CS 189 Introduction to Machine Learning
Fall 2017 HW7

This homework is due **Saturday, October 14 at 10pm.**

1 Getting Started

You may typeset your homework in latex or submit neatly handwritten and scanned solutions. Please make sure to start each question on a new page, as grading (with Gradescope) is much easier that way! Deliverables:

1. Submit a PDF of your writeup to assignment on Gradescope, "HW[n] Write-Up"
2. Submit all code needed to reproduce your results, "HW[n] Code".
3. Submit your test set evaluation results, "HW[n] Test Set".

After you've submitted your homework, be sure to watch out for the self-grade form.

(a) Before you start your homework, write down your team. Who else did you work with on this homework? List names and email addresses. In case of course events, just describe the group. How did you work on this homework? Any comments about the homework?

None. I work alone
Comments The skeleton code is terrible! It is still buggy and not updated until last minutes. MAKE SURE IT WORKS ON YOUR COMPUTER BEFORE RELEASING IT.

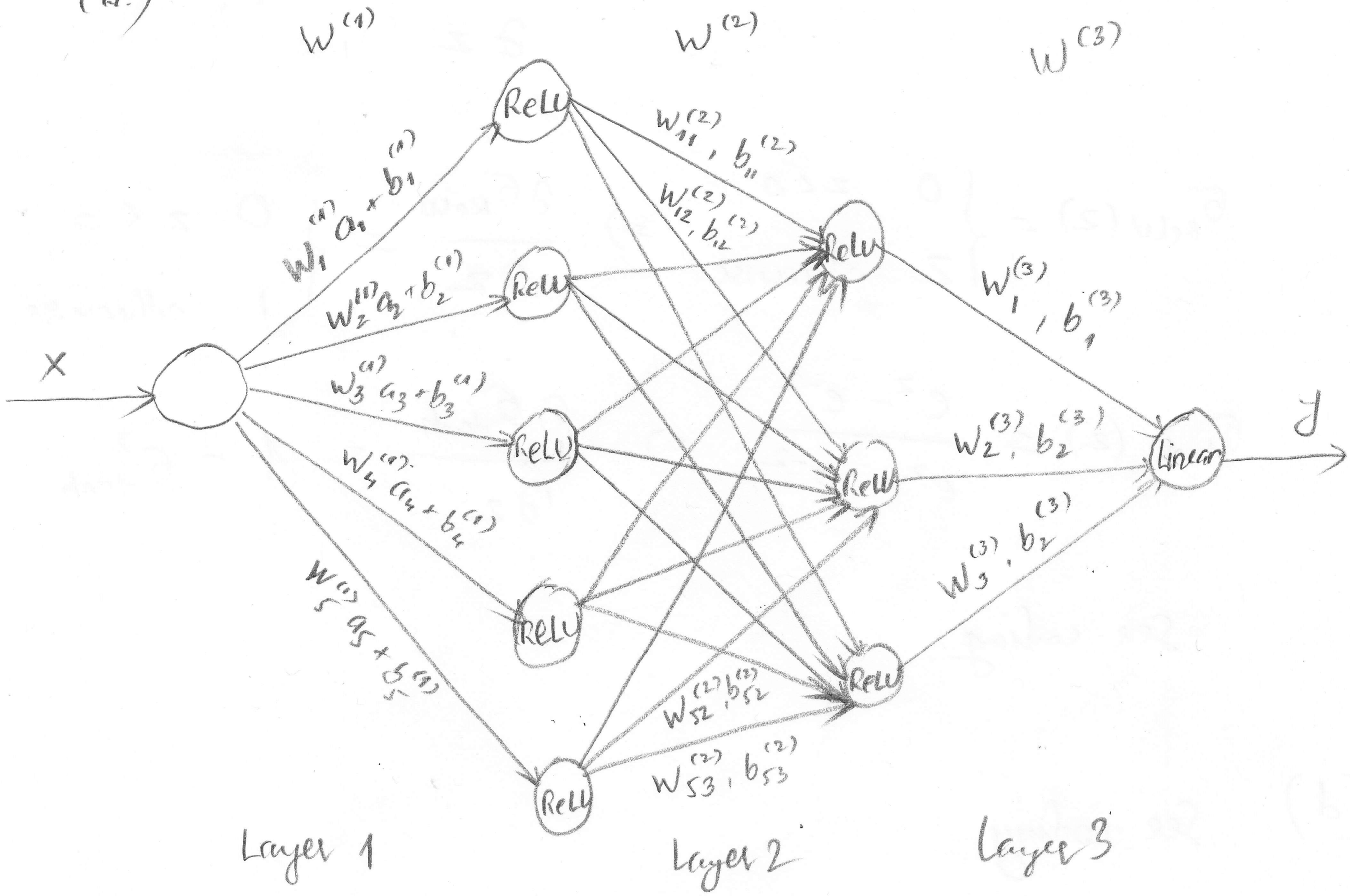
(b) Please copy the following statement and sign next to it:

