Boston University
Department of Electrical and Computer Engineering

ENG EC 414 Introduction to Machine Learning

HW 10 Solution

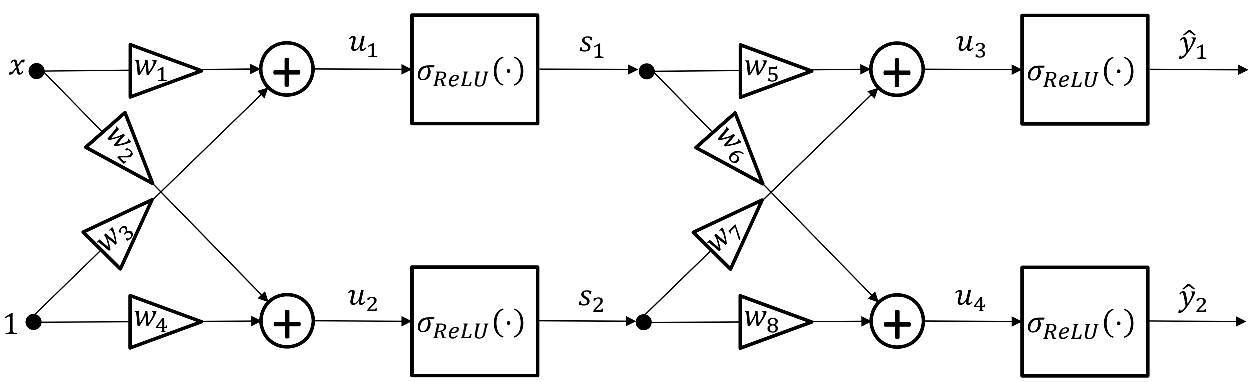
Issued: Fri 20 Nov 2020 **Due:** 10:00am Fri 4 Dec 2020 in Gradescope

Important: Before you proceed, please read the documents pertaining to *Homework formatting and submission guidelines* in the Homeworks section of Blackboard. **In particular, for computer assignments you are prohibited from using any online code or built-in MATLAB functions except as indicated in the problem or skeleton code (when provided).**

Important: To obtain full grade, please clearly motivate all your answers.

Note: Problem difficulty = number of coffee cups ☕

Problem 10.1 [13pts] *(Backprop by Hand)* A two-layer feed-forward neural network with ReLU activation function, $\sigma_{ReLU}(t) = \max\{0, t\}$, is shown in the figure below. Let $\ell(\hat{y}_1, \hat{y}_2, y) = (\hat{y}_1 + \hat{y}_2 - y)^2$ be the loss function, $(x = 0.5, y = 1)$ be a training sample, and initial weights $w_1 = -1, w_2 = 1, w_3 = 1, w_4 = 1, w_5 = -1, w_6 = 1, w_7 = 2, w_8 = 2$. Note that the second input is fixed to 1, it is the bias term.



(a) [3pts] Compute the values of $u_1, u_2, s_1, s_2, u_3, u_4, \hat{y}_1, \hat{y}_2$, and $\ell(\hat{y}_1, \hat{y}_2, y)$ in the first forward pass iteration of the Backprop algorithm.

(b) [10pts] Compute the values of the partial derivatives of the loss with respect to $\hat{y}_1, \hat{y}_2, u_3, u_4, w_5, w_6, w_7, w_8, s_1, s_2$ in the first backward pass iteration of the Backprop algorithm.

Solution: Training example: $(x = 0.5, y = 1)$. Initial weights: $w_1 = -1, w_2 = 1, w_3 = 1, w_4 = 1, w_5 = -1, w_6 = 1, w_7 = 2, w_8 = 2$.

(b) [10pts] Backward Pass First Iteration

Step #	Quantity	Expression	Value
1	$\frac{\partial \ell}{\partial \hat{y}_1}$	$2(\hat{y}_1 + \hat{y}_2 - y)$	10
2	$\frac{\partial \ell}{\partial \hat{y}_2}$	$2(\hat{y}_1 + \hat{y}_2 - y)$	10
3	$\frac{\partial \ell}{\partial u_3}$	$\sigma'(u_3) \cdot \frac{\partial \ell}{\partial \hat{y}_1} = 1(u_3 > 0) \cdot \frac{\partial \ell}{\partial \hat{y}_1}$	10
4	$\frac{\partial \ell}{\partial u_4}$	$\sigma'(u_4) \cdot \frac{\partial \ell}{\partial \hat{y}_2} = 1(u_4 > 0) \cdot \frac{\partial \ell}{\partial \hat{y}_2}$	10
5	$\frac{\partial \ell}{\partial w_5}$	$s_1 \cdot \frac{\partial \ell}{\partial u_3}$	5
6	$\frac{\partial \ell}{\partial w_6}$	$s_1 \cdot \frac{\partial \ell}{\partial u_4}$	5
7	$\frac{\partial \ell}{\partial w_7}$	$s_2 \cdot \frac{\partial \ell}{\partial u_3}$	15
8	$\frac{\partial \ell}{\partial w_8}$	$s_2 \cdot \frac{\partial \ell}{\partial u_4}$	15
9	$\frac{\partial \ell}{\partial s_1}$	$w_5 \cdot \frac{\partial \ell}{\partial u_3} + w_6 \cdot \frac{\partial \ell}{\partial u_4} = -\frac{\partial \ell}{\partial u_3} + \frac{\partial \ell}{\partial u_4}$	0
10	$\frac{\partial \ell}{\partial s_2}$	$w_7 \cdot \frac{\partial \ell}{\partial u_3} + w_8 \cdot \frac{\partial \ell}{\partial u_4} = 2\frac{\partial \ell}{\partial u_3} + 2\frac{\partial \ell}{\partial u_4}$	40

