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.

1.4 d **1 / 1**

- ✓ **- 0 pts** Correct
- **- 0.5 pts** Minor error
- **- 1 pts** Incorrect

Your answer:

$$E = \frac{1}{2}\sum_k (c_k - t_k)^2$$

$$c_k = g(h_k) \;\text{ and }\; h_k = f_k + \sum_j w_{jk} \cdot b_j$$

$$b_j = g(z_j) \;\text{ and }\; z_j = e_j + \sum_i a_i u_{ij}$$

Hence,

$$\psi_j = \frac{\partial E}{\partial z_j} = \sum_k (c_k - t_k) \cdot g'(h_k) \cdot w_{jk} \cdot g'(z_i) = g'(z_j) \sum_k \delta_k \cdot w_{jk}$$

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

Your answer:

$$\frac{\partial E}{\partial u_{ij}} = \sum_k (c_k - t_k) \cdot g'(h_k) \cdot w_{jk} \cdot g'(z_i) \cdot a_i = \psi_j \cdot a_i$$

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

Your answer:

$$\frac{\partial E}{\partial e_j} = \sum_k (c_k - t_k) \cdot g'(h_k) \cdot w_{jk} \cdot g'(z_i) \cdot 1 = \psi_j$$

1.5 e **2 / 2**

- ✓ **- 0 pts** Correct
- **- 1 pts** Miss the summation
- **- 1 pts** Incorrect
- **- 2 pts** Incorrect

Your answer:

$$E = \frac{1}{2}\sum_k (c_k - t_k)^2$$

$$c_k = g(h_k) \text{ and } h_k = f_k + \sum_j w_{jk} \cdot b_j$$

$$b_j = g(z_j) \text{ and } z_j = e_j + \sum_i a_i u_{ij}$$

Hence,

$$\psi_j = \frac{\partial E}{\partial z_j} = \sum_k (c_k - t_k) \cdot g'(h_k) \cdot w_{jk} \cdot g'(z_i) = g'(z_j) \sum_k \delta_k \cdot w_{jk}$$

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

Your answer:

$$\frac{\partial E}{\partial u_{ij}} = \sum_k (c_k - t_k) \cdot g'(h_k) \cdot w_{jk} \cdot g'(z_i) \cdot a_i = \psi_j \cdot a_i$$

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

Your answer:

$$\frac{\partial E}{\partial e_j} = \sum_k (c_k - t_k) \cdot g'(h_k) \cdot w_{jk} \cdot g'(z_i) \cdot 1 = \psi_j$$

1.6 f 1 / 1

- ✓ **- 0 pts** Correct
- **- 0.5 pts** Incorrect
- **- 1 pts** Incorrect

Your answer:

$$E = \frac{1}{2}\sum_k (c_k - t_k)^2$$

$$c_k = g(h_k) \text{ and } h_k = f_k + \sum_j w_{jk} \cdot b_j$$

$$b_j = g(z_j) \text{ and } z_j = e_j + \sum_i a_i u_{ij}$$

Hence,

$$\psi_j = \frac{\partial E}{\partial z_j} = \sum_k (c_k - t_k) \cdot g'(h_k) \cdot w_{jk} \cdot g'(z_i) = g'(z_j) \sum_k \delta_k \cdot w_{jk}$$

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

Your answer:

$$\frac{\partial E}{\partial u_{ij}} = \sum_k (c_k - t_k) \cdot g'(h_k) \cdot w_{jk} \cdot g'(z_i) \cdot a_i = \psi_j \cdot a_i$$

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

Your answer:

$$\frac{\partial E}{\partial e_j} = \sum_k (c_k - t_k) \cdot g'(h_k) \cdot w_{jk} \cdot g'(z_i) \cdot 1 = \psi_j$$

1.7 g **1 / 1**

- ✓ **- 0 pts** Correct
- **- 0.5 pts** Incorrect
- **- 1 pts** Incorrect