I certify that all solutions are entirely in my words and that I have not looked at another student's solutions. I have credited all external sources in this write up.

I certify that all solutions are entirely in my words and that I habe not looked at another student's solutions. I have credited all external sources in this write up.

Problem #2

(a.)



(b.)

$$MSE(\vec{\hat{y}}) = \frac{1}{2} \sum_{i=1}^{n} \| y_i - \hat{y}_i \|_2^2$$

$$\Rightarrow \frac{\partial MSE}{\partial \vec{\hat{y}}} = \vec{\hat{y}} - \vec{y}$$

See coding

(c) $\sigma_{linear}(z) = z \Rightarrow \dfrac{\partial \sigma_{linear}}{\partial z} = 1$

$$\sigma_{ReLU}(z) = \begin{cases} 0 & z < 0 \\ z & \text{otherwise} \end{cases} \Rightarrow \frac{\partial \sigma_{ReLU}}{\partial z} = \begin{cases} 0 & z < 0 \\ 1 & \text{otherwise} \end{cases}$$

$$\sigma_{tanh}(z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}} \Rightarrow \frac{\partial \sigma_{tanh}}{\partial z} = 1 - \sigma_{tanh}^2$$

See coding

(d) See coding

(e) $$\frac{\partial MSE}{\partial \vec{a}_i} = \frac{\partial MSE}{\partial \vec{a}_{i+1}} \underbrace{\frac{\partial \vec{a}_{i+1}}{\partial \vec{a}_i}}_{\frac{\partial \sigma(\vec{z}_i)}{\partial \vec{z}_i} \frac{\partial \vec{z}_i}{\partial a_i}}$$

$\vec{z}_i = W_i \vec{a}_i + b_i$

$$\Rightarrow \frac{\partial MSE}{\partial \vec{a}_i} = \frac{\partial MSE}{\partial \vec{a}_{i+1}} \frac{\partial \sigma(\vec{z}_i)}{\partial \vec{z}_i} W_i$$

See coding

f.)

$$\frac{\partial MSE}{\partial w_i} = \frac{\partial MSE}{\partial \vec{a}_{i+1}} \underbrace{\frac{\partial \vec{a}_{i+1}}{\partial w_i}}_{\frac{\partial \sigma(\vec{z}_i)}{\partial \vec{z}_i} \underbrace{\frac{\partial \vec{z}_i}{\partial w_i}}_{\vec{a}_i}}$$

$$\vec{z}_i = w_i \vec{a}_i + \vec{b}_i$$

$$\frac{\partial MSE}{\partial w_i} = \frac{\partial MSE}{\partial \vec{a}_{i+1}} \frac{\partial \sigma(\vec{z}_i)}{\partial \vec{z}_i} \vec{a}_i$$

Similarly

$$\frac{\partial MSE}{\partial \vec{b}_i} = \frac{\partial MSE}{\partial \vec{a}_{i+1}} \frac{\partial \sigma(\vec{z}_i)}{\partial \vec{z}_i} \cancel{\frac{\partial \vec{z}_i}{\partial \vec{b}_i}} \to I$$