(a) [3pts] Forward Pass First Iteration

Step #	Quantity	Expression	Value
1	u_1	$w_1 x + w_3 = (1 - x)$	0.5
2	u_2	$w_2 x + w_4 = (1 + x)$	1.5
3	s_1	$\max\{0, u_1\}$	0.5
4	s_2	$\max\{0, u_2\}$	1.5
5	u_3	$w_5 s_1 + w_7 s_2 = (-s_1 + 2s_2)$	2.5
6	u_4	$w_6 s_1 + w_8 s_2 = (s_1 + 2s_2)$	3.5
7	$\hat{y}_1$	$\max\{0, u_3\}$	2.5
8	$\hat{y}_2$	$\max\{0, u_4\}$	3.5
9	ℓ	$(\hat{y}_1 + \hat{y}_2 - y)^2$	25

Problem 10.2 [13pts] *(Function Approximation with Neural Network)* Let

$$h(x) = \begin{cases} 2x + 2 & -1 \le x \le 0 \\ -x + 2 & 0 \le x \le 2 \\ 0 & \text{else} \end{cases}$$

Let $\sigma_{ReLU}(t) = \max\{0, t\}$ be the Rectifier Linear Unit activation function. Find values of (β_1, γ_1), $(\alpha_2, \beta_2, \gamma_2)$, and (β_3, γ_3) such that for all x,

$$h(x) = \underbrace{\sigma_{ReLU}(\beta_1 + \gamma_1 x)}_{h_1(x)} + \underbrace{\alpha_2 \cdot \sigma_{ReLU}(\beta_2 + \gamma_2 x)}_{h_2(x)} + \underbrace{\sigma_{ReLU}(\beta_3 + \gamma_3 x)}_{h_3(x)}$$

Sketch the graphs of $h_1(x), h_2(x), h_3(x)$, and $h(x)$ and properly label axes and key points.

Solution: **Problem 10.3** [13pts] $\boxed{\beta_1 = 2, \gamma_1 = 2, \alpha_2 = -3, \beta_2 = 0, \gamma_2 = 1, \beta_3 = -2, \gamma_3 = 1}$ The solution can be visualized in the figure below. The function $h(x)$ is continuous and piecewise linear.

The graph of $h(x)$ is zero for $(-\infty, -1]$ and then increases linearly with a positive slope of 2 in the interval $(-1, 0]$. Thus in the range $(-\infty, 0]$ the graph of $h(x)$ exactly matches the graph of a ReLU activation function shifted to the left by 1 and scaled to have slope 2, i.e., $2\sigma_{ReLU}(x+1) = 2\max\{0, (x+1)\} = \max\{2 \times 0, 2 \times (x+1)\} = \max\{0, 2x+2\} = \sigma_{ReLU}(2x+2) =: h_1(x)$.

Over the interval $(0, 2]$ the graph of $h(x)$ decreases linearly with slope -1. However, in this interval $h_1(x)$ is already increasing linearly with slope 2. To counteract this increase, we need to add to $h_1(x)$ a function $h_2(x)$ which linearly decreases with a slope equal to $-1-2=-3$ (which would yield a net slope of -1 for the sum) and is also zero for all $x \leq 0$ so as to not change the values of $h_1(x)$ in the range $(-\infty, 0]$. This is nothing but the graph of a ReLU activation with slope -3, i.e., $-3\sigma_{ReLU}(x) =: h_2(x)$.

Finally, over the interval $(2, \infty)$ the graph of $h(x)$ is zero whereas the sum $h_1(x) + h_2(x)$ is decreasing linearly with a slope equal to -1. To counteract this decrease, we need to add to $h_1(x) + h_2(x)$ a function $h_3(x)$ which linearly increases with a slope equal to $+1$ (which would yield a net slope of 0 for the sum) and is also zero for all $x \leq 2$ so as to not change the values of $h_1(x) + h_2(x)$ in the range $(-\infty, 2]$. This is nothing but the graph of a ReLU activation shifted to the right by 1, i.e., $\sigma_{ReLU}(x-2) =: h_3(x)$.