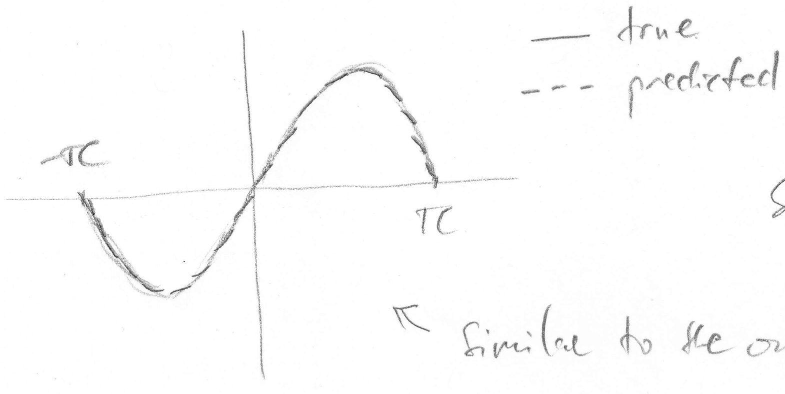
g.) Output

ReLU MSE : 0.00074328

linear MSE : 0.12079632

tanh MSE : 0.00173546

Plot :



see coding

similar to the output

h.) not enough time to run

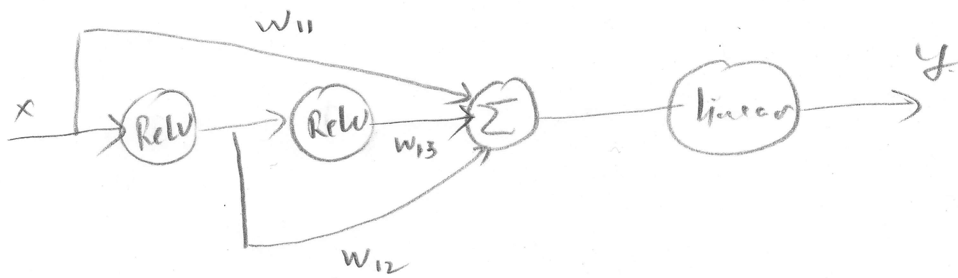
i.) not enough time to run

j.) skip

k.) skip

Problem #3

Consider the net-work:



1.) Write the out-put for each layers (from $x \to y$)

solution:

$$a_1 = w_{11} x + b_{11}$$
$$a_2 = w_{12}\,\sigma_{ReLU}(x) + b_{12}$$
$$a_3 = w_{13}\,\sigma_{ReLU}(\sigma_{ReLU}(x)) + b_{13}$$

$$a_4 = w_4 (a_1 + a_2 + a_3) + b_{14}$$

$$y = \sigma_{linear}(a_4) = a_4$$

where b's are noise

2.) Define error & take the derivative of erro w.r.t x given a_1, a_2, a_3 & a_4

Solution:

$$MSE(\vec{y_p}) = \frac{1}{2}\|\vec{y_p} - y\|_2^2$$

$$\frac{\partial MSE}{\partial \vec{x}} = \frac{\partial MSE}{\partial a_4}\underbrace{\frac{\partial a_4}{\partial a_1}\frac{\partial a_1}{\partial x}}_{w_{11}} + \frac{\partial MSE}{\partial a_4}\frac{\partial a_4}{\partial a_2}\underbrace{\frac{\partial a_2}{\partial x}}_{w_{12}\frac{\partial \sigma_{REW}}{\partial x}} + \frac{\partial MSE}{\partial a_4}\frac{\partial a_4}{\partial a_3}\underbrace{\frac{\partial a_3}{\partial x}}_{w_{13}\frac{\partial ReLU}{\partial a_2}\underbrace{\frac{\partial a_2}{\partial x}}_{w_{12}\frac{\partial \sigma_{ReLU}}{\partial x}}}